

## ONTOLOGÍAS PARA LA BÚSQUEDA DE IMÁGENES MÉDICAS EN BASE DE DATOS

Matías Agüero<sup>1</sup>, Carlos Alvez<sup>2</sup>, Marcela Vegetti<sup>1</sup>

<sup>1</sup>Facultad Regional Santa Fe – Universidad Tecnológica Nacional

Lavaise 610, Santa Fe, Santa Fe, S3002GJC, Argentina

<sup>2</sup>Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos

Av. Tavella 1424, Concordia, Entre Ríos - CP 3200

mjaguero@gmail.com, caralv@fcad.uner.edu.ar, mvegetti@santafe-conicet.gov.ar

### Resumen

Las imágenes digitales en medicina están reemplazando cada día más a los antiguos formatos que ocupaban inmensos espacios físicos destinados a almacenar estas placas, papeles, etc. Los antiguos formatos de imágenes una vez almacenadas en depósitos no se volvían a consultar, ocasionaban gastos y ocupaban lugar a las instituciones médicas. Ese fue uno de los motivos por los que surgió el estándar DICOM, que establece un mismo formato de imagen digital para todos los estudios en medicina, y a la vez, incluye información del paciente y del estudio realizado. Al contar con tanta información digital, surge el problema de encontrar formas de relacionar esta información, de manera que sea útil para los profesionales en medicina. De aquí la necesidad de la utilización de Ontologías para relacionar todas las imágenes de manera que el profesional pueda recuperar imágenes/diagnósticos similares a través del uso de las Ontología creadas sobre las imágenes DICOM.

**Palabras clave:** Base de datos OR.

DICOM. Ontologías. OWL.

### Contexto

Este proyecto da continuidad al trabajo realizado en la tesis doctoral “Modelos para la recuperación de imágenes por

similitud en Bases de Datos Objeto-Relacionales” [1], cuyo objetivo fue la recuperación de imágenes por similitud, sea por contenido físico como semántico, basado principalmente en modelos de representación de imágenes en Bases de Datos Objeto-Relacionales (BDOR) [2].

### Introducción

Las imágenes en medicina son una base fundamental para poder brindar un diagnóstico acertado sobre el estado de salud de los pacientes. Su adquisición, almacenamiento, traslado y visualización constituyen un desafío por la gran cantidad de éstas que existen en los centros de salud. Las imágenes digitales proporcionan diferentes ventajas sobre los antiguos formatos físicos desde el momento de la adquisición hasta el de la visualización de las mismas. Algunos de estas ventajas son:

- Utilización de estándares en el formato de las imágenes.
- Reducción de espacios para el almacenamiento.
- Mejores tiempos de respuesta para la disponibilidad de las mismas debido a una mejor organización.
- Posibilita llevar una mejor historia clínica de los pacientes.
- Las mismas pueden ser visualizadas, por diferentes especialistas, al mismo tiempo, en un lugar diferente al de donde se realizó el estudio, gracias a la

posibilidad de comunicación que proporcionan.

- El paciente no necesariamente tiene que cargar con los estudios que se realizó.

DICOM (Digital Image and Communications in Medicine / Tratamiento Digital de Imágenes y Comunicaciones en Medicina) [3], es el estándar que describe detalladamente los medios para dar formato e intercambiar imágenes e información entre diferentes dispositivos o equipos médicos de diagnóstico por imágenes. Una imagen médica no tiene sentido sola. Es importante la información del contexto en el que se ha tomado la misma. En el contexto de una imagen DICOM podemos encontrarnos con datos del paciente, del estudio que se le está realizando, del equipo que ha efectuado el estudio de las imágenes tomadas (número de tomas realizadas, separación entre cada imagen, dimensión de las imágenes, etc.). Los formatos genéricos de imágenes (TIFF, JPEG, PNG, etc.) no cubren las necesidades anteriormente nombradas ya que no contemplan en su composición esta información. Asimismo, distintos estudios médicos e imágenes de un paciente están relacionados entre sí. Un estudio que se realiza a un paciente puede haber sido derivado de uno anteriormente realizado. Lo más importante es poder identificar a quien corresponde cada examen, el momento en el que se lo realizó y quien lo solicitó. Todo esto es posible a través del estándar DICOM, el cual propone organizar las imágenes de un paciente como se muestra en la Fig. 1.

Actualmente existen sistemas que realizan la gestión de imágenes, conocidos como PACS (Picture Archiving and Communications Systems / Sistemas de Almacenamiento y Comunicación de Imágenes), en donde la comunicación en ambiente de red es la

parte medular para el diseño de aplicaciones y se apoya en el protocolo DICOM para la gestión de la imagen diagnóstica, donde en los últimos años se ha producido una generalización a nivel mundial del uso de este protocolo.

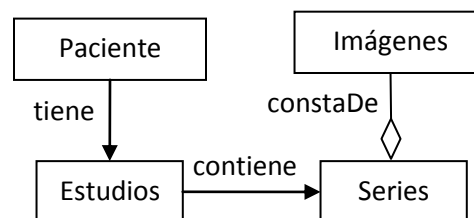


Fig. 1. Organización Imágenes DICOM

A pesar de contar con las imágenes en formato DICOM, también es necesario poder establecer relaciones entre las mismas de manera de a partir de ciertas reglas o datos poder recuperar las imágenes y diagnósticos similares que se encuentran en la base de datos, y de esta manera proporcionar una herramienta útil a los profesionales en medicina para el diagnóstico en los pacientes. Estas relaciones o búsquedas por similitud se logran a través de la utilización de Ontologías [4].

Una ontología describe un conjunto de entidades y sus relaciones, en donde se especifica formalmente las entidades en un dominio y sus propiedades (atributos). Permite que las personas y los sistemas de información accedan al conocimiento almacenado en su interior. Grandes y sofisticadas ontologías se han desarrollado para la genética y la anatomía humana. Actualmente hay cerca de 200 ontologías diferentes para una variedad de disciplinas biomédicas (<http://bioportal.bioontology.org>). Las ontologías son "computables", es decir la información altamente estructurada que contienen puede ser utilizada para mejorar la recuperación de información, el descubrimiento de conocimiento y razonamiento automatizado.

## Líneas de Investigación, Desarrollo e Innovación

En este plan confluyen varias líneas de investigación, y entre las más importantes encontramos: tratamiento de imágenes en bases de datos Objeto-Relacionales (OR) y Ontologías.

La primera de las líneas mencionada está relacionada con la construcción de modelos para la representación de datos relacionados con imágenes médicas, de manera que los mismos puedan consultarse, recuperarse y compararse de manera simple y eficiente. Esto además, debe respetar las especificaciones del estándar adoptado.

Estos modelos, tienen estructuras de datos complejas y pueden ocasionar algunos inconvenientes si se trabaja con el modelo de datos relacional. En primer lugar, los metadatos relacionados a las imágenes son estructuras no atómicas, por lo que en las bases de datos relacionales (para respetar la primera forma normal), se deben separar en varias tablas, lo que hace menos eficiente el tratamiento. Por ejemplo, en la comparación de dos conjuntos de metadatos, necesita realizar varias operaciones de join. En segundo lugar, el procesamiento de los datos se debe hacer desde lenguajes anfitriones (C++, java, etc.), que pueden presentar problemas de compatibilidad, dado que se deben transferir los datos generados por métodos implementados en distintos lenguajes a la base de datos para ser almacenados. Este tema no es trivial, y se debe tratar para cada lenguaje en particular. Además, para que un lenguaje pueda tratar con los datos, se deben crear cursores que comúnmente deben ser transferidos a través de una red. Esto puede provocar problemas de tráfico de red innecesarios, dado que muchos de

estos procesos retornan un conjunto pequeño de datos.

Frente a estos inconvenientes de las bases de datos relacionales, aquí se ha optado por utilizar el modelo OR [2]. Este modelo permite dar soluciones a las limitaciones de las bases de datos. En primer lugar, permite definir tipos de datos que pueden contener estructuras complejas como colecciones, objetos grandes, etc., sin la limitación de la primera forma normal. En segundo lugar, permite definir e implementar el comportamiento de los datos creando los métodos para la gestión de las instancias de los tipos definidos, y así facilitar el acceso seguro desde las aplicaciones que los utilicen. Por último, los sistemas de bases de datos OR, proveen diferentes paquetes para la gestión de tecnologías utilizar en este trabajo, como para la gestión de imágenes DICOM [5] [6] y tecnologías para el tratamiento de datos semánticos [7].

Las facilidades mencionadas, permiten crear infraestructuras que extienden los servicios del sistema de gestión de base de datos [8]. Estas le permiten a este sistema tratar con datos de un dominio específico, en este caso, para aplicaciones en medicina.

Este plan de trabajo aborda también una línea de investigación relacionada con las ontologías. Una ontología es “una especificación formal y explícita de una conceptualización compartida” [9]. Una conceptualización es un modelo abstracto de algún fenómeno de la realidad que identifica los conceptos más relevantes de dicho fenómeno. Estos conceptos, así como las restricciones que pudieran existir sobre su utilización y sus relaciones deben estar explícitamente definidos en un lenguaje entendible por máquinas. Asimismo, la noción de compartido que menciona la definición se refiere a que una ontología captura

conocimiento que no es privado a un único individuo sino consensuado por un grupo en un dominio determinado [10].

Desde las últimas décadas las ontologías se han vuelto relevantes y están siendo ampliamente utilizadas en diversos campos como la Ingeniería del Conocimiento, la Inteligencia Artificial y las Ciencias de la Computación, en aplicaciones relacionadas con la gestión del conocimiento, el procesamiento del lenguaje natural, el comercio electrónico, la información de integración inteligente, recuperación de información, diseño e integración de bases de datos, la bioinformática, la educación y la web Semántica. El crecimiento que se observa en el uso de las ontologías se debe en gran medida en la capacidad que brindan para la definición de la semántica de los términos en un dominio y para la inferencia de nuevo conocimiento basado en la conceptualización explícita y formalmente definida.

Las ontologías, entonces, permitirán enriquecer semánticamente los metadatos DICOM para poder inferir nuevo conocimiento sobre estas imágenes y lograr hacer más eficiente la búsqueda de imágenes por similitud de contenido tanto física como semántico. Para lograr esto, esta línea de investigación abordará un conjunto de actividades que van desde la búsqueda de ontologías, el anotado semántico de las imágenes hasta la especificación de las consultas necesarias para la búsqueda por similitud.

La búsqueda, análisis y selección de ontologías que permitan el anotado semántico de las imágenes es una meta central de este trabajo. En este sentido, se deberá enfocar la búsqueda en ontologías que permitan representar el estándar DICOM y en ontologías que permitan extender los metadatos DICOM con información respecto a enfermedades y diagnósticos. Se pretende focalizar esta

búsqueda en ontologías que representen estándares, tesauros y catálogos ampliamente aceptados en el ámbito biomédico. Entre ellas se puede mencionar RadLex (Radiology Lexicon), la clasificación internacional de enfermedades (ICD - International Classification of Diseases), el tesoro NCI, la ontología FMA (Foundational Model of Anatomy).

Una vez seleccionada las ontologías se deberá definir la manera en que se vincularan las ontologías seleccionadas (la que representa al estándar y las de enfermedades y diagnósticos). Asimismo se deberá especificar la manera en que las imágenes DICOM almacenadas serán semánticamente anotadas con los conceptos de las ontologías. Finalmente, se deberán escribir las consultas semánticas que permitirán las búsquedas por similitud

### Resultados y Objetivos

El objetivo general de esta investigación es crear una arquitectura genérica que permita la recuperación de imágenes en medicina (formato DICOM) por similitud semántica, a través de la utilización de Ontologías.

Los objetivos específicos son:

- Crear modelos en una base de datos Objeto-Relacional que soporten el estándar DICOM.
- Crear ontologías en la base de datos para la búsqueda semántica de imágenes y diagnósticos.
- Crear un software para la inserción y recuperación de imágenes DICOM.

### Formación de Recursos Humanos

Como ya se mencionara antes, este trabajo se encuentra enmarcado en el desarrollo de una tesis de maestría, con lo cual este plan de trabajo contribuirá a la

formación del tesista en diferentes dominios: el tratamiento de imágenes y los estándares que describen el formato para el intercambio de las mismas, así como también, diversos aspectos en el campo de las ontologías y tecnologías semánticas.

El equipo de trabajo está compuesto por el tesista junto con el Director y Co-Director de la misma. En cuanto al integrante Matías Agüero, este se encuentra realizando el proyecto final o tesis para la obtención de la Maestría en Ingeniería de Sistemas, la cual ya cursó en la Universidad Tecnológica Nacional (UTN).

Los integrantes participarán de reuniones científicas y técnicas que permitan actualizar los conocimientos en el tema de interés. También se trabajará en la presentación de trabajos en congresos nacionales/internacionales relacionados con el área Sistemas de Información, así como en revistas internacionales con referato. Estos trabajos servirán para divulgar los conocimientos obtenidos durante el trabajo de investigación.

### Referencias

- [1] Alvez Carlos E. “Modelos para la recuperación de imágenes por similitud en Bases de Datos Objeto-Relacionales”. Tesis Doctoral. Santa Fe, Argentina, Abril de 2012. ISBN 978-987-33-2249-5.
- [2] Melton Jim, “(ISO-ANSI Working Draft) Foundation (SQL /Foundation)”, ISO/IEC 9075-2:2003 (E), United States of America (ANSI), 2003.
- [3] Página oficial de NEMA. <http://medical.nema.org/>
- [4] Charles E. Kahn, Curtis P. Langlotz, David S. Channin and Daniel L. Rubin, “Informatics in Radiology: An Information Model of the DICOM Standard” Radiographics. 2011 Jan-Feb; 31(1): 295–304.
- [5] Sue Pelski,. Oracle Multimedia. DICOM Developer's Guide. 11g Release 2 (11.2) E10778-03. August 2010.
- [6] Alvez C. y Vecchiotti A. “Representación y Recuperación de Imágenes Médicas en Bases de Datos Objeto-Relacionales”. Jornadas Argentinas de Informática e Investigaciones Operativas: 36<sup>a</sup> JAIIO– Simposio de Informática y Salud: SIS-2007. Mar del Plata. Agosto de 2007.
- [7] Chuck Murray. Oracle Database. Semantic Technologies Developer's Guide. 11g Release 2 (11.2) E25609-06. January 2014.
- [8] Carlos E. Alvez, Aldo R. Vecchiotti. Combining Semantic and Content Based Image Retrieval in ORDBMS. KES'2010, Lecture Notes in Computer Science, 2010, Volume 6277/2010, 44-53. Editors Rossitza Setchi. Ivan Jordanov, Robert J. Howlett, Lakhmi C. Jain. Springer-Verlag Berlin Heidelberg, 2010.
- [9] Borst, W.N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. PhD. Thesis. University of Twente. 1997.
- [10] Studer, R, V. Benjamins, y D. Fensel. Knowledge Engineering: Principles and Methods. IEEE Transactions on Data and Knowledge Engineering, 25, 161-197. 1998.