

Un Repositorio RDF para la Integración y Consulta de Datos de Pacientes Hepáticos*

María del Mar Roldan-García and Jose F. Aldana-Montes

Departamento de Lenguajes y Ciencias de la Computación
Universidad de Málaga, España
{mmar, jfam}@lcc.uma.es

Abstract. Los casos médicos pasados, y por lo tanto, la experiencia clínica, son recursos de valor incalculable para apoyar la práctica clínica, la investigación y la formación. Los profesionales médicos deben poder intercambiar información sobre casos médicos de pacientes y explorarlos desde distintas perspectivas subjetivas. Esto requiere de una metodología sistemática y flexible para la representación de los casos médicos que soporte el intercambio de información procesable del paciente. En este artículo presentamos un enfoque basado en ontologías para modelar casos médicos de pacientes que utiliza pacientes con enfermedades hepáticas como ejemplo. Para este fin, se propone una nueva ontología, LiCO, que utiliza estándares médicos bien conocidos para representar casos de pacientes con enfermedades hepáticas. La utilidad del enfoque propuesto se demuestra con consultas semánticas y razonamiento utilizando datos recopilados de pacientes reales. Los resultados preliminares son prometedores con respecto al potencial de la representación de casos médicos basada en ontologías para la construcción de sistemas de búsqueda y recuperación de información de casos médicos, allanando el camino hacia una plataforma de intercambio de experiencias clínicas para comparar diagnósticos, para investigación y para formación.

Keywords: Ontologías, Integración de datos, RDF y Representación de casos médicos

1 Introducción

Históricamente, la disciplina de la medicina se ha desarrollado sobre la experiencia colectiva reunida durante siglos, incluso durante milenios. Esta experiencia ha pasado a través de generaciones en una relación de maestro-aprendiz y más tarde a través de relatos escritos de casos anteriores, hasta que la medicina moderna se esforzó por explicar los resultados clínicos por medio de relaciones causales. A pesar del gran avance que esto supuso, el estudio de casos anteriores no ha perdido su importancia, ya que siguen quedando muchas cosas que la

* Este trabajo ha sido parcialmente financiado por los Proyectos TIN2014-58304 y TIN2017-86049-R (Ministerio de Economía, Industria y Competividad), y P12-TIC-1519 (Plan Andaluz de Investigación, Desarrollo e Innovación) y ha sido realizado en el contexto de la acción COST Keystone 1302.

medicina moderna aún no ha conseguido explicar. Para capturar y transmitir dicho conocimiento, es necesario un ecosistema eficiente para compartir la experiencia clínica (*Clinical Experience Sharing* (CES) en inglés), que permita una representación eficiente y que ofrezca una herramienta de intercambio de casos médicos [2]. Un ecosistema CES se compone esencialmente de tres componentes principales, aplicaciones, representación semántica y repositorio. El componente más crítico en dicho ecosistema es la *representación semántica*, que es la que hace de puente semántico entre los datos y los usuarios. La representación semántica debe capturar el significado de los datos en un formato legible por una máquina para que las computadoras puedan procesar/analizar los datos de acuerdo con las necesidades del usuario. Cuando hablamos de una aplicación CES para la búsqueda/recuperación de casos médicos similares, las solicitudes de los usuarios pueden ser subjetivas. Por ejemplo, para ayudar con las decisiones sobre un tratamiento, síntomas similares con diferentes tratamientos pueden ser de interés. Alternativamente, para ayudar a la formación médica, pueden necesitarse casos con diagnósticos similares pero con síntomas diferentes. Por otro lado, en el contexto de una aplicación CES de informes médicos estructurados, el usuario tiene que ser guiado para rellenar el informe, de manera que la aplicación se adapte dinámicamente en base a las observaciones/anotaciones proporcionadas por el usuario. La representación semántica que está en el centro del ecosistema CES debe ser un modelo de dominio centrado en el caso médico y basado en ontologías y vocabularios médicos. Su objetivo es representar un modelo integrado y completo de casos médicos (por ejemplo, pacientes).

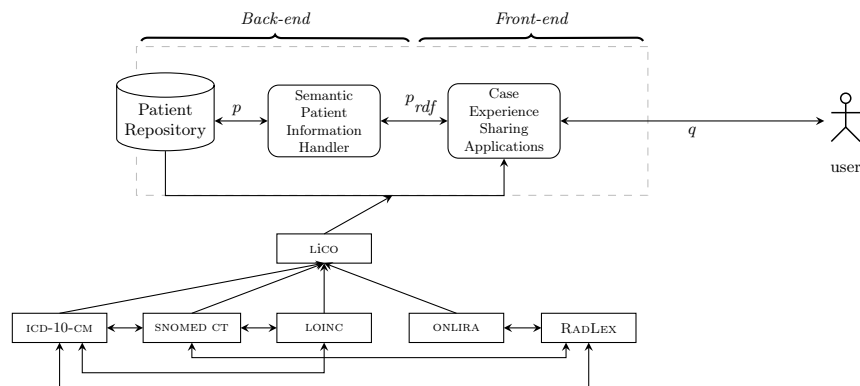


Fig. 1. Principales componentes de un ecosistema de Intercambio de experiencias clínicas (CES). LiCO es una ontología para representar casos de pacientes con referencias a vocabularios médicos bien conocidos.

En este estudio, presentamos una ontología centrada en casos médicos de pacientes hepáticos llamada Lico (*Liver Case Ontology*) que integra varias ontologías y vocabularios desarrollados para subdisciplinas de la medicina. Hasta

donde llega nuestro conocimiento, no existe ninguna ontología similar en la literatura, a pesar de que se han propuesto numerosas ontologías en el campo de la medicina. La figura 1 representa una visión general del enfoque propuesto. En nuestra propuesta, la ontología Lico utiliza varios vocabularios y ontologías médicas que especifican términos y conceptos relacionados con el registro de observaciones médicas. Las principales contribuciones de este artículo son: (1) Lico, una ontología para casos médicos de pacientes hepáticos que utiliza ontologías/vocabularios médicos bien conocidos¹, (2) la validación de Lico en términos de representación y procesamiento de datos de pacientes reales, mediante un repositorio de casos reales de pacientes hepáticos (ejemplos de representaciones RDF disponibles en el sitio web.), y (3) presentación de una prueba de concepto para búsquedas/recuperación CES utilizando un razonador semántico sobre el repositorio de datos de pacientes reales.

2 Modelo Semántico

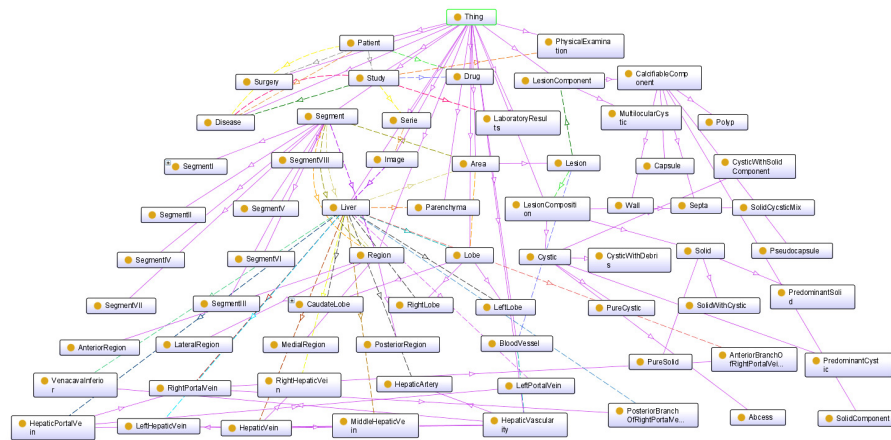


Fig. 2. Visión general de la ontología LICO. Las flechas sólidas representan la relación de subclase mientras que las flechas punteadas indican propiedades específicas.

Uno de los objetivos principales de este trabajo es desarrollar un modelo semántico para la representación de casos médicos de pacientes hepáticos. La descripción de dichos casos médicos ha sido realizada por radiólogos de la Facultad de Medicina de la Universidad de Estambul, Departamento de Radiodiagnóstico. Para formalizar dichos casos médicos, se ha desarrollado una ontología (Lico), que describe las características principales del caso médico del paciente

¹ LICO está disponible en <http://vavlab.ee.boun.edu.tr/pages.php?p=research/CARERA/preDownloadForm.php>

hepático. La ontología está desarrollada en OWL (W3C Web Ontology Language) siguiendo el proceso estándar de desarrollo Ontology 101 de siete pasos [7]:

1. Determinar el dominio y el alcance de la ontología. El alcance de nuestra ontología es el caso médico de los pacientes hepáticos utilizado en la Facultad de Medicina de la Universidad de Estambul. La ontología describirá la información del paciente hepático, incluidos datos demográficos, historia clínica, resultados de pruebas de laboratorio, uso de medicamentos y hallazgos radiológicos.

2. Considerar la reutilización de ontologías ya existentes. Para describir la información radiológica, se utiliza una versión revisada de la ontología ONLIRA [4]. ONLIRA fue desarrollada por investigadores de la Universidad Bogazici de Estambul y posteriormente fue modificada y mejorada por los autores de este trabajo en colaboración con dichos investigadores. Las clases de ONLIRA se mapearon con RADLEX [5]. RADLEX es una terminología controlada desarrollada por la *Radiological Society of North America* (RSNA) con el propósito de proporcionar una fuente unificada de términos de radiología. Para las enfermedades de los pacientes se usan las categorías de ICD-10-CM [8]. La ICD-10-CM está publicada por la Organización Mundial de la Salud (OMS), e incluye aproximadamente 15,000 códigos para enfermedades, signos y síntomas, hallazgos anormales, quejas, circunstancias sociales, y causas externas de lesiones o enfermedades. Los resultados de laboratorio están representados mediante códigos LOINC [3]. LOINC es un sistema de codificación para documentar observaciones clínicas y de laboratorio. Finalmente, siempre que es posible mapeamos las clases de la ontología con códigos de SNOMED CT [1], por ejemplo, las clases que describen los resultados del examen físico del paciente. SNOMED CT es la terminología clínica integral, multilingüe y codificada de mayor amplitud, precisión e importancia desarrollada en el mundo y pretende ser el estándar utilizado en los historiales clínicos electrónicos (EHR). El resto de información relativa a los pacientes y a sus estudios médicos se ha definido en base a la descripción del caso médico y se ha optado por no reutilizar ninguna otra terminología médica, ya que ninguna se adapta completamente a los datos descritos.

3. Enumerar los términos importantes en la ontología. Los conceptos importantes se identificaron durante la fase de definición de nuestra representación de los casos de pacientes hepáticos. En esta fase se realizaron consultas a expertos en el dominio y búsqueda de literatura. Los términos clave incluyen paciente, enfermedad, cirugía, medicamento, estudio, examen físico, resultados de laboratorio, diagnósticos, hígado, lóbulo, segmento, lesión y vascularidad.

4. Definir las clases y la jerarquía de clases. Las clases corresponden a los términos identificados para casos médicos de pacientes hepáticos. La figura 2 muestra las clases principales y sus relaciones con otras clases. Por ejemplo, la clase `onlira:HepaticVascularity` tiene varias subclases, incluidas `onlira:HepaticArtery`, `onlira:HepaticPortalVein`, `onlira:HepaticVein` y `onlira:VenacavaInferior` y la clase `onlira:Lobe` tiene las subclases `onlira:CaudateLobe`, `onlira:LeftLobe` y `onlira:RightLobe`.

5. Definir las relaciones entre clases y sus atributos. Las clases tienen relaciones con otras clases y también atributos. Al igual que las clases, las rela-

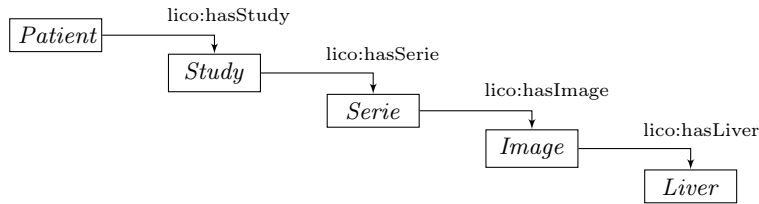


Fig. 3. Cuando un paciente visita a un médico se crea un estudio. Una exploración relacionada con este estudio da como resultado una serie de imágenes, de la cual se selecciona una para hacer las observaciones sobre el hígado.

ciones se identificaron en función de nuestra representación de los casos médicos del paciente hepático. Algunos ejemplos de propiedades de objetos son: `lico:hasStudy` con dominio `lico:Patient` y rango `lico:Study`, `lico:hasGeneticDisease` con dominio `lico:Patient` y rango `lico:Disease`, `lico:hasLiver` con dominio `lico:Image` y rango `onlira:Liver`, y `onlira:hasLobe` con dominio `onlira:Liver` y rango `onlira:Lobe`. Entre los ejemplos de propiedades del tipo de datos se incluyen: `lico:hasAge` con dominio `lico:Patient` y rango `xsd:nonNegativeInteger` y `lico:hasAST` con dominio `lico:LaboratoryResults` y rango `xsd:double`. Además, `lico:hasAST` se define como equivalente a `loinc:1920-8`.

6. Definir las propiedades de los atributos. Este paso incluye la definición de las restricciones de cardinalidad y de valor (*value restrictions*). Las restricciones de valor se usan en nuestra ontología para restringir los segmentos hepáticos en cada lóbulo hepático o región hepática. Por ejemplo, el lóbulo caudado está segmentado solo por el segmento I, y el segmento II está ubicado solo en la región lateral. Las restricciones de cardinalidad se usan para especificar el número de segmentos en cada lóbulo. Por ejemplo, el lóbulo caudado, el lóbulo izquierdo y el lóbulo derecho están segmentados por uno, tres y cuatro segmentos, respectivamente.

7. Crear instancias. Las instancias (individuos en OWL) corresponden a los datos específicos obtenidos de un paciente específico. Los datos se obtienen de los pacientes de la Facultad de Medicina de la Universidad de Estambul, Departamento de Radiodiagnóstico. Los médicos insertan manualmente los datos utilizando formularios web generados en base a las clases y propiedades de la ontología. Posteriormente, los datos recopilados se exportan a RDF.

A lo largo de este artículo, se usan varias figuras para describir nuestro enfoque y distintos aspectos de la ontología. En estas figuras, las clases se representan con rectángulos, los datos con paralelogramos, los individuos con elipses y las propiedades con flechas. Los prefijos *onlira* y *lico* se usan para referirse a las ontologías ONLIRA y LICO respectivamente. Como se mencionó anteriormente, ONLIRA es una ontología existente desarrollada en trabajos anteriores que ha sido mejorada y extendida por los autores de este trabajo en colaboración con sus creadores, mientras que LICO se define en este documento como una nueva extensión de ONLIRA que incluye no solo la descripción de las imágenes radioló-

gicas de los pacientes, sino del caso médico completo. Debido a limitaciones de espacio, en algunas figuras se usan los prefijos o y l para referirse a onlira y lico, respectivamente.

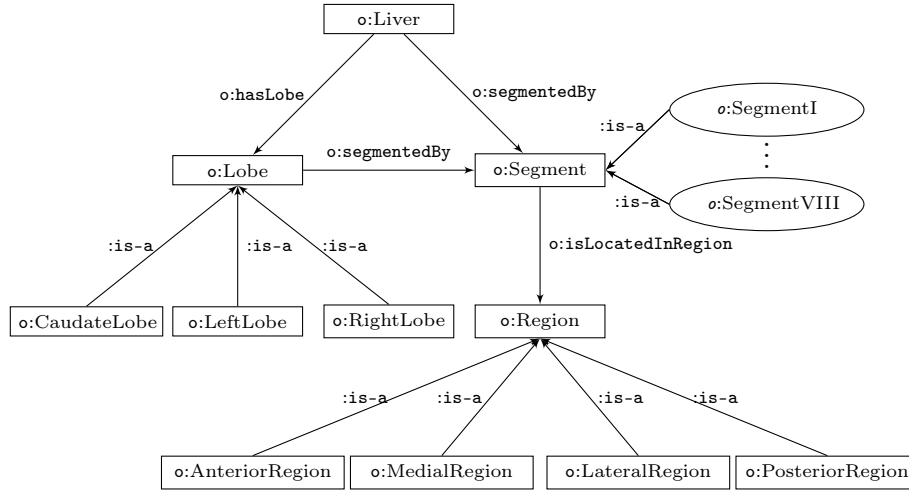


Fig. 4. Varias definiciones utilizadas para referirse a partes de un hígado

2.1 Ontología

Los casos médicos de pacientes hepáticos se modelan mediante los múltiples conceptos interrelacionados definidos en la ontología Lico, como por ejemplo, los que definen los datos demográficos, el historial médico, los resultados de pruebas de laboratorio, el uso de medicamentos y los hallazgos radiológicos. La ontología propuesta, resultante del proceso de desarrollo descrito anteriormente, tiene un total de 93 clases, 36 propiedades de objetos (relaciones binarias entre individuos), 119 propiedades de datos (atributos individuales) y 474 axiomas lógicos. En este artículo, se describe un subconjunto de las clases principales de la ontología junto con sus propiedades. Nos vamos a centrar en un paciente hepático, por lo que el caso médico debe incluir la descripción del paciente, así como todas las investigaciones médicas relacionadas con los distintos estudios realizados al paciente, y sus distintas series de imágenes radiológicas. La figura 3 muestra las relaciones de alto nivel entre las clases de la ontología, que comienzan en el paciente y van hasta las imágenes donde se realizan las observaciones del hígado. Cada vez que un paciente visita a un médico con una nueva incidencia, se crea un estudio para ese tema. Un estudio a menudo implica la realización de imágenes médicas (por ejemplo, tomografía computarizada y resonancia magnética), que consisten en una serie de imágenes relacionadas con la región del

cuerpo relevante para el estudio. Se selecciona una imagen apropiada de la serie para documentar las observaciones sobre las imágenes relacionadas con los órganos y las anomalías (lo que llamamos lesiones). Las clases principales que se definen en lico son: *Patient*, *Study*, *Serie*, *Image*, *Liver*, *Area* y *Lesion*.

Table 1. lico:Patient: Propiedades de objeto y de tipo de datos

Propiedades de Objeto	Lógica de descripciones
lico:hasGeneticDiseases	$\top \sqsubseteq \forall \text{ lico:hasGeneticDisease.lico:Disease}$
lico:hasNonLiverDisease	$\top \sqsubseteq \forall \text{ lico:hasNonLiverDisease.lico:Disease}$
lico:hasRegularDrug	$\top \sqsubseteq \forall \text{ lico:hasRegularDrug.lico:Drug}$
lico:hasSurgery	$\top \sqsubseteq \forall \text{ lico:hasSurgery.lico:Surgery}$
lico:hasStudy	$\top \sqsubseteq \forall \text{ lico:hasStudy.lico:Study}$
Propiedades de tipo de datos	Lógica de descripciones
lico:hasAge	$\top \sqsubseteq \forall \text{ lico:hasAge.Datatype nonNegativeInteger}$
lico:hasGender	$\top \sqsubseteq \forall \text{ hasGender.}\{ "F"^\wedge\text{string}, "M"^\wedge\text{string}\}$
lico:hasName	$\top \sqsubseteq \forall \text{ lico:hasName.Datatype string}$
lico:hasPatientID	$\top \sqsubseteq \forall \text{ lico:hasPatientID.Datatype string}$

Patient es la clase principal de lico. Como se muestra en la tabla 1, un paciente tiene una edad (lico:hasAge), un género, un nombre y un ID. Características importantes de los pacientes son sus enfermedades genéticas (lico:hasGeneticDisease) y sus enfermedades no hepáticas (lico:hasGeneticDisease) representadas mediante códigos ICD-10-CM, los medicamentos que toma el paciente de forma regular y sus cirugías previas.

Table 2. lico:Study: Propiedades de objeto y de tipo de datos

Propiedades de Objeto	Lógica de descripciones
lico:hasLiverPreDiagnosis	$\top \sqsubseteq \forall \text{ lico:hasLiverPreDiagnosis.lico:Disease}$
lico:hasNonRegularDrugs	$\top \sqsubseteq \forall \text{ lico:hasNonRegularDrugs.lico:Drug}$
lico:hasLaboratoryResults	$\top \sqsubseteq \forall \text{ lico:hasLaboratoryResults.lico:LaboratoryResults}$
lico:hasPhysicalExamination	$\top \sqsubseteq \forall \text{ lico:hasPhysicalExamination.lico:PhysicalExamination}$
lico:hasFinalDiagnosis	$\top \sqsubseteq \forall \text{ lico:hasFinalDiagnosis.lico:Disease}$
lico:hasSerie	$\top \sqsubseteq \forall \text{ lico:hasSerie.lico:Serie}$
Propiedades de tipo de datos	Lógica de descripciones
lico:hasComplaints	$\top \sqsubseteq \forall \text{ lico:hasComplaints.Datatype string}$
lico:hasDate	$\top \sqsubseteq \forall \text{ lico:hasDate.Datatype date}$
lico:hasStudyID	$\top \sqsubseteq \forall \text{ lico:hasStudyID.Datatype string}$

Study. Cada paciente tiene al menos un estudio. Un estudio tiene una fecha, un ID y posibles incidencias. Además, tiene un prediagnóstico, los medicamentos no regulares que se consumen, un diagnóstico final (que también es un código ICD-10-CM), resultados de laboratorio y exámenes físicos. La tabla 2 muestra las propiedades de objeto y de tipo de datos para la clase lico:Study. Las propiedades

de tipo de datos de la clase `lico:PhysicalExamination` se mapean con códigos SNOMED CT. Las propiedades de tipo de datos de la clase `lico:LaboratoryResults` se mapean con códigos LOINC. Finalmente, un estudio tiene series, cada una de las cuales consiste en un conjunto de imágenes radiológicas. Una imagen incluye, entre otras cosas, el hígado del paciente.

Table 3. `onlira:Lesion`: Propiedades de objeto y de tipo de datos

Propiedades de Objeto	Lógica de descripciones
<code>onlira:hasLesionComponent</code>	$\top \sqsubseteq \forall \text{onlira:hasLesionComponent.onlira:LesionComponent}$
<code>onlira:hasLesionComposition</code>	$\top \sqsubseteq \forall \text{onlira:hasLesionComposition.onlira:LesionComposition}$
<code>lico:isCloseToBloodVessel</code>	$\top \sqsubseteq \forall \text{lico:isCloseToBloodVessel.lico:BloodVessel}$
Propiedades de tipo de datos	Lógica de descripciones
<code>onlira:hasDebrisLocation</code>	$\top \sqsubseteq \forall \text{onlira:hasDebrisLocation.}\{\text{"floating inside"}^{\wedge}\text{string, "located on dependent position"}^{\wedge}\text{string, "other"}^{\wedge}\text{string}\}$
<code>onlira:hasLesionContrastUptake</code>	$\top \sqsubseteq \forall \text{onlira:hasLesionContrastUptake.}\{\text{"dense"}^{\wedge}\text{string, "heterogeneous"}^{\wedge}\text{string, "homogeneous"}^{\wedge}\text{string, "minimal"}^{\wedge}\text{string, "moderate"}^{\wedge}\text{string, "other"}^{\wedge}\text{string}\}$
<code>onlira:hasLesionQuantity</code>	$\top \sqsubseteq \forall \text{onlira:hasLesionQuantity.}\{\text{"1"}^{\wedge}\text{integer, "2"}^{\wedge}\text{integer, "3"}^{\wedge}\text{integer, "4"}^{\wedge}\text{integer, "5"}^{\wedge}\text{integer, "multiple"}^{\wedge}\text{string}\}$
<code>onlira:isDebrisObserved</code>	$\top \sqsubseteq \forall \text{onlira:isDebrisObserved}$ Datatype boolean
<code>onlira:isLevelingObserved</code>	$\top \sqsubseteq \forall \text{onlira:isLevelingObserved}$ Datatype boolean

Liver. El hígado es un concepto clave en `Lico` ya que nos centramos en pacientes con enfermedades hepáticas. Existen numerosas propiedades para describir el hígado, que se importan de la ontología `ONLIRA`. El hígado se describe con un tamaño, una densidad y un contorno. Por ejemplo, la dimensión *cranio-caudal* del hígado se define con la propiedad `onlira:hasCranioCaudalDimension` cuyo rango es un entero que corresponde al tamaño en milímetros. Además, el cambio en el tamaño del hígado durante un periodo de tiempo es una característica médicamente significativa, y se define con la propiedad `onlira:hasSizeChange`, que puede asociarse con valores como *increased*. La clase hígado se relaciona con una clase `onlira:Parenchyma` mediante la propiedad `onlira:hasParenchyma`. Los lóbulos del hígado se representan con las clases: `onlira:CaudateLobe`, `onlira:RightLobe` y `onlira:LeftLobe`. La propiedad `onlira:hasLobe` relaciona un hígado con sus lóbulos. Los lóbulos y los hígados se segmentan por segmentos y los segmentos se ubican en regiones (ver Figura 4). Finalmente, algunas propiedades de objeto describen la vascularidad de un hígado. Todas las propiedades de tipo de datos de `onlira:Liver` y la propiedad `onlira:hasParenchyma` se definen como *propiedades funcionales* para garantizar que solo se aserte un valor de propiedad para un hígado dado.

Area y Lesion. Para describir un área anormal observada en una imagen radiológica del hígado se usan distintas propiedades. El concepto general se representa con la clase `onlira:Area`. La propiedad `onlira:hasAreaShape` describe la forma de un área, que puede tomar los valores: *band*, *fusiform*, *linear*, *nodular*,

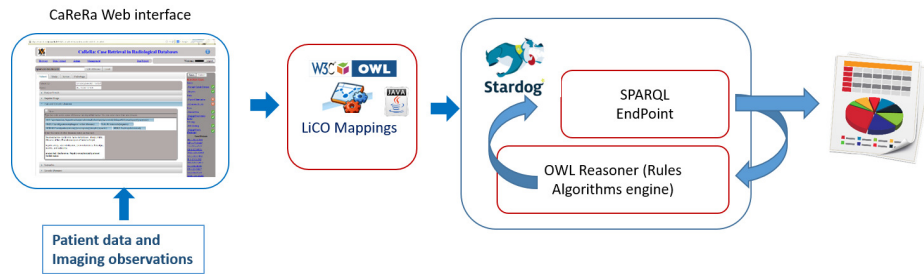


Fig. 5. Descripción general del enfoque semántico para consultar casos médicos de pacientes hepáticos.

ovoid, *serpiginious*, and *other*. La propiedad `onlira:hasAreaDensity` describe la densidad del área que puede tomar los valores *hyperdense*, *hypodense*, or *isodense*. La propiedad `onlira:isCalcified` indica si un área está calcificada y si lo está, entonces la propiedad `onlira:hasCalcification` especifica el tipo de calcificación que puede ser *coarse*, *focal*, *millimetric*, *punctate*, or *scattered*. La ubicación de un área se especifica con las propiedades `onlira:isLocatedInSegment` y `onlira:isLocatedInLobe`. Un tipo de área particularmente importante es una lesión. Una lesión es un daño patológico a un órgano o tejido debido a una lesión o enfermedad. Se puede manifestar como un absceso, un tumor, una úlcera o una herida. La clase `onlira:Lesion` es una subclase de `onlira:Area`. Las lesiones tienen propiedades adicionales a las definidas para el área (ver tabla 3). Por ejemplo, las lesiones pueden tener componentes (`onlira:hasComponent`) y composición (`onlira:hasComposition`).

3 Caso de Uso

Para validar el enfoque propuesto, se han usado 46 casos reales de pacientes hepáticos que se han representado usando LICO, y que se han usado para realizar consultas y examinar la utilidad de la representación propuesta para obtener información útil. Los datos de los pacientes se adquirieron durante el proyecto CaReRa de una manera heterogénea, mientras que algunos datos críticos se recopilaban en base a la ontología ONLIRA, algunos otros datos se recogieron con métodos más estándares. Todos los datos fueron recolectados en base a estándares médicos aceptados. La tarea inicial fue crear casos médicos descritos según LICO a partir de los datos recopilados de los pacientes. Estos casos se exportaron a RDF con funciones que traducen todos los datos del paciente a las tripletas RDF correspondientes. Más específicamente, los datos se serializan en formato RDF con funciones que usan la API RDF de Apache Jena [6], que permite manipular gráficos RDF. Estas funciones crean instancias de las clases que describen la información del paciente y de las relaciones entre estas instancias a través de las propiedades de objeto y las propiedades de tipo de datos de LICO. Las representaciones RDF de los casos médicos de pacientes hepáticos se almacenan

usando el repositorio RDF y razonador *Stardog*², que proporciona un endpoint SPARQL para ejecutar consultas. La figura 5 ilustra la adquisición de datos, su transformación a RDF y la creación de endpoint SPARQL para permitir la consulta de los casos médicos.

3.1 Datos de Pacientes hepáticos

Los datos de los pacientes se obtuvieron de registros de pacientes existentes en la Facultad de Medicina de la Universidad de Estambul, Departamento de Radiodiagnóstico, con la aprobación de la Facultad de Medicina de la Universidad de Estambul, Comité de Evaluación de ética³. **CaReRa-Web**⁴ es una herramienta de recopilación y recuperación de datos dirigida por la ontología para pacientes hepáticos. Exporta la información del paciente expresada con Lico como tripletas RDF que se utilizan para realizar consultas y razonamientos semánticos. Los datos de las tomografías computarizadas (*CT scan*) se anonimizaron en el sitio web del hospital. El personal médico autorizado solo accedió a los registros impresos del paciente y a la información de diagnóstico (excluyendo todos los datos demográficos, excepto el sexo y la edad) que ingresaron manualmente en una base de datos segura a través de una herramienta web interna. Los datos son tanto de pacientes sanos como de personas con enfermedades hepáticas. Los datos de los paciente consisten en datos demográficos, de laboratorio, de exploración física y de observación de imágenes radiológicas todos ellos registrados según estándares bien conocidos. Las observaciones de imágenes radiológicas incluyen información general sobre las tomografías computarizadas realizadas al paciente (por ejemplo el *Slice Thickness* es de 1.0mm y el *KVP/Current* es de 120/300v/ma). Las observaciones de las imágenes describen las características del hígado (por ejemplo, *Density Type: heterogeneous* y *Right Portal Vein Lumen Diameter: reduced*). Los detalles de la patología se especifican para una imagen seleccionada de la serie. Dada una imagen, se especifica una región de interés y se documentan las observaciones relacionadas con ella.

Los individuos RDF para modelar la anatomía hepática se crean con Jena. Por lo tanto, se generan individuos RDF para el hígado, así como para la vascularización hepática, los segmentos, los lóbulos, las regiones, etc. de cada paciente. Estos individuos están relacionados con el hígado del paciente por medio de las propiedades de objeto de LICO como *onlira:hasPortalVein*, *onlira:isSegmentedBy*, etc. Por ejemplo, para el paciente *Capa – 00029*, se crea el *liver579*. Las tripletas RDF para relacionar el *liver579* con su anatomía son entre otras:

```
o:liver579 o:isSegmentedBy o:SegmentII.  
o:liver579 o:hasPortalVein o:liver579PortalVein.
```

² <http://www.stardog.com/>

³ Ethics committee approval #: 09/06/2010 - 01

⁴ Se puede acceder a las capturas de pantalla de **CaReRa-Web** y a una versión demo en: <http://vavlab.ee.boun.edu.tr/pages.php?p=research/CARERA/crrweb/careraweb.html>



Fig. 6. Parte de la información del caso médico para el paciente con id: *Capa-00029* representada con un grafo RDF generado a partir de los datos del paciente recopilados con la herramienta *CaReRa-Web*. Las elipses denotan individuos y los paralelogramos indican valores de datos.

La figura 6 muestra parte de los datos de un paciente de forma gráfica (ID del paciente: *Capa* – 00029) que está relacionado con las observaciones de imágenes radiológicas de una lesión en el hígado. Las elipses apiladas indican que un paciente puede tener múltiples estudios (correspondientes a cada investigación) y cada estudio puede tener varias series (para cada exploración). Para permitir el razonamiento con los datos de diferentes pacientes y para evitar inconsistencias, necesitamos generar diferentes individuos para diferentes partes de la anatomía del hígado en diferentes hígados. Posteriormente se crean individuos RDF para modelar la relación entre el paciente y el hígado que se muestra en la figura 3, es decir, se generan las tripletas RDF que enlazan *Capa* – 00029 con *liver579* para el ejemplo del paciente mostrado en la figura 6. También se generan las tripletas RDF para conectar los estudios del paciente con los resultados de laboratorio correspondientes, los exámenes físicos y el diagnóstico final. Finalmente, el hígado de cada paciente está vinculado a sus lesiones y las correspondientes observaciones de las imágenes se le asignan por medio de las correspondientes propiedades de tipo de datos de Lico.

Algunos ejemplos de tripletas RDF para *Capa* – 00029 en la figura 6 son:

```

l:Study521 l:hasPhysicalExamination l:Study521PhysicalExam.
l:Study521PhysicalExam l:hasPulse 78.
o:liver579 o:hasArea o:lesion117.
o:lesion117 o:hasAreaDensity "hypodense".

```

Consulta 1:

```

select ?patient ?study ?lesion ?vessel ?ldh ?pulse
where {
  ?patient l:hasStudy ?study.
  ?study l:hasSerie ?serie;
         l:hasLaboratoryResults ?lb;
         l:hasPhysicalExamination ?fe;
  ?serie l:hasImage ?img.
  ?img l:hasLiver ?liver.
  ?liver o:hasArea ?lesion.
  ?lesion l:isCloseToBloodVessel ?bv.
  ?bv l:hasBloodVesselName ?vessel.
  ?lb l:hasLDH ?ldh.
  ?fe l:hasPulse ?pulse }

```

Resultado:

patient	study	lesion	vessel	ldh	pulse
l:Capa-00342	l:Study780	o:Lesion212	o:Liver849VenacavaInferior	3.07E2	84
l:Capa-00342	l:Study780	o:Lesion212	o:Liver849RightPortalVein	3.07E2	84
l:Capa-00050	l:Study522	o:Lesion122	o:Liver589HepaticVein	4.66E2	78
l:Capa-00024	l:Study509	o:Lesion115	o:Liver576RightHepaticVein	3.85E2	88
l:Capa-00024	l:Study509	o:Lesion115	o:Liver576MiddleHepaticVein	3.85E2	88
l:Capa-00024	l:Study509	o:Lesion115	o:Liver576RightPortalVein	3.85E2	88
l:Capa-00352	l:Study790	o:Lesion216	o:Liver859LeftPortalVein	2.95E2	70
l:Capa-00352	l:Study790	o:Lesion216	o:Liver859RightPortalVein	2.95E2	70
l:Capa-00021	l:Study507	o:Lesion135	o:Liver574PosteriorBranchOfRightPo...	3.2E2	92
l:Capa-00241	l:Study679	o:Lesion178	o:Liver748RightPortalVein	2.73E2	80

Fig. 7. ID del estudio, lesión, vaso sanguíneo, ldh y pulso de pacientes que tienen una lesión cerca de un vaso sanguíneo.

3.2 Consultas Semánticas

En esta sección presentamos varias consultas SPARQL para demostrar la utilidad de nuestro enfoque. Dichas consultas han sido definidas por los autores de este trabajo en colaboración con los autores de la ontología ONLIRA. El objetivo de estas consultas es demostrar la utilidad de la ontología propuesta para consultar datos integrados de casos médicos de pacientes hepáticos así como del la explotación de las relaciones con otras terminologías médicas. Estas consultas se han evaluado utilizando el endpoint SPARQL del razonador *Stardog*. El repositorio RDF de stardog almacena los datos descritos en la Sección 3.1.

El rango de la propiedad `l:hasBloodVesselName` es la clase `onlira:HepaticVascularity` que tiene varias subclases, por ejemplo `onlira:RightPortalVein`, `onlira:VenacavaInferior`, etc. Según la semántica de OWL, los individuos de una subclase son siempre individuos de su clase padre. Es decir, los individuos de `onlira:RightPortalVein` son también individuos de `onlira:HepaticVascularity`. Por lo tanto, la consulta “*Obtener todas las lesiones cercanas a cualquier vaso sanguíneo*” devuelve las lesiones que

Consulta 2:

```
PREFIX radlex: <http://www.radlex.org/RID/#>
select ?patient ?study ?lesion ?ldh ?pulse
where{
  ?img    l:hasLiver ?liver .
  ?patient l:hasStudy ?study .
  ?study  l:hasSerie ?serie .
          l:hasLaboratoryResults ?lb;
          l:hasPhysicalExamination ?fe;
  ?serie  l:hasImage ?img.
  ?liver  o:hasArea ?lesion .
  ?lesion o:isLocatedInSegment ?sg.
  ?sg     rdf:type radlex:RID66.
  ?lb     l:hasLDH ?ldh.
  ?fe     l:hasPulse ?pulse }
```

Resultado:

patient	study	lesion	ldh	pulse
l:Capa-00152	l:Study599	o:Lesion149	4.93E2	88
l:Capa-00024	l:Study509	o:Lesion115	3.85E2	88
l:Capa-00021	l:Study507	o:Lesion135	3.2E2	92
l:Capa-00321	l:Study759	o:Lesion207	4.06E2	88
l:Capa-00230	l:Study669	o:Lesion174	3.17E2	70
l:Capa-00255	l:Study693	o:Lesion185	2.73E2	70
l:Capa-00273	l:Study711	o:Lesion194	2.99E2	70
l:Capa-00267	l:Study705	o:Lesion192	3.52E2	80
l:Capa-00055	l:Study525	o:Lesion125	9.4E2	80
l:Capa-00072	l:Study532	o:Lesion127	3.89E2	80
l:Capa-00354	l:Study792	o:Lesion217	3.58E2	88
l:Capa-00306	l:Study744	o:Lesion202	3.31E2	88

Fig. 8. ID del estudio, lesión, ldh y pulso de pacientes que tienen una lesión en el Segmento VI, usando el código RADLEX R66 (segment VI).

estén próximas a vasos sanguíneos sin tener que especificar el nombre de todos los posibles vasos sanguíneos, gracias al razonador semántico. La figura 7 presenta la Consulta 1, que busca pacientes que tienen lesiones cerca de un vaso sanguíneo junto con su diagnóstico final, la ID del estudio, el LDH ⁵ y el pulso, así como parte del resultado de esta consulta. Como puede observarse, la consulta devuelve los códigos ICD-10-CM de las enfermedades, por ejemplo *Q50.5* (Quiste hidatídico), *K75.9* (Hepatitis) y *C22.0* (Hepatocarcinoma). Podemos observar también que algunos pacientes tienen una lesión cerca de más de un vaso sanguíneo, como por ejemplo *l:capa-00024*, o más de una lesión como por ejemplo *l:capa-00342*. Por otro lado, algunos pacientes, como *l:capa-00050*, tienen más de un diagnóstico final. Con este tipo de consultas, un médico puede obtener información integrada del paciente, es decir, datos de la observación de imágenes, de los resultados de laboratorio y del examen físico que generalmente se almacenan en diferentes repositorios. Finalmente, con nuestro enfoque, los radiólogos o los médicos pueden usar los códigos RADLEX en las consultas si se sienten más cómodos. Por ejemplo, el código RID66 (SegmentVI) se puede emplear para encontrar el ID del estudio, LDH y pulso para pacientes con una lesión localizada en el segmento VI del hígado. Esto es posible porque el razonador semántico es capaz de realizar inferencias basadas en owl:equivalentClass. La consulta correspondiente (Consulta 2) se muestra en la Figura 8.

⁵ Lactato deshidrogenasa (LDH)

4 Conclusiones

En este trabajo se ha desarrollado una ontología, LICO para el dominio de pacientes hepáticos como una prueba de concepto para la representación y el razonamiento basados en casos médicos compuestos. LICO se alinea con varios vocabularios médicos muy conocidos (SNOMED CT, LOINC, RADLEX, e ICD-10-CM). Se demuestra con datos de pacientes reales que LICO representa con éxito los casos médicos de pacientes hepáticos y que el razonador *Stardog* puede realizar consultas semánticas no triviales sobre esta representación, explotando las relaciones con otros vocabularios. En conclusión, los resultados presentados son muy alentadores para la adopción de tecnologías de la Web Semánticas para modelar casos médicos. La investigación futura incluiría una validación de LICO más exhaustiva mediante la definición de consultas reales por parte del personal médico, la extensión de LICO a otros órganos (subdominios médicos) y el desarrollo de interfaces de usuario intuitivas que permitirían a los usuarios finales crear consultas subjetivas de casos médicos de una manera efectiva y eficiente.

References

1. SNOMED CT website. <http://www.snomed.org/snomed-ct/>. Accessed: 17-August-2017.
2. Neda Marvasti et. al. Clinical experience sharing by similar case retrieval. In *Proceedings of the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare*, MIIRH '13, pages 67–74, New York, NY, USA, 2013. ACM.
3. Stanley M Huff, Roberto A Rocha, Clement J McDonald, Georges JE De Moor, Tom Fiers, W Dean Bidgood Jr, Arden W Forrey, William G Francis, Wayne R Tracy, Dennis Leavelle, et al. Development of the logical observation identifier names and codes (loinc) vocabulary. *Journal of the American Medical Informatics Association*, 5(3):276–292, 1998.
4. Nadin Kokciyan, Rustu Turkay, Suzan Uskudarli, Pinar Yolum, Baris Bakir, and Burak Acar. Semantic description of liver ct images: An ontological approach. *IEEE journal of biomedical and health informatics*, 18(4):1363–1369, 2014.
5. Sanjoy Kundu, Maxim Itkin, Debra A. Gervais, Venkataramu N. Krishnamurthy, Michael J. Wallace, John F. Cardella, Daniel L. Rubin, and Curtis P. Langlotz. The {IR} radlex project: An interventional radiology lexicon - a collaborative project of the radiological society of north america and the society of interventional radiology. *Journal of Vascular and Interventional Radiology*, 20(7, Supplement):S275 – S277, 2009.
6. Brian McBride. Jena: Implementing the rdf model and syntax specification. In *Proceedings of the Second International Conference on Semantic Web-Volume 40*, pages 23–28. CEUR-WS. org, 2001.
7. Natalya F. Noy and Deborah L. McGuinness. Ontology development 101: A guide to creating your first ontology. Technical report, Stanford University Knowledge Systems Laboratory Technical Report KSL-01-05, 2001.
8. World Health Organization. Who | international classification of diseases. <http://www.who.int/classifications/icd/en/>. Accessed: 2017-08-05.