

Escenarios de vinculación de las bases de datos relacionales y las ontologías: un mapeo sistemático

(Linkage scenarios of relational databases and ontologies: a systematic mapping)

Yoan Antonio López Rodríguez¹, Yusniel Hidalgo Delgado², Nemury Silega Martínez³

Resumen

Las bases de datos relacionales constituyen una de las fuentes de datos más utilizadas. Sin embargo, como fuente de almacenamiento presenta un grupo de limitaciones. En las bases de datos relacionales se dificulta almacenar conocimiento semántico. Para solventar las carencias en la representación de conocimiento de las bases de datos relacionales, una tendencia ha sido el uso de las ontologías. Las ontologías poseen una mayor riqueza semántica y son más cercanas al vocabulario del usuario final que los esquemas de las bases de datos relacionales. El objetivo de la presente investigación consiste en realizar un mapeo sistemático acerca de los escenarios en que se vinculan las bases de datos relacionales y las ontologías para brindar una mejor integración, consulta y visualización de los datos almacenados. El mapeo se llevó a cabo aplicando una propuesta metodológica establecida en la literatura. Como resultado de la investigación, se detectó que el mapeo de bases de datos relacionales a ontologías y la utilización de las ontologías para la integración de fuentes de datos heterogéneas fueron los escenarios de mayor presencia. Igualmente, se identificaron tendencias y desafíos en cada uno de los escenarios que pueden merecer mayores esfuerzos de investigación en el futuro.

Palabras clave

Bases de datos relacionales, mapeo sistemático, ontologías.

Abstract

Relational databases are one of the most used data sources. However, as a storage source, they present a group of shortcomings. It is complex to store semantic knowledge in relational databases. To solve the deficiencies in knowledge representation of relational databases, one trend has been to use ontologies. Ontologies possess a richer semantic and are closer to the end user vocabulary than relational database schemas. The objective of the present research was to carry out a systematic mapping about the scenarios where relational databases and ontologies are linked to provide a better integration, query, and visualization of stored data. The mapping was carried out by applying a methodological proposal established in the literature. As outcomes of the research, it was detected that the mapping of relational databases to ontologies and the ontologies usage for the integration of heterogeneous data sources were the most common scenarios. Likewise, trends and challenges were identified in each scenario, which might deserve further research efforts in the future.

Keywords

Relational databases, ontologies, systematic mapping.

1 Universidad de las Ciencias Informáticas. La Habana, Cuba. [yalopez@uci.cu, <https://orcid.org/0000-0001-5615-375X>]
2 Universidad de las Ciencias Informáticas. La Habana, Cuba. [yhdelgado@uci.cu, <https://orcid.org/0000-0001-5763-0669>]
3 Universidad de las Ciencias Informáticas. La Habana, Cuba. [nsilega@uci.cu, <https://orcid.org/0000-0002-8436-5650>]

1. Introducción

Las bases de datos relacionales son una de las fuentes de datos más utilizadas. Su gran aceptación se debe entre, otras razones, a la robustez, flexibilidad, alto rendimiento, escalabilidad, seguridad y protección con que almacenan los datos (Sujatha y Raju, 2016). Sin embargo, como fuente de almacenamiento, también presenta un grupo de limitaciones. En las bases de datos relacionales se dificulta almacenar conocimiento semántico (Liu y Gao, 2018). Para operar sobre los datos se requiere aprender un lenguaje de consulta estructurado y, por tanto, el usuario común que solo conoce el lenguaje natural no puede acceder directamente a la información requerida (Sujatha y Raju, 2016). El formato de almacenamiento no es adecuado para ser procesado directamente por herramientas analíticas. En ambientes donde coexisten varias bases de datos se dificulta su integración y, con ello, el intercambio de los datos (May, 2017). La Web Semántica, extensión de la web actual, necesita que la información y el conocimiento se publiquen de manera comprensible y procesable por las máquinas. Motivo por el cual, actualmente, muchas organizaciones están en la obligación de transformar sus grandes volúmenes de datos hacia otros formatos (Urrutia et al. 2017).

Para solventar las carencias semánticas de las bases de datos relacionales, una tendencia ha sido el uso de las ontologías. Múltiples trabajos publicados destacan la complementariedad entre las bases de datos relacionales y las ontologías para resolver problemas de integración, consulta y visualización de datos (Liu y Gao, 2018; Zdravkovi et al.; Abbes y Gargouri, 2017; Agapito et al., 2015). Un área de investigación que ha recibido el esfuerzo de la comunidad científica ha sido la transformación de bases de datos relacionales a ontologías. Asimismo, las ontologías se han usado para integrar fuentes de datos heterogéneas, logrando que las bases de datos subyacentes resulten transparentes para los usuarios y las consultas puedan emitirse como si estuvieran sobre una sola base de datos (Liu y Gao, 2018; Wang et al., 2017).

Las ontologías ofrecen una conceptualización explícita y comprensible para los elementos del dominio (Abbes y Gargouri, 2017). Una ontología es una especificación formal y explícita de una conceptualización compartida (Studer, Benjamins y Fensel, 1998). Conceptualización se refiere a una forma de comprender y describir un dominio de conocimiento compartido que es consensuado por el conjunto de interesados expertos en ese dominio y explícita porque está descrita en un lenguaje formal (Urrutia et al., 2017). En la representación abstracta de un dominio en una ontología se identifican los conceptos relevantes (clases), atributos (propiedades), relaciones (roles, asociaciones) y las restricciones (axiomas) (Soylu et al., 2016).

Las ontologías son más ricas y cercanas al vocabulario del usuario final que los esquemas de las bases de datos relacionales. Una ontología conecta conceptos del dominio, mientras que un esquema de base de datos relacional utiliza un método de aplanamiento y dispersión no natural basado en la unión entre las tablas que la componen (Soylu et al., 2016). Ante el hecho de que la mayor parte de los datos reside en bases de datos relacionales y que en la actualidad estos datos se requieren con mayor carga semántica, la vinculación de las bases de datos relaciones y las ontologías es un área de investigación en crecimiento.

Por todo lo anterior, y considerando que la mayoría de los trabajos sobre este campo que se presentan en la literatura se enfocan en un escenario concreto del uso de las ontologías como complemento de las bases de datos relacionales, un estudio con enfoque holístico sobre esta área del conocimiento pudiera considerarse de gran utilidad para el uso de las lecciones aprendidas por la comunidad científica. Por tanto, como objetivo de la presente investigación

se plantea realizar un mapeo sistemático sobre los escenarios en que se vinculan las bases de datos relacionales y las ontologías para brindar una mejor integración, consulta y visualización de los datos almacenados. Un mapeo sistemático ofrece una visión general del tema estudiado, así como la identificación de tendencias y brechas que son aprovechadas por los investigadores para orientar sus estudios (Petersen, Vakkalanka y Kuzniarz, 2015).

El resto del trabajo se ha organizado de la siguiente forma: seguidamente, se explica la metodología utilizada para la realización del mapeo. Luego, se hace un análisis de los resultados de la clasificación obtenida en cuanto a: 1) los escenarios de vinculación, 2) las fuentes en que fueron publicados los trabajos y 3) el propósito de cada uno de los escenarios. Por último, se ofrecen las conclusiones del trabajo.

2. Metodología

Para cumplir con el objetivo del presente trabajo se realizó un mapeo sistemático aplicando la propuesta metodológica establecida por Petersen (Petersen, Vakkalanka y Kuzniarz, 2015) que consistió en los siguientes pasos: 1) determinación de la necesidad y el alcance del mapeo que incluye la definición de las preguntas de investigación, 2) identificación de los trabajos, 3) extracción y clasificación de los datos y 4) análisis y visualización de los resultados.

Determinación de la necesidad y el alcance del mapeo

El propósito del estudio consistió en identificar los escenarios de la vinculación entre las bases de datos relacionales y las ontologías para la integración, consulta y visualización de los datos almacenados. Para orientar el mapeo sistemático se definieron las siguientes preguntas de investigación:

PI 1 ¿Cuáles son los estudios que se han realizado en que se vinculan las bases de datos relacionales y las ontologías? Esta pregunta permite establecer un inventario sobre los trabajos realizados en que se utiliza un enfoque que complementa los modelos relacionales con las ontologías.

PI 2 ¿Cuáles son los escenarios abordados sobre la vinculación de las bases de datos relacionales y las ontologías en cada uno de los trabajos? Esta pregunta permite identificar los escenarios que se presentan en cada uno de los trabajos, así como las tendencias y desafíos que presentan cada uno de esos escenarios.

PI 3 ¿Dónde fue publicado cada uno de los trabajos?

PI 4 ¿Cuál es el propósito de cada uno de los trabajos?

Al responder las preguntas PI 2, PI 3 y PI 4 se pueden clasificar los trabajos según el escenario que abordan, la fuente de publicación y el propósito.

Identificación de los trabajos

La búsqueda de los estudios primarios se realizó en las siguientes bases de datos bibliográficas: IEEEExplore, ACM Digital Library, SpringerLink y ScienceDirect. En cada una de ellas se aplicó la cadena de búsqueda: (*ontology AND database*) OR (*ontologies AND database*). Se realizaron las búsquedas sobre el campo 'resumen' y se seleccionaron los 10 artículos más relevantes en los últimos seis años de cada una de las fuentes para un total de 40 documentos.

Del total de documentos se seleccionaron aquellos que cumplieron con los criterios de inclusión: artículos de revistas científicas en que se vinculan las bases de datos relacionales con las ontologías. Para depurar el listado de documentos se aplicaron los criterios de exclusión que se listan a continuación:

- Duplicados
- Publicados en revistas no arbitradas.
- Enfocados en el almacenamiento de ontologías difusas.
- Basados únicamente en la construcción de ontologías.
- Enfocados en la vinculación de ontologías con bases de datos no relacionales.
- Enfocados en la creación automática de ontologías a partir de textos u otras fuentes de datos no relacionales.

Luego de aplicados los criterios de inclusión y exclusión, de los 40 artículos, fueron seleccionados 16. En la Tabla 1 se relaciona la cantidad de artículos por cada una de las fuentes consultadas.

Tabla 1. Cantidad de artículos seleccionados por cada una de las fuentes consultadas

Base de datos	Encontrados	Seleccionados	Porcentaje
ACM Digital Library	10	5	50 %
IEEEExplore	10	3	30 %
ScienceDirect	10	5	50 %
SpringerLink	10	3	30 %
Total	40	16	40 %

Extracción y clasificación de los datos

La clasificación de los artículos se realizó tomando como referente las preguntas de investigación establecidas que permitieron llegar a las siguientes categorías: 1) escenarios en que se vinculan las bases de datos relacionales y las ontologías, 2) fuentes donde fueron publicados los artículos y 3) propósitos de las investigaciones.

Los resultados de las preguntas de la investigación se encuentran detallados en la Tabla 2, la cual constituye un inventario de los trabajos más relevantes en los últimos seis años sobre la vinculación de las bases de datos relacionales y las ontologías. En la primera columna aparece el título del artículo, le sigue la fuente en la segunda columna y el año de publicación en la tercera columna. En la cuarta columna se representan los tipos de escenarios, los cuales se han representado con los posibles valores: mapeo de bases de datos relacionales a ontologías y viceversa (E1), ontologías para la integración de fuentes de datos heterogéneas (E2), acceso a datos basado en ontologías (OBDA) (E3), consultas en lenguaje natural basado en ontologías (E4), aprendizaje ontológico a partir de bases de datos relacionales (E5) y aporte semántico mediante ontologías (E6). Los tipos de escenarios surgieron de la lectura de los artículos, lo cual según Petersen et al. (2015) ocurre comúnmente en este tipo de investigaciones. En este caso, un artículo puede abordar más de un escenario. En la quinta columna se representa el propósito del trabajo que puede tomar uno de los valores establecidos por Tonella et al. (2007): propuesta

conceptual (PC), prueba de concepto (PRC), cuantificación (C), comparación (CO), comparación condicional (CC) y revisión (R).

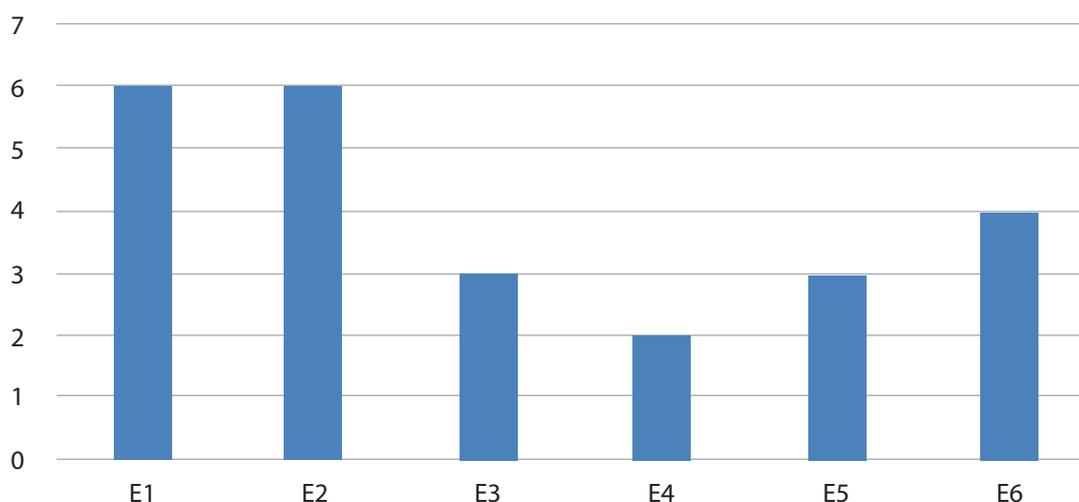
Tabla 2. Listado de los artículos seleccionados

Título del artículo	Fuente	Año	Escenarios	Propósito
RDB2RDF: Completed Transformation from Relational Database into RDF Ontology (Thi et al., 2014).	ACM Digital Library	2014	E1, E5	C
Mapping Relational Databases to Ontology Representation: A Review (Haw y May, 2017).	ACM Digital Library	2017	E1, E2	R
Using Linked Data in the Data Integration for Maternal and Infant Death Risk of the SUS in the GISSA Project (Freitas et al, 2017).	ACM Digital Library	2017	E2	C
Database Ontology-Supported Query for Ubiquitous Environments (Maran, Medeiros y Machado, 2017).	ACM Digital Library	2017	E2	PRC
An Approach for Learning Ontology from Relational Database (Liu y Gao, 2018).	ACM Digital Library	2018	E5	C
Automatic Mapping Magnetic Resonance Images into Multimedia Database Using SIFT (Reynoso et al., 2015).	IEEEExplore	2015	E6	PRC
An Ontology to Assess Data Quality Domains, A Case Study Applied to a Health Care Entity (Urrutia, et al., 2017).	IEEEExplore	2017	E2, E6	PRC
Pay-As-You-Go Methodology for Ontology-Based Data Access (Capsenta y Miranker, 2017)	IEEEExplore	2017	E3, E1	PC
Ontology Based Natural Language Interface for Relational Databases (Sujatha y Raju, 2016)	ScienceDirect	2016	E4	C
Database analysis for ontology learning (Gorskis, Aleksejeva y Polaka, 2016).	ScienceDirect	2016	E3, E5	PC
Ontology-based Data Semantic Management and Application in IoT- and 2 Cloud-Enabled Smart Homes (Tao, Ota y Dong, 2016).	ScienceDirect	2016	E1, E2	C
Automatic approach to enrich databases using ontology: Application in medical domain (Nakhla y Noura, 2017).	ScienceDirect	2017	E6	C
A Learning-based Ontology Alignment Approach Using Inductive Logic Programming (Karimi y Kamandi, 2019).	ScienceDirect	2019	E1	C
Explication and semantic querying of enterprise information systems (Zdravkovi et al., 2013).	SpringerLink	2014	E1, E2	PC
Development of Korean spine database and ontology for realizing e-Spine (Seo et al., 2014).	SpringerLink	2014	E6	PC
Ontology-based end-user visual query formulation: Why, what, who, how, and which? (Soylu et al., 2016).	SpringerLink	2016	E4, E3	PC

3. Resultados y discusión

A continuación, en correspondencia con las preguntas de la investigación se realiza el análisis y visualización de los resultados. Petersen et al. (2015) sugieren usar gráficos para mostrar los resultados de las clasificaciones en un mapeo sistemático. La principal clasificación que se realizó en el presente trabajo tuvo que ver con los tipos de escenarios de vinculación de las bases de datos relacionales y las ontologías, clasificación que se realizó con el objetivo de estructurar el área del conocimiento y encontrar tendencias que sirvieran como punto de partida para investigaciones posteriores. En el gráfico de columnas agrupadas de la Figura 1 se representa la cantidad de artículos que abordan cada uno de los escenarios. Tal como se muestra, el mapeo de bases de datos relacionales a ontologías y viceversa (E1), así como la utilización de las ontologías para la integración de fuentes de datos heterogéneas (E2) fueron los escenarios con mayor presencia en el área de investigación de la vinculación de las bases de datos relacionales y las ontologías, con un 38 % cada uno. El escenario de las consultas en lenguaje natural basado en ontologías (E4) fue el escenario con menor presencia en los artículos de la revisión con un 13 %. Un motivo de lo anterior pudiera ser que el escenario E4, usualmente, se combina con el escenario E3 en las investigaciones y aquí se ha analizado de manera independiente.

Figura 1. Cantidad de artículos por escenario



En el gráfico circular de la Figura 2 se aprecia que el 62 % de los artículos seleccionados para el mapeo fueron publicados en las bases de datos ScienceDirect y ACM Digital Library. En cuanto al tipo de propósito de los artículos revisados se clasificaron siete como cuantificación (C) para un 44 %, cinco como propuesta conceptual (PC) para 31 %, tres como prueba de concepto (PRC) para un 19 % y uno como revisión (R) para un 6 % (ver gráfico de barras agrupadas de la Figura 3). La clasificación de las investigaciones en cuanto al tipo de propósito confirmó la existencia de una investigación robusta sobre el tema en cuestión. Además, se pudo constatar que dentro de los últimos seis años, el 2017 fue un año prominente en cuanto a la cantidad de publicaciones relevantes en el tema (ver gráfico de líneas de la Figura 4).

Aprovechando las clasificaciones realizadas en el mapeo sistemático y con el propósito de encontrar tendencias y desafíos, seguidamente se realiza un análisis en cada uno de los escenarios, fundamentado en los resultados estadísticos obtenidos.

Figura 2. Cantidad de artículos por fuente

■ ACM Digital Library ■ IEEEExplore
■ ScienceDirect ■ SpringerLink

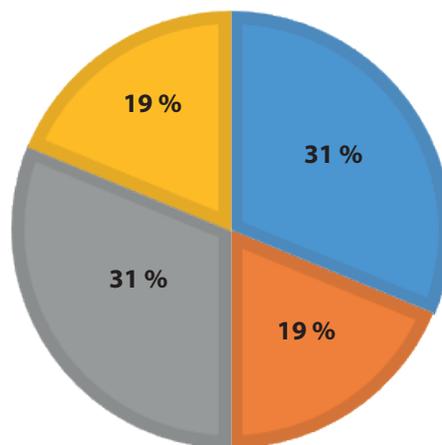


Figura 3. Cantidad de artículos por tipo de propósito

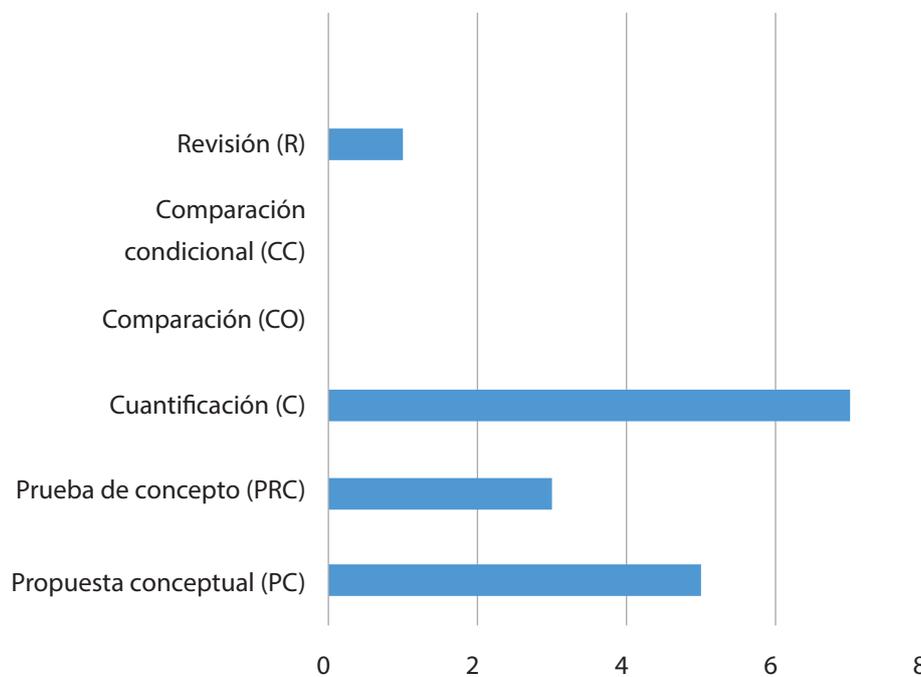
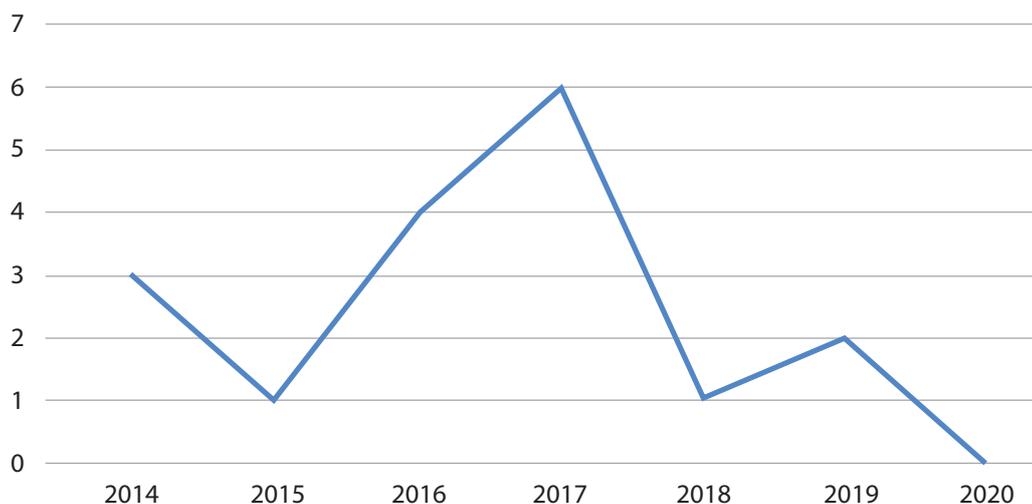


Figura 4. Cantidad de artículos por año

Escenarios de vinculación de las bases de datos relacionales y las ontologías

Los seis escenarios presentados en este trabajo provienen de la apreciación de los autores. Los escenarios no son disjuntos, pueden estar contenidos unos dentro de otros, como es el caso del escenario E1, imprescindible para la correlación entre los dos modelos de datos que se abordan. A continuación, se analizan las tendencias encontradas en cada uno de los escenarios.

Mapeo de bases de datos relacionales a ontologías y viceversa (E1)

El mapeo surgió por la necesidad de transformar las bases de datos relacionales a ontologías y así proporcionarles semántica formal a los datos explícitamente (May, 2017). La Web Semántica ha sido uno de los grandes impulsores de la transformación de bases de datos relacionales a ontologías y, con ello, de los algoritmos de mapeo. El mapeo de una base de datos hacia una ontología es un proceso en el que la semántica implícita de un esquema de base de datos se correlaciona con la estructura de conocimiento explícita y formal de la ontología (Zdravkovi et al., 2013). Aunque se han desarrollado múltiples algoritmos de mapeo el problema principal continúa siendo su automatización. La mayoría de los enfoques que aún se ofrecen son mapeos semiautomáticos (May, 2017).

Uno de los temas que merece especial atención dentro del mapeo es la estrategia a utilizar para el poblado de los datos de la base de datos relacional en la ontología. Dos enfoques se presentan en la literatura: población impulsada por consultas y exportación masiva o volcado masivo (Zdravkovi et al., 2013). En el primer enfoque, se realizan consultas que permiten poblar la ontología con instancias en el momento en que se realiza la consulta semántica sobre la ontología. El segundo enfoque, mantiene la correspondencia completa entre los datos de las tablas de la base de datos y los individuos de la ontología. Este segundo enfoque origina mayores problemas de rendimiento en los procesos de razonamiento debido al gran tamaño de la ontología (Zdravkovi et al., 2013). Para el poblado de las ontologías se han desarrollado marcos

de trabajo tales como: Jena (Ameen et al., 2014), OWL API (Horridge y Bechhofer, 2011) y M2O (Pop et al., 2015).

El mapeo en el sentido inverso también ha sido trabajado con el propósito, principalmente, de utilizar las bases de datos relacionales como medio de almacenamiento de las ontologías. Una base de datos relacional proporciona capacidades de almacenamiento, consultas eficientes y una buena gestión de transacciones, por lo que ha sido un foco de atención durante años utilizar las bases de datos relacionales como forma de almacenamiento para las instancias de las ontologías. Existen tres tipos de métodos de almacenamiento: horizontal, vertical y de descomposición. Tao et al. (2016) propusieron un nuevo método de almacenamiento de tipo descomposición, argumentando que la tendencia actual es desarrollar ese tipo de métodos.

Entre los lenguajes y herramientas utilizados para establecer la alineación (*mapping*) entre la base de datos relacional y las ontologías se encuentran D2RQ, Virtuoso RDF Graphs, Ultrawrap, Spyder, ontop, el estándar R2RML y RDF2OWL (Čerāns y Būmans, 2015). Hazber et al. (2019) presentaron una tabla comparativa de enfoques y herramientas de alineación entre bases de datos relacionales y ontologías. Entre los enfoques existentes, RDF2OWL y su correspondiente herramienta, han demostrado ser útiles para ontologías grandes. Sin embargo, para chequear la completitud y la validez de la alineación se recomienda utilizar otras herramientas tales como D2RQ (Bizer y Seaborne, 2004). El lenguaje RDB2OWL se basa en representar las correspondencias entre los elementos de la base de datos relacional y la ontología a través de anotaciones en los ficheros OWL. A continuación, en las Figuras 5 y 6 se ilustra un ejemplo de alineación entre una base de datos relacional y una ontología. A través de la anotación *rdb2owl:DBExpr* en el fichero OWL se hace corresponder la clase *Teacher* de la ontología a la tabla *XTeacher* de la base de datos.

Figura 5. Tabla XTeacher

	Column Name	Data Type	Allow Nulls
🔑	Teacher_ID	numeric(10, 0)	<input type="checkbox"/>
	Level_Code	varchar(30)	<input checked="" type="checkbox"/>
	IDCode	varchar(30)	<input checked="" type="checkbox"/>
	Name	varchar(40)	<input checked="" type="checkbox"/>
	Nat_Code	varchar(2)	<input checked="" type="checkbox"/>
	Salary	decimal(18, 2)	<input checked="" type="checkbox"/>

Figura 6. Anotación RDB2OWL en el fichero OWL

```
<!-- http://lumii.lv/ontologies/UnivExample.owl#Teacher -->

<owl:Class rdf:about="http://lumii.lv/ontologies/UnivExample.owl#Teacher">
  <rdfs:subClassOf rdf:resource="http://lumii.lv/ontologies/UnivExample.owl#Person" />
  <rdb2owl:DBExpr>XTeacher</rdb2owl:DBExpr>
</owl:Class>
```

En la revisión del presente trabajo se pudo constatar que el 38 % de los artículos analizados abordan el mapeo de bases de datos relacionales a ontologías y viceversa. En ese grupo de artículos clasificados por el tipo de propósito se tienen: tres cuantificaciones (C), dos propuestas conceptuales (PC) y una revisión (R). Clasificados por el tipo de fuente en que fueron publicados se tienen: dos en ACM Digital Library, uno en IEEEExplore, dos en ScienceDirect y uno en SpringerLink (ver Tabla 3).

Tabla 3. Artículos del escenario E1

	ACM Digital Library	IEEEExplore	ScienceDirect	SpringerLink
E1 (2014)	1 (C)			1 (PC)
E1 (2016)			1 (C)	
E1 (2017)	1 (R)	1 (PC)		
E1 (2019)			1 (C)	
Total	2	1	2	1

Ontologías para la integración de fuentes de datos heterogéneas (E2)

Hacia los últimos 10 años una enorme cantidad de información crece y se produce continuamente a un ritmo muy acelerado. La mayoría de esas fuentes de datos son heterogéneas, lo que dificulta la integración y el intercambio de los datos (May, 2017).

La integración de datos consiste en integrar datos que comparten semánticas comunes, pero fueron originados desde fuentes no relacionadas. Para lograr la integración es preciso resolver los problemas de heterogeneidad. La integración de datos implica, principalmente, dos tareas principales. La primera tarea es la integración a nivel de esquema, la cual resuelve las diferencias de homogeneización en los esquemas y la nomenclatura utilizada para representar los datos. La segunda tarea es la integración a nivel de datos, la cual resuelve los diferentes registros a integrar que se refieren a la misma entidad del mundo real. Las ontologías permiten resolver las dos tareas (Liu y Gao, 2018). El mapeo de bases de datos relacionales a ontologías (E1) se considera un caso de integración de datos (Liu y Gao, 2018).

Las ontologías se utilizan ampliamente en la integración de datos, ya que permiten representar el conocimiento como una descripción formal de un dominio de interés (Liu y Gao, 2018). Dentro de la Web Semántica se definen los conceptos y las relaciones que hacen posible la interoperabilidad global (Urrutia et al., 2017).

La Web Semántica puede considerarse como un sistema abierto y distribuido. Como tal, el problema de la heterogeneidad ha sido crítico para ser abordado. Se ha tenido en cuenta la interoperabilidad entre las diversas metodologías de mapeos para soportar diferentes fuentes de datos (May, 2017).

En la revisión del presente trabajo se pudo constatar la importancia de las ontologías para lograr la integración entre bases de datos relacionales heterogéneas. El 38 % de los artículos analizados abordan ese tema. En ese grupo de artículos clasificados por el tipo de propósito se tienen: dos cuantificaciones (C), una propuesta conceptual (PC), una revisión (R) y dos pruebas de conceptos (PRC). Clasificados por el tipo de fuente en que fueron publicados se tienen: tres en ACM Digital Library, uno en IEEEExplore, uno en ScienceDirect y uno en SpringerLink (ver Tabla 4).

Tabla 4. Artículos del escenario E2

	ACM Digital Library	IEEEExplore	ScienceDirect	SpringerLink
E2 (2014)				1(PC)
E2 (2016)			1(C)	
E2 (2017)	1(R), 1(C), 1(PRC)	1(PRC)		
Total	3	1	1	1

Acceso a datos basado en ontologías (OBDA) (E3)

Los sistemas OBDA se encuentran en la categoría de sistemas de información impulsados por ontologías (Soylu et al., 2016). El paradigma de Acceso a Datos Basado en Ontologías (OBDA) proporciona una comunicación cómoda entre los usuarios del negocio y los desarrolladores. Las ontologías proporcionan una abstracción lógica del negocio independiente de cómo y dónde se almacenan los datos físicamente. Consiste en tres componentes: una ontología objetivo, un fichero de mapeo y las consultas. Las consultas son definidas en términos de la abstracción lógica de la ontología en lugar de la estructura física de la base de datos (Capsenta y Miranker, 2017).

Los sistemas OBDA tienen dos enfoques para su implementación: materialización y virtualización. La materialización consiste en crear un almacén de tripletas en formato RDF a partir de las bases de datos de origen. Las consultas se realizan con el lenguaje SPARQL sobre el almacén de tripletas RDF (Sequeda y Miranker, 2013). En el enfoque de virtualización, las consultas en SPARQL se reescriben en consultas SQL y los datos permanecen en su lugar con su estructura original (Gorskis et al., 2016). Los datos apropiados son virtualizados en formato RDF mediante una capa adicional de ontologías. En caso de existir más de una fuente de datos se realizan consultas federadas (Capsenta y Miranker, 2017; Soyly et al., 2016). Específicamente, la federación se considera otro beneficio del paradigma OBDA que consiste en integrar fuentes de datos distribuidas con esquemas diferentes. El uso de las tecnologías y estándares de la Web Semántica hace posible aplicar e integrar el enfoque OBDA en un contexto más amplio como el de los datos públicos disponibles en la Web (por ejemplo los datos enlazados) (Soyly et al., 2016). Con el enfoque OBDA, un sistema tradicional de bases de datos se transforma en una base de conocimiento con una arquitectura de múltiples capas coherente con el principio de independencia de los datos (Capsenta y Miranker, 2017).

Los desafíos actuales que presenta OBDA coinciden con dos áreas de investigación abiertas en este ámbito: ingeniería de ontologías e ingeniería de mapeo (Capsenta y Miranker, 2017). Incluso, ambos en el enfoque OBDA tienen mayor relevancia debido al gran tamaño que tienen las bases de datos relacionales. Dentro de la ingeniería ontológica, uno de los desafíos está relacionado con las capacidades de razonamiento distribuido en ontologías modulares con importaciones dinámicas (Soyly et al., 2016).

En la revisión del presente trabajo se pudo constatar que este campo es un área de investigación abierta. El 19 % de los artículos analizados abordan explícitamente el paradigma de Acceso a Datos Basado en Ontologías (OBDA). En ese grupo de artículos clasificados por el tipo de propósito se tienen tres propuestas conceptuales (PC). Clasificados por el tipo de fuente en que fueron publicados se tienen: uno en IEEEExplore, uno en ScienceDirect y uno en SpringerLink (ver Tabla 5).

Tabla 5. Artículos del escenario E3

	IEEEExplore	ScienceDirect	SpringerLink
E3 (2016)		1(PC)	1(PC)
E3 (2017)	1(PC)		
Total	1	1	1

Consultas en lenguaje natural basado en ontologías (E4)

Las consultas en lenguaje natural constituyen una de las vías más sugerentes para ampliar el acceso a los datos. Google es un ejemplo fehaciente de ello. Sin embargo, para consultar una base de datos relacional es preciso conocer un lenguaje estructurado. Los esfuerzos actuales que se realizan en ese sentido se basan en utilizar las ontologías para permitir la realización de consultas en lenguaje natural y, con ello, ampliar el acceso a los datos.

En la formulación de interfaces en lenguaje natural se utilizan las ontologías como un medio natural de acceso a las fuentes de datos tradicionales (Soylu et al., 2016). En los sistemas para el procesamiento de consultas en lenguaje natural sobre bases de datos relacionales, las ontologías se construyen específicas para cada base de datos y su aporte principal está en el análisis semántico. Las ontologías permiten reconocer las relaciones entre las palabras de la cadena de búsqueda y la base de datos (Sujatha y Raju, 2016).

A su vez las ontologías son responsables de convertir las palabras ingresadas por el usuario en expresiones del lenguaje formal. Normalmente, este tipo de sistemas operan con *corpus* de palabras. Todos los idiomas tienen dos categorías de palabras: las de contenido y las de función. Las palabras de contenido son: verbos, sustantivos, adjetivos y adverbios. Mientras que las palabras de función son: preposiciones, conjunciones, pronombres e interjecciones. Cada idioma tiene repositorios de palabras de contenido y de función (Sujatha y Raju, 2016). Además, las ontologías deben procesar las frases que requieren un procesamiento matemático, por ejemplo: menor que, igual a, entre otras (Sujatha y Raju, 2016).

La creación de interfaces en lenguaje natural basada en ontologías sobre bases de datos relacionales con altos niveles de recuperación y precisión continúa siendo un problema de investigación. Se necesitan sistemas que puedan abordar la robustez y la escalabilidad de las aplicaciones, así como la internalización en diferentes idiomas (Soylu et al., 2016). Este tipo de sistema está obligado a mitigar la cobertura incompleta de los lenguajes, así como otros múltiples desafíos (Sujatha y Raju, 2016).

En la revisión del presente trabajo se pudo constatar que este campo es un área de investigación abierta. El 13 % de los artículos analizados abordan el tema de las consultas en lenguaje natural basado en ontologías. Uno de los motivos por los cuales el porcentaje no es alto es que este tema, usualmente, se combina en las publicaciones con el paradigma de Acceso a Datos Basado en Ontologías (OBDA) y en esta investigación se ha querido analizar por separado. En ese grupo de artículos clasificados por el tipo de propósito se tienen: una propuesta conceptual (PC) y una cuantificación (C). Clasificados por el tipo de fuente en que fueron publicados se tienen: uno en ScienceDirect y uno en SpringerLink (ver la Tabla 6).

Tabla 6. Artículos del escenario E4

	ScienceDirect	SpringerLink
E4 (2016)	1(C)	1(PC)
Total	1	1

Aprendizaje ontológico a partir de bases de datos relacionales (E5)

El objetivo principal de los métodos de aprendizaje de ontología es derivar automáticamente una ontología de los datos existentes. La creación automática de ontologías putativas a partir de bases de datos relacionales está implícita en los esfuerzos de mapeo (escenario E1) y en la mayoría de los sistemas de acceso a datos basados en ontologías (escenario E3) (Gorskis et al., 2016).

Los estudios más recientes sobre la generación de ontologías a partir de bases de datos relacionales se han orientado a reducir al mínimo posible la intervención del usuario. En uno de los enfoques analizados en la revisión se logró que la transformación pueda ser reversible, manteniendo las relaciones de claves externas y primarias de las tablas. En ese enfoque se usa el vocabulario RDFS en lugar de RDF, lo que le proporciona un mayor nivel de expresividad a la ontología generada. La transformación en el enfoque se logra sin la intervención del usuario (Thi et al., 2014).

Cuando se utiliza como entrada del proceso de generación de ontologías únicamente a la base de datos relacional, dos elementos que en ella no estén relacionados nunca podrán relacionarse en la ontología. En ese sentido, un elemento importante ha sido utilizar otras ontologías o tesauros externos que puedan aportarle semántica a la base de datos relacional que se desea transformar. En uno de los artículos analizados se utilizó el tesauro WordNet (Miller, 1995) para obtener otras relaciones (Thi et al., 2014).

En la revisión del presente trabajo se pudo constatar que, en la mayoría de las investigaciones en que se obtienen ontologías a partir de bases de datos relacionales, están presentes esfuerzos para minimizar al máximo posible la intervención del usuario y, con ello, lograr el aprendizaje ontológico. El 19 % de los artículos analizados abordan ese tema. En ese grupo de artículos clasificados por el tipo de propósito se tienen: una propuesta conceptual (PC) y dos cuantificaciones (C). Clasificados por el tipo de fuente en que fueron publicados se tienen: dos en ACM Digital Library y uno en ScienceDirect (ver la Tabla 7).

Tabla 7. Artículos del escenario E5

	ACM Digital Library	ScienceDirect
E5 (2014)	1(C)	
E5 (2016)		1(PC)
E5 (2018)	1(C)	
Total	2	1

Aporte semántico mediante ontologías (E6)

En este escenario se agrupan los trabajos en los cuales el aporte de las ontologías lo recibe directamente la base de datos relacional. Uno de los escenarios en que se vinculan las ontologías

para aportarle semántica a las bases de datos relacionales es en el proceso de enriquecimiento de una base de datos relacional. En una organización, usualmente, se tiene una base de datos por cada uno de los sistemas de información. Por ejemplo, en un hospital donde hay dos sistemas informáticos, uno para la atención a los pacientes y otro para la gestión contable, se tienen dos bases de datos respectivas que a la vez tienen ciertas similitudes. Utilizando el enriquecimiento se pueden integrar ambas en una única base de datos objetivo o se pueden actualizar una a partir de la otra. El proceso de enriquecimiento es guiado por ontologías (Nakhla y Nouira, 2017).

Nakhla y Nouira (2017) propusieron un enfoque novedoso de enriquecimiento de bases de datos que permite, a diferencia de los enfoques que le precedieron, realizar un enriquecimiento completo de la base de datos de forma automática. Para ello, utilizaron un conjunto de reglas incorporadas a una ontología que le permitieron comparar los registros y la estructura de la base de datos.

Las ontologías también permiten hacer evaluaciones de calidad de las bases de datos relacionales. La limpieza de datos se define como el proceso de detectar y corregir errores en bases de datos, implementando una ontología para clasificar los tipos de errores sintácticos y semánticos. El proceso de limpieza de los datos tiene gran significación en entornos de integración de datos. Urrutia y Chávez (2017) definieron un proceso flexible de eliminación de anomalías en los datos. En ese proceso se utilizaron las métricas de calidad: existencia, unicidad y consistencia incorporadas en una ontología, se insertaron los datos en la ontología y mediante un proceso de inferencia se realizó una evaluación automática de la base de datos (Urrutia et al., 2017).

En sentido general, las ontologías son muy útiles para la representación del conocimiento. Reynoso y Cuevas utilizaron ontologías para representar conocimiento sobre resonancias magnéticas de diferentes partes del cuerpo humano. Las imágenes del inventario de resonancias se encontraban almacenadas en una base de datos. El uso de las ontologías automatizó la clasificación de las imágenes y facilitó su integración, consulta y visualización (Reynoso et al., 2015). Asimismo, López, Hidalgo y Silega (2016), presentaron un método para la integración de ontologías y sistemas relacionales que utiliza las ontologías para enriquecer semánticamente una base de datos relacional de créditos bancarios. El método consta de siete actividades que comprenden: el desarrollo en paralelo del modelo relacional y el modelo ontológico, el poblado del modelo ontológico a partir de los datos almacenados en la base de datos relacional, la inferencia de conocimiento en el modelo ontológico y su inserción en el modelo relacional. El método constituye un ejemplo de la complementariedad que se obtiene de la integración de las bases de datos relacionales y las ontologías.

En la revisión del presente trabajo se pudo constatar que las ontologías son una técnica poderosa para la representación del conocimiento. El 25 % de los artículos analizados abordan el tema de utilizar las ontologías para aportarle semántica a las bases de datos relacionales. En ese grupo de artículos clasificados por el tipo de propósito se tienen: una propuesta conceptual (PC), una cuantificación (C) y dos pruebas de conceptos (PRC). Clasificados por el tipo de fuente en la que fueron publicados se tienen: dos en IEEExplore, uno en ScienceDirect y uno en SpringerLink (ver la Tabla 8).

Tabla 8. Artículos del escenario E6

	IEEEExplore	ScienceDirect	SpringerLink
E6 (2014)			1(PC)
E6 (2015)	1(PRC)		
E6 (2017)	1(PRC)	1(C)	
Total	2	1	1

4. Conclusiones y recomendaciones

En el mapeo sistemático realizado en la presente investigación se identificaron artículos que vinculan las bases de datos relacionales y las ontologías. Esos artículos fueron clasificados en cuanto a: 1) el escenario de vinculación de las bases de datos relacionales y las ontologías abordado, 2) el tipo de propósito de la investigación y 3) la fuente en la que fue publicado. Las respuestas de las preguntas de la investigación son presentadas seguidamente.

PI 1: ¿Cuáles son los estudios que se han realizado en que se vinculan las bases de datos relacionales y las ontologías? De los 40 artículos encontrados se seleccionaron 16 que cumplieron con los criterios de inclusión y exclusión definidos. La mayor parte de los artículos desechados se enfocaron únicamente a las ontologías o las relacionaban con bases de datos no relacionales.

PI 2: ¿Cuáles son los escenarios abordados sobre la vinculación de las bases de datos relacionales y las ontologías en cada uno de los trabajos? A partir de la lectura de los artículos seleccionados se definieron seis escenarios en que se vinculan las bases de datos relacionales y las ontologías con el objetivo de brindar una mejor integración, consulta y visualización de los datos almacenados. Los escenarios identificados fueron: mapeo de bases de datos relacionales a ontologías y viceversa (E1), ontologías para la integración de fuentes de datos heterogéneas (E2), acceso a datos basado en ontologías (OBDA) (E3), consultas en lenguaje natural basado en ontologías (E4), aprendizaje ontológico a partir de bases de datos relacionales (E5) y aporte semántico mediante ontologías (E6). En la clasificación realizada, el mapeo de bases de datos relacionales a ontologías y viceversa (E1), así como la utilización de las ontologías para la integración de fuentes de datos heterogéneas (E2) fueron los escenarios con mayor presencia para un 38 % cada uno, mientras que el escenario de las consultas en lenguaje natural basada en ontología (E4) fue el escenario con menor presencia para un 13 %. Por cada uno de los escenarios se delimitaron tendencias (T) y desafíos (D) que a continuación se mencionan.

Mapeo de bases de datos relacionales a ontologías y viceversa (E1)

- El 38 % de los artículos revisados abordaron este escenario, lo que evidencia la vigencia del tema.
- (D) aunque se han desarrollado múltiples algoritmos de mapeo, la mayoría de los enfoques ofrecen mapeos semiautomáticos y el problema principal continúa siendo su automatización.
- (T) dentro de los dos tipos de poblado de ontologías analizados se considera el poblado impulsado por consultas como el más factible.
- (T) dentro de los tres métodos analizados para el almacenamiento de ontologías en bases de datos relacionales: vertical, horizontal y de descomposición, la tendencia actual está enfocada en el método de descomposición.

Ontologías para la integración de fuentes de datos heterogéneas (E2)

- El 38 % de los artículos revisados abordaron este escenario, lo que evidencia la vigencia del tema.
- (T) las ontologías permiten resolver las dos tareas que abarca la integración de las bases de datos relacionales: integración a nivel de esquema y a nivel de datos.

Acceso a datos basado en ontologías (OBDA) (E3)

- El 19 % de los artículos revisados abordaron este escenario, lo que evidencia la vigencia del tema.
- (D) los desafíos actuales que presenta OBDA coinciden con dos áreas de investigación abiertas en este ámbito: ingeniería de ontologías e ingeniería de mapeo, aquí con mayor relevancia debido al gran tamaño que tienen las bases de datos relacionales.

Consultas en lenguaje natural basado en ontologías (E4)

- El 13 % de los artículos revisados abordaron este escenario, lo que evidencia la vigencia del tema.
- (D) se necesitan sistemas robustos y escalables, así como la internalización en diferentes idiomas.
- (D) este tipo de sistemas está obligado a mitigar la cobertura incompleta de los lenguajes.

Aprendizaje ontológico a partir de bases de datos relacionales (E5)

- El 19 % de los artículos revisados abordaron este escenario, lo que evidencia la vigencia del tema.
- (T) los estudios más recientes sobre la generación de ontologías a partir de bases de datos relacionales se han orientado a reducir al mínimo posible la intervención del usuario.
- (T) un elemento importante ha sido utilizar otras ontologías o tesauros externos para obtener relaciones que complementen las explícitas de la base de datos relacional.

Aporte semántico mediante ontologías (E6)

- El 25 % de los artículos revisados abordaron este escenario, lo que evidencia la vigencia del tema.
- (T) en el empleo de las ontologías como complemento semántico de las bases de datos relacionales se constataron diferentes usos en la revisión, dentro de ellos: el enriquecimiento, la evaluación de la calidad y la clasificación de los datos.

PI 3 ¿Dónde fue publicado cada uno de los estudios? En ACM Digital Library y en ScienceDirect fueron publicados el 31 % de los artículos seleccionados en cada una de ellas, mientras que en IEEExplore y en SpringerLink fueron publicados el resto para un 19 % en cada una.

PI 4 ¿Cuál es el propósito de cada estudio? De los artículos revisados se clasificaron siete como cuantificación (C) para un 44 %, cinco como propuesta conceptual (PC) para un 31 %, tres como prueba de concepto (PRC) para un 19 % y uno como revisión (R) para un 6 %.

Luego de respondidas las preguntas de la investigación se puede concluir que, la vinculación de las bases de datos relacionales y las ontologías parten de la existencia de una gran cantidad de sistemas legados con bases de datos relacionales y la necesidad de manipular semánticamente esos datos con apoyo de las ontologías. En el trabajo, se presentaron ejemplos que evidencian el esfuerzo dedicado por la comunidad científica para eliminar las brechas entre esos dos modelos de datos con el objetivo de aprovechar cada vez más su complementariedad.

Referencias

- Abbes, H.; Gargouri, F. (2017). MongoDB-Based Modular Ontology Building for Big Data Integration. *Journal on Data Semantics*, 7: 1-27. <https://doi.org/10.1007/s13740-017-0081-z>
- Ameen, A. et al. (2014). Reasoning in Semantic Web Using Jena. *Computer Engineering and Intelligent Systems*, 5(4): 39-48. <https://core.ac.uk/download/pdf/234644794.pdf>
- Bizer, C.; Seaborne, A. (2004). D2RQ-treating non-RDF databases as virtual RDF graphs. In Proceedings of the 3rd international semantic web conference (ISWC2004) (Vol. 2004). Springer. <https://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.126.2314>
- Capsenta, J. F. S.; Miranker, D. P. (2017). A Pay-As-You-Go Methodology for Ontology-Based Data Access. *IEEE Internet Computing*, 21(2): 92-96. <https://doi.org/10.1109/MIC.2017.46>
- Čerāns, K.; Būmans, G. (2015). RDB2OWL: a language and tool for database to ontology mapping. In Proceedings of the CAISE 2015 Forum at the 27th International Conference on Advanced Information Systems Engineering (CAISE 2015), Kista, Sweden (81-88). <http://ceur-ws.org/Vol-1367/paper-11.pdf>
- Freitas, R., et al. (2017). Using linked data in the data integration for maternal and infant death risk of the SUS in the GISSA Project. In Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web (193-196). <https://doi.org/10.1145/3126858.3131606>
- Gorskis, H.; Aleksejeva, L.; Polaka, I. (2016). Database Analysis for Ontology Learning. *Procedia Computer Science*, 102: 113-120. <https://doi.org/10.1016/j.procs.2016.09.377>
- Haw, S. C.; May, J. W. (2017). Mapping Relational Databases to Ontology Representation: A Review. In Proceedings of the International Conference on Digital Technology in Education, 54-58. <https://doi.org/10.1145/3134847.3134852>
- Hazber, M. A., et al. (2019). A survey: Transformation for integrating relational database with semantic Web. In Proceedings of the 2019 3rd International Conference on Management Engineering, Software Engineering and Service Sciences (66-73). <https://dl.acm.org/doi/10.1145/3312662.3312692>
- Horrige, M.; Bechhofer, S. (2011). The OWL API: A Java API for OWL Ontologies. *Semantic Web*, 2(1): 11-21. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.678.6080&rep=rep1&type=pdf>
- Karimi, H.; Kamandi, A. (2019). PT US CR. Expert Systems With Applications, 125: 412-424. <https://doi.org/10.1016/j.eswa.2019.02.014>
- Liu, X.; Gao, F. (2018). An Approach for Learning Ontology from Relational Database. Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence. <https://doi.org/10.1145/3302425.3302495>
- López Rodríguez, Y. A.; Hidalgo Delgado, Y.; Silega Martínez, N. (2016). Método para la integración de ontologías en un sistema para la evaluación de créditos. *Revista Cubana de Ciencias Informáticas*, 10(4): 97-111. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992016000400007
- Maran, V.; Medeiros, G.; Machado, A. (2017). Database Ontology-Supported Query for Ubiquitous Environments. In Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web, 185-188. <https://doi.org/10.1145/3126858.3131575>
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11): 39-41. <https://doi.org/10.1145/219717.219748>

- Nakhla, Z.; Nouria, K. (2017). Automatic approach to enrich databases using ontology: Application in medical domain. *Procedia Computer Science*, 112: 387-396. <https://doi.org/10.1016/j.procs.2017.08.221>
- Petersen, K.; Vakkalanka, S.; Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering : An update. *Information and software technology*, 64: 1-18. <https://doi.org/10.1016/j.infsof.2015.03.007>
- Pop, C., et al. (2015). M2O: A Library for Using Ontologies in Software Engineering. *Intelligent Computer Communication and Processing (ICCP)*: 69-75. <https://doi.org/10.1109/ICCP.2015.7312608>
- Reynoso, J. L., et al. (2015). Automatic Mapping Magnetic Resonance Images into Multimedia Database Using SIFT. *IEEE Latin America Transactions*, 13(8): 2709-2714. <https://doi.org/10.1109/TLA.2015.7332153>
- Seo, D., et al. (2014). Development of Korean spine database and ontology for realizing e-Spine. Cluster computing. Recuperado de <https://link.springer.com/article/10.1007/s10586-013-0344-x>
- Sequeda, J. F.; Miranker, D. P. (2013). Ultrawrap: SPARQL execution on relational data. *Journal of Web Semantics*, 22: 19-39. <https://doi.org/10.1016/j.websem.2013.08.002>
- Soylu, A., et al. (2016). Ontology-based end-user visual query formulation: Why, what, who, how, and which? *Universal Access in the Information Society*, 16: 435-467 <https://doi.org/10.1007/s10209-016-0465-0>
- Studer, R.; Benjamins, V. R.; Fensel, D. (1998). Knowledge Engineering : Principles and Methods. *Data and Knowledge engineering*, 25(1): 161-197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- Sujatha, B.; Raju, S. V. (2016). Ontology Based Natural Language Interface for Relational Databases. *Procedia Computer Science*, 92: 487-492. <https://doi.org/10.1016/j.procs.2016.07.372>
- Tao, M.; Ota, K.; Dong, M. (2016). Ontology-based Data Semantic Management and Application in IoT- and Cloud-Enabled Smart Homes. *Future Generation Computer Systems*, 76: 528-539 <https://doi.org/10.1016/j.future.2016.11.012>
- Thi, P., et al. (2014). RDB2RDF : Completed Transformation from Relational Database into RDF Ontology. In Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, 88. <https://doi.org/10.1145/2557977.2558083>
- Tonella, P., et al. (2007). Empirical studies in reverse engineering : state of the art and future trends. *Empirical Software Engineering*, 12(5): 551-571. <https://doi.org/10.1007/s10664-007-9037-5>
- Urrutia, A., et al. (2017). An Ontology to Assess Data Quality Domains. A Case Study Applied to a Health Care Entity. *IEEE Latin America Transactions*, 15(8): 1506-1512. <https://doi.org/10.1109/TLA.2017.7994799>
- Zdravkovi, M., et al. (2013). Explication and semantic querying of enterprise information systems. *Knowledge and information systems*, 40(3): 697-724. <https://doi.org/10.1007/s10115-013-0650-x>