

# FRAMEWORK BASADO EN ONTOLOGÍAS PARA LA RECUPERACIÓN DE IMÁGENES MÉDICAS

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 son una base fundamental para poder brindar un diagnóstico acertado sobre el estado de salud de los pacientes. Su adquisición, almacenamiento, traslado y visualización son un desafío por la gran cantidad de éstas que existen en los centros de salud. Por esto es importante la definición de modelos de almacenamiento en bases de datos, tanto de las imágenes como de sus metadatos para una adecuada gestión de las mismas, principalmente en lo que refiere a recuperación e interoperabilidad. El estándar DICOM establece un mismo formato de imagen digital para todos los estudios en medicina, y a la vez, incluye información del paciente, del estudio realizado, equipos y mucha otra información relacionada con la imagen. La finalidad de este trabajo es la utilización de Ontologías para relacionar la información de todas las imágenes y recuperar las similares a través del uso de consultas semánticas sobre las Ontologías creadas con las imágenes DICOM.

**Palabras clave:** Base de datos OR, DICOM, Ontologías.

## Contexto

Este trabajo consiste en el desarrollo de una tesis de maestría, que 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 digitales en medicina están reemplazando cada día más a los antiguos formatos físicos que ocupaban inmensos espacios destinados a almacenar placas, hojas de papel, y otros medios. Una vez almacenadas en depósitos no se volvían a consultar, ocasionando gastos y ocupando lugar en las instituciones médicas. Las imágenes digitales proporcionan diferentes ventajas sobre los antiguos formatos físicos desde el momento de la adquisición hasta el de su visualización. Entre ellas:

- 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.

Una imagen médica no tiene sentido sola, sino que es importante la información del contexto en el que se ha tomado la misma. En este sentido, el estándar DICOM (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. Además, 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.). 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.

Actualmente existen sistemas que realizan la gestión de imágenes, conocidos como PACS (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. En los últimos años se ha producido una generalización a nivel mundial del uso de este protocolo.

Una herramienta útil para los profesionales en medicina al momento de efectuar el diagnóstico de un paciente, es lograr la recuperación de imágenes similares a la imagen obtenida en un estudio efectuado a dicho paciente. Para lo cual, además de contar con las imágenes en formato DICOM, también es necesario establecer relaciones entre las mismas de manera que, a partir de ciertas reglas o datos, poder recuperar otras imágenes y diagnósticos similares que se encuentren almacenados en la base de datos. La capacidad de las ontologías [4] de inferir nuevo conocimiento a partir de información explícita, hacen de las mismas una ayuda importante para la realización de estas búsquedas por similitud.

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://biportal.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**

La principal línea de investigación que se está trabajando es la gestión de imágenes en bases de datos OR (Objeto Relacionales). En este trabajo se trata la creación 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. El uso de la tecnología OR se debe a que los modelos para la recuperación de imágenes, 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.

El modelo OR [2] permite dar soluciones a las limitaciones de las bases de datos relacionales. 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 paquetes para la gestión de imágenes DICOM [5] [6] y el tratamiento de datos semánticos [7].

Las facilidades mencionadas, posibilitan la creación de infraestructuras que extienden los servicios del sistema de gestión de base de datos [8].

Otra línea de investigación que abarca este trabajo, está 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 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.

El relevamiento, 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.

Una vez seleccionada las ontologías se deberá definir la manera en que se vincularan las mismas. 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 se han alcanzado de manera parcial como se indica a continuación.

### Objetivos alcanzados

- *Crear modelos en una base de datos Objeto-Relacional que soporten el estándar DICOM.* Se diseñaron y construyeron tablas, objetos y ontologías de la base de datos que

serán los encargados de almacenar y permitir la recuperación de las imágenes médicas por contenido semántico.

- *Crear ontologías en la base de datos para la búsqueda semántica de imágenes y diagnósticos.* Durante el proceso de investigación y diseño de una ontología para este trabajo se basó principalmente en la línea definida por el estándar DICOM, define como principales entidades a *Estudio, Serie e Imagen* (Fig. 1). Además se utilizaron ontologías ya existentes desarrolladas por otros investigadores y organizaciones, y de esta manera poder tomar ventajas de las mismas e integrarlas a lo desarrollado en esta investigación. Las ontologías utilizadas como complemento de lo desarrollado en este trabajo son las ontologías de Síntomas y Enfermedades Humanas. Las mismas las podemos encontrar en los sitios Berkeley.org<sup>1</sup> y BioPortal<sup>2</sup>.

### Objetivos pendientes

- *Crear un software para la inserción y recuperación de imágenes DICOM.* Este objetivo se logro de manera parcial. Hasta aquí se han implementado los modelos definidos en un sistema de gestión de base de datos objeto-relacional. Queda pendiente, finalizar con el desarrollo de una aplicación que permitirá almacenar imágenes DICOM y sus metadatos en el modelo semántico en la base de datos, y así, poder recuperar tanto imágenes como información con contenido semántico. Esta aplicación, permitirá visualizar las imágenes

---

<sup>1</sup> Berkeley Bioinformatics Open-source Projects  
<http://www.berkeleybop.org/>

<sup>2</sup> Repositorio de ontologías bio-médicas:  
<http://bioportal.bioontology.org/>

almacenadas en la base de datos, especificar los criterios de búsquedas y mostrar los resultados.

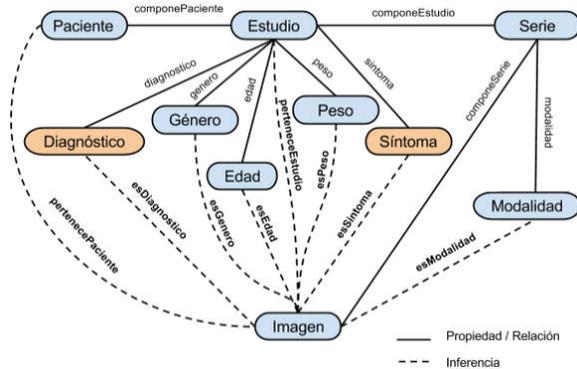


Fig. 1: Ontología en base al estándar DICOM.

## Formación de Recursos Humanos

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 maestrando, junto con el Director y Co-Director de la misma. El tesista 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). El estado de la tesis se encuentra en un alto porcentaje de avance, y los resultados finales, serán presentados en congresos nacionales o internacionales relacionados con el área Sistemas de Información, así como también, 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, 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] Mabotuwana, T., Lee, M.C., Cohen-Solal, E.V. (2013). An ontology-based similarity measure for biomedical data - Application to radiology reports. *Journal of Biomedical Informatics*, 46, 857-868.
- [5] Sue Pelski,. Oracle Multimedia DICOM Developer's Guide, 12c Release 1 (12.1) E17698-09. July 2014.
- [6] Alvez C. y Vecchietti A. “Representación y Recuperación de Imágenes Médicas en Bases de Datos Objeto-Relacionales”. Jornadas Argentinas de Informática e Investigaciones Operativas: 36ª 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. Vecchietti. 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.