

Extracción de respuestas a partir de ontologías utilizando patrones estructurales en SQWRL

Gabriela A. García-Robledo, Beatriz A. González-Beltrán,
José A. Reyes-Ortiz, Maricela Bravo

Universidad Autónoma Metropolitana,
Unidad Azcapotzalco,
México

{a12181800075,bgonzalez,jaro,mcbc}@azc.uam.mx

Resumen. La extracción de respuestas es una tarea que involucra un reto para la disciplina de Extracción de Información (*EI*). El uso de ontologías aminora la complejidad de esta tarea. Sin embargo, los desafíos vienen con el procesamiento de la petición del usuario (consulta) y su transformación en un lenguaje formal para la extracción de información de las ontologías. Por ello, en este artículo se propone un método para procesar una consulta semiestructurada, transformada automáticamente en el lenguaje *SQWRL* (Semantic Query-Enhanced Web Rule) y extraer respuestas de un sistema de ontologías del dominio académico utilizando patrones. El proceso completo consiste en, primero, extender la consulta semiestructurada con los elementos necesarios para la extracción de información del sistema de ontologías; después, se crean patrones estructurales para mapear entidades identificadas en la consulta a entidades existentes en el sistema de ontologías; finalmente, la respuesta es obtenida con la ejecución de los patrones generados en *SQWRL*. El objetivo principal de las consultas es contestar a preguntas de tipo ¿quién?, ¿dónde? y ¿cuándo? en un dominio académico. Una evaluación del método propuesto ha sido realizada con un conjunto de 258 preguntas formuladas y evaluadas por expertos en el dominio académico, obteniendo resultados alentadores con una precisión cercana al 97 %.

Palabras clave: Extracción de información, ontologías, reconocimiento de entidades, patrones estructurales en SQWRL.

Responses Extraction from Ontologies using Structural Patterns in SQWRL

Abstract. Question-answering is a challenge for the Information Extraction (*EI*) task. The use of ontologies reduces the complexity of this task. However, the challenges are presented in the processing of the user request through a query and how it is converted into a formal language for information extraction from ontologies. This paper proposes a method

that starts with a semi-structured query in order to translate it into the *SQWRL* language (Semantic Query-Enhanced Web Rule) and extract a suitable answer from an ontology system of the academic domain using patterns. The whole process consists of extending the semi-structured query with the necessary elements to extract relevant information from the ontology system; structural patterns are created to match entities identified in the query with existing entities in the ontology system; the answer is obtained with the execution of patterns generated in *SQWRL*. The queries are able to answer questions about who? where? and when? in an academic domain. An evaluation process has been carried out with a set of 258 questions, which were expressed and evaluated by experts in the academic domain, obtaining promising results with precision close to 97%.

Keywords: Information extraction, ontologies, entity recognition, structural patterns in *SQWRL*.

1. Introducción

En un ambiente académico se realizan diariamente actividades de docencia, investigación y difusión, tales como: asesorías, congresos, seminarios, clases, talleres, cursos, entre otros. También existen instalaciones principales, tales como: laboratorios de cómputo, cafetería, librería, salones de clase, auditorios, entre otros. Finalmente, diversos participantes realizan actividades dentro de ellas como: administrativos, profesores y alumnos.

Una de las formas de almacenamiento y representación de la información generada de estas actividades o eventos, son haciendo uso de ontologías, las cuales permiten estructurar dicha información de tal manera que se facilite su consulta. Sin embargo, la consulta de información sobre eventos y sus aspectos relacionados (personas, tiempo y lugar) involucra una petición de un usuario. Debido a la gran cantidad de información generada en el dominio académico, este proceso de consulta a partir de una petición en lenguaje natural o en un formato estructurado, requiere de un procesamiento computacional y de su transformación hacia lenguajes formales.

Los lenguajes formales de consulta ayudan a obtener información dentro de una ontología, dentro de la literatura los principales son: a) *SPARQL* (Protocol and RDF Query Language), que es un lenguaje de consulta “orientado a objetos” porque solo consulta la información contenida en los modelos[1]; b) *SWRL* (Semantic Web Rule Language) es un formalismo para integrar reglas con ontologías en la Web Semántica y esta basado en una combinación de dos lenguajes *OWL* (Web Ontology Language) y *RuleML* (Rule Markup Language)[3]. *SWRL* introduce las reglas al adaptar la semántica existente por un lenguaje de reglas directamente en la capa de ontología[2]. La interacción entre ontologías y el lenguaje de consulta se integran en átomos antecedentes evaluados por una conjunción y los átomos consecuentes, ambos pueden ser vacíos[3].

El proceso de transformar una consulta estructurada que represente la petición de un usuario es un reto hoy en día. Por ello, el objetivo principal de este trabajo es partir de una petición de un usuario expresada en una consulta semiestructurada obtenida en [15] y extraer una respuesta de un sistema de ontologías del dominio académico. Este proceso involucra el enriquecimiento de una consulta semiestructurada, mapeo de entidades a elementos del sistema de ontologías, la identificación de patrones estructurales (sucesión de elementos estructurados identificados con frecuencia dentro de distintas preguntas) y la ejecución de la consulta en *SQWRL* para obtener una respuesta específica sobre eventos, espacios físicos y actores involucrados (personas) que pertenecen al dominio académico. La elección de este lenguaje de consulta se basó en la afinidad entre la estructura resultante de [15] y la sintaxis de átomos antecedentes y consecuentes. El sistema de ontologías utilizado modela, entre otros conceptos, la noción de personas, tiempo y lugares; con ello es posible contestar preguntas del tipo ¿quién?, ¿dónde? y ¿cuándo?.

El resto del artículo se organiza de la siguiente manera. En la Sección 2 se presentan los avances más importantes y recientes respecto a las áreas de investigación de este trabajo, tales como: a) el uso de ontologías para representar información de eventos; b) el uso de lenguajes formales para extraer información de ontologías; c) métodos de transformación de consultas en lenguaje natural (no estructurada), semi-estructurada o estructurada. Por su parte, en la Sección 3 se expone el método propuesto para la transformación de la consulta semiestructurada en una consulta en *SQWRL* con la finalidad de extraer información del sistema de ontologías y obtener una respuesta. La evaluación del método propuesto es presentada en la Sección 4 mediante una experimentación con 258 preguntas y mostrando resultados alentadores. Finalmente, en la Sección 5 se presentan las conclusiones de este artículo y el trabajo a futuro.

2. Trabajos relacionados

En esta sección se presenta una revisión de los avances dentro del área de investigación de este trabajo. Existen diversos trabajos que se fundamentan en el uso de ontologías para las consultas de información en lenguaje natural, pero pocos lo han aplicado usando un método de extracción en consultas *SQWRL*.

López, V., et al.[4] en 2005, reportan a Aqualog, donde proponen una arquitectura en cascada, la consulta es traducida a un conjunto de tripletas compatibles con la ontología. El modelo de tripletas que utilizan es de la forma “sujeto, predicado, verbo”. Requiere de los recursos de GATE (General Architecture for Text Engineering) para tareas de procesamiento de lenguaje natural. El sistema identifica términos, relaciones, indicadores de preguntas como lo son “qué/quién/cuándo”, patrones o tipos de preguntas. Por otro lado, en el mismo año Muhammad, A., et al.[5] desarrollan un método de extracción de información biológica basada en una ontología y sistema de respuesta a consultas con cuatro componentes: una etiqueta basada en una ontología, un etiquetador de nuevos

textos, una base de conocimiento potenciador que mejora la ontología e incorpora nuevos conocimientos y un procesador de consultas de usuarios.

La propuesta de Bernstein, A., et al.[6] en el 2006 describe a Ginseng, una interfaz de lenguaje natural que utiliza reglas gramaticales estáticas que proporcionan las estructuras y frases básicas del idioma. Ginseng utiliza el vocabulario definido por las ontologías, no necesita interpretar las consultas, maneja una gramática multinivel, un analizador incremental y una capa de consulta a la ontología a través de SPARQL.

En el 2007, Tianyong, H., et al.[7] proponen un tipo de patrón semántico, que se puede usar para publicar preguntas y respuestas en sistemas QA interactivos para el usuario. El procedimiento incluye un análisis de estructura de preguntas, coincidencia, generación y clasificación de patrones para la extracción de respuesta. Implementan una interfaz de usuario para hacer uso del patrón semántico en un sistema de control de calidad que permite a los usuarios publicar y responder preguntas.

Tablan, V., et al.[8] en el 2008 presentan QuestIO System, que es una interfaz de lenguaje natural sin un vocabulario predefinido, realiza como primera parte un análisis lingüístico por medio de un tokenizador y etiquetado, después ejecuta el diccionario ontológico que crea anotaciones acerca de la identificación de las partes importantes de la ontología, posteriormente se inicia una transformación de la consulta para generar una formal.

Damljanovic, D., et al.[9] en el 2010 crean FREYA, una interfaz de lenguaje natural que utiliza búsquedas en *SPARQL* en ontologías, ejecuta preguntas en lenguaje natural y genera una respuesta en forma de un grafo. La interfaz también utiliza el conocimiento de la ontología para la identificación de términos, con los que genera tripletas para realizar la búsqueda. Encuentra el tipo de respuesta donde se realiza un análisis sintáctico y localiza la frase relevante que define a la pregunta y lo que se busca.

También, existe un método en el 2015 desarrollado por Eka, A., et al.[10] que se enfoca en analizar preguntas de tipo ¿Por qué? La propuesta adapta el modelo simple de bolsa de palabras, utiliza entidades semánticas para representar una consulta que al ser ejecutada se enriquecen con ayuda de *SQWRL* sobre la ontología de dominio.

Existen desarrollos en la geomática y geografía, Bereta, K., et al.[11] presentan un sistema capaz de responder consultas *GeoSPARQL* (A Geographic Query Language for RDF) sobre base de datos relacionales, traducen *GeoSPARQL* a *SQL* utilizando ontologías y asignaciones.

Una implementación relacionada también con *SPARQL* se presenta en Solís, A., et al.[12]. Una interfaz de lenguaje natural para deducción de información almacenada en ontologías. Utiliza solo *SPARQL* como lenguaje de consulta en ontologías. El módulo procesador obtiene etiquetas gramaticales y lemas del texto utilizando Freeling. La información obtenida es enviada a un módulo de conocimiento capaz de generar consultas y ejecutarlas.

En el 2019 Setyo, F., et al.[13] desarrollan un clasificador de instancias para corpus pequeños aplicados a un sistema de respuesta de preguntas semánticas.

Consta de un pre procesamiento, análisis morfológico y semántico, extracción de características, clasificación de instancias con un algoritmo de redes de función de base radial y un módulo de transformación.

Finalmente, en el sector de la salud, Mourad, S. y Said, O.E.A. [14] en este año desarrollan SemiBioNLQA, que consiste en clasificar preguntas, recuperar documentos y extraer una respuesta. El sistema trata cuatro tipos de preguntas y se basa en: patrones artesanales sintácticos de México, un algoritmo de aprendizaje automático para la clasificación de preguntas, un motor de búsqueda y similitud para la recuperación de documentos, análisis de sentimientos y una métrica de frecuencia para la extracción de la respuesta.

Con la revisión de los trabajos relacionados a las disciplinas involucradas en este artículo, se puede observar la importancia de realizar una adecuada transformación de una consulta no estructurada (lenguaje natural) o semiestructurada hacia una consulta estructurada, apoyándose de un lenguaje formal con la finalidad de extraer información de algún medio por ejemplo las ontologías. Además, se puede observar que la mayoría de los trabajos se enfocan en el inglés, esto conlleva a una necesidad latente de contar, con un método de extracción de información partiendo de una consulta en español, lenguaje en el cual existe una carencia de herramientas computacionales. Por ello, en la siguiente sección se presenta el método propuesto para transformar una consulta semiestructurada en consulta SQWRL y obtener una respuesta de un sistema de ontologías del dominio académico.

3. Método propuesto

En esta sección se presenta el método propuesto para transformar una consulta semiestructurada que expresa la petición del usuario en una consulta en SQWRL. Este proceso involucra un mapeo de entidades ontológicas, una identificación de patrones estructurales y una ejecución de una consulta SQWRL para obtener una respuesta de un sistema de ontologías. El método propuesto se observa en la Figura 1.

3.1. Consulta semiestructurada creada en [15]

El método propuesto en este trabajo tiene como punto de partida una consulta semiestructurada, la cual es creada en [15] mediante identificación de entidades y componentes de una consulta en lenguaje natural del dominio académico con una precisión del 96 %. El método propuesto en [15] es realizado a partir de lexicones de eventos, espacios físicos y personas que pertenecen a un ambiente académico, disponibles en: <https://github.com/UamAISII/AppVoz.git>. Ellos analizan un sistema de ontologías llamado *Ambiente Inteligente* desarrollado en [16] con las ontologías Persona, Evento, Red de Sensores, Espacio Físico y Tiempo. En [15] se crea una estructura de recuperación de información y una *tupla* ontológica con cierta estructura, pero que sigue conservando elementos no estructurados como texto libre.

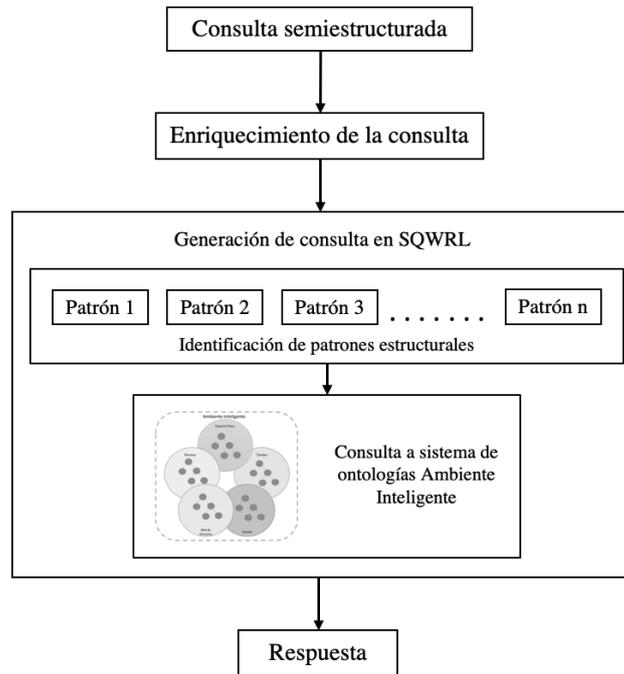


Fig. 1. Método de solución propuesto.

Esta consulta semiestructurada es mejorada en este trabajo para considerar elementos necesarios para responder a la petición del usuario.

La consulta semiestructurada obtenida consta de diversos elementos, tales como el tipo de pregunta expresada y el conjunto de entidades relevantes que contiene la petición del usuario. Las entidades relevantes son aquellas que contienen la información necesaria para responder a la consulta del usuario. Este conjunto de información es representada en una *5-tupla* elementos con la estructura mostrada en la representación mostrada (1):

$$[TP, S, TS, CS, I^*]. \quad (1)$$

Las preguntas en lenguaje natural (texto no estructurado) fueron generadas por miembros del cuerpo estudiantil de la UAM Azcapotzalco (aproximadamente 40 personas formadas por ayudantes del departamento de sistemas y estudiantes de ingeniería en computación e ingeniería industrial), quienes apoyaron mediante la expresión de diversas formas de preguntar a una interfaz de consulta, por ejemplo, la dinámica es: “Preguntar por la ubicación del auditorio W001”, un ejemplo de petición mostrado a la comunidad es: “¿Dónde esta el auditorio W001?”.

De esta manera, se generó un corpus de 300 preguntas, las cuales arrojan la misma estructura general de la representación mostrada (1), donde:

- TP es el tipo de pregunta;
- S el sujeto del que se habla en la pregunta;
- TS el tipo de sujeto, por ejemplo, el nombre de una persona, la matrícula, el nombre del lugar, entre otros;
- CS se refiere a la clase del sujeto como, profesor, edificio, auditorio, estudiante, entre otros;
- I^* es un dato que tiene una ocurrencia de 0 o más, es decir, la información extra que puede existir o no dentro de la pregunta.

Si la consulta es: “¿Cuál es la matrícula de Pedro García?”, la estructura generada siguiendo con las características de la representación mostrada (1) se muestra en la Tabla 1.

Tabla 1. Ejemplo de representación 1.

[Persona, pedro garcía, nombre, alumno, matrícula]

En la Tabla 1, *Persona* es TP , “*pedro garcía*” es S , *nombre* equivale a TS , *alumno* es CS y por último nuestra I^* es *matrícula*.

Debido a la gran variedad de preguntas generadas en el idioma español, se ha observado la necesidad de extender la representación mostrada (1) con la finalidad de obtener la mayor cantidad de entidades involucradas en la pregunta para que ayuden a encontrar una respuesta pertinente. Este proceso de enriquecimiento de la consulta semiestructurada se muestra a continuación.

3.2. Enriquecimiento de la consulta estructurada mediante mapeo ontológico

La representación mostrada (1), es mejorada en este trabajo mediante un mapeo de elementos ontológicos (relaciones, nombres de clase e instancias) obteniendo la representación mostrada (2). Esta corresponde a una *7-tupla* de elementos que cumple con todos los datos necesarios para realizar la transformación y ejecución de una consulta en SQWRL sobre el sistema de ontologías del dominio académico:

$$[O, S, P, TP, C, P_2, TP_2], \quad (2)$$

donde:

- O es la ontología u ontologías a las que debe consultarse;
- S es el sujeto que contiene la pregunta;

- P es la propiedad del sujeto, por ejemplo, si es el nombre propio de una persona deberá decir *hasName* o si habla del nombre de un auditorio dirá *hasNamePhysicalSpace*, estas propiedades pertenecen al sistema de ontologías;
- TP se refiere al tipo de propiedad *DataProperty* u *ObjectProperty* según sea el caso;
- C es la clase del sujeto dentro de la ontología que consulta y que contiene la información para responder, por ejemplo, *Professor*, *Building*;
- P_2 es la propiedad que se desea saber, por ejemplo, si queremos saber la matrícula de un alumno este dato contendrá *hasStudentID*;
- TP_2 se refiere al tipo de P_2 y puede ser *DataProperty* u *ObjectProperty*.

Continuando con el ejemplo de consulta mostrado en la sección anterior: “¿Cuál es la matrícula de Pedro García?”, la estructura generada siguiendo con las características de la representación mostrada (2) se muestra en la Tabla 2.

Tabla 2. Ejemplo de representación 2.

[[person, pedro garcía, hasName, DataProperty, RegularStudent, hasStudentID, DataProperty]]

En la Tabla 2, *person* es O , “*pedro garcía*” es S , *hasName* equivale a P , *DataProperty* es TP , *RegularStudent* es C , en el caso de *hasStudentID* que pertenece a P_2 y por último TP_2 es *DataProperty*.

Las clases y relaciones de tipo de dato (propiedades) del sistema de ontologías que pueden ocupar un lugar dentro de la representación mostrada (2) se muestran en la Tabla 3 y Tabla 4 respectivamente.

Las ontologías contienen, también, relaciones de tipo objeto o relaciones semánticas. De esta manera los elementos que conforman la 7 -*tupla* en la representación mostrada (2) puede contener una relación de este tipo que involucre a más de una ontología. La Tabla 5 muestra estas relaciones de objeto del sistema de ontologías del dominio académico.

3.3. Identificación e implementación de patrones estructurales

La representación mostrada en la sección anterior es transformada en consultas en *SQWRL* como lenguaje formal para la ejecución de la consulta sobre el conocimiento existente en el sistema de ontologías. Mediante un análisis de 300 peticiones con la información semiestructurada en la 7 -*tupla*, se identificaron tres patrones estructurales con sintaxis *SQWRL*. Los cuales se presentan a continuación.

El Patrón 1 que se centra en las preguntas relacionadas a las categorías de espacio físico y persona, tiene la siguiente sintaxis:

$$\text{Patrón 1: } ON : CL(?var_1) \wedge ON : DP(?var_1, "VS") \wedge ON : DPS(?var_1, ?res) \rightarrow sqwrl : select(?res),$$

Tabla 3. Clases del sistema de ontologías del dominio académico.

Ontología	Nombre de clase
Espacio Físico	ClassRoom
	Cubicle
	CommonArea
	Building
	ComputerLab
	AdministrativeOffice
	Auditorium
	GreenArea
PublicArea	
Tiempo	Instant
Persona	Profesor
	AssistantStudent
Evento	RegularStudent
	DiplomaCourse
	PostgraduateCourse
	UndergraduateCourse
	Colloquium
	Congress
	Demo
	DiscussionPanel
	Presentation
	Seminar
Workshop	

donde: ON es el nombre de la ontología de referencia; CL es el nombre de la clase involucrada; DP es el nombre de la relación de tipo de dato (propiedad); VS es el nombre del sujeto de la relación; DPS es la relación de tipo de dato (propiedad) de la respuesta. La relación entre las variables del patrón estructural 1 y la representación mostrada (2) pueden observarse en la Tabla 6.

El Patrón 2 es el resultado de un análisis de las variantes de preguntas relacionadas a los eventos. Para este tipo de pregunta en particular, se puede contestar sobre su descripción o el tiempo en que es realizado dicho evento. Esto se debe a la información representada en el sistema de ontologías, el cual contiene una ontología *Evento* y una ontología *Tiempo* de propósito general que ayuda a saber los datos que se refieren a cuándo ocurren los eventos académicos. Por esta razón si solo se pregunta por la descripción del evento es utilizado el Patrón 1. Sin embargo, si la pregunta se refiere al tiempo del evento, por ejemplo, “¿Cuándo es el Seminario de Física?”, se necesitan más especificaciones y como resultado de ello, se genera el Patrón estructural 2 que se muestra a continuación:

$$\begin{aligned}
 \text{Patrón 2: } & ON_1 : Event(?var_1) \wedge ON_2 : I(?var_2) \wedge On_1 : \\
 & DP(?var_1, "VS") \wedge ON_1 : DPS(?var_1, ?var_2) \wedge \\
 & ON_2 : PT_1(?var_2, ?res_1) \wedge ON_2 : PT_2(?var_2, ?res_2) \rightarrow sqwrl : \\
 & select(?res_1) \wedge sqwrl : select(?res_2),
 \end{aligned}$$

Tabla 4. Relaciones de tipo propiedad del sistema de ontologías del dominio académico.

Ontología	Nombre de propiedad
Espacio Físico	hasLevel
	hasNamePhysicalSpace
	hasDay
Tiempo	hasHour
	hasMinute
	hasMonth
Persona	hasAcademicTitle
	hasCategory
	hasEconomicNumber
	hasEmail
	hasName
	hasStudentID
Evento	hasEventName
	hasDescription

Tabla 5. Relaciones semánticas (tipo de objeto) del sistema de ontologías.

Propiedad de objeto	Dominio	Rango
isAssignedTo	Persona	Espacio Físico
hasPersonAssigned	Espacio Físico	Persona
isBesideOf	Espacio Físico	Espacio Físico
isLocatedInto	Espacio Físico	Espacio Físico
containsPhysicalSpace	Espacio Físico	Espacio Físico
participatesIn	Persona	Evento
hasPersonInvolved	Evento	Persona
hasTemporalEntity	Evento	Tiempo
happensIn	Evento	Espacio Físico
eventIsPartOf	Evento	Evento
isTemporalEntityOf	Tiempo	Evento
hasEventProgrammed	Espacio Físico	Evento

donde ON_1 es el nombre de la ontología detectada en la tupla ontológica; ON_2 se refiere al nombre de la ontología *Tiempo* (*time*); I es la clase que contiene la información del tiempo del evento (clase *Instant* de ontología *Tiempo*); PT_1 es la relación de tipo de dato (propiedad) del tiempo que se refiere al día del evento (*hasDay*); PT_2 es la relación de tipo de dato (propiedad) del tiempo que se refiere al mes del evento (*hasMonth*); DP es el nombre de la relación de tipo de dato (propiedad); VS es el nombre del sujeto de la relación; DPS es la relación de tipo de dato (propiedad) de la respuesta. La relación entre las variables del patrón estructural 2 y la 7-tupla ontológica de la representación mostrada (2) pueden observarse en la Tabla 7.

En el caso de $?var_1$ y $?var_2$ son variables dentro de la consulta. $?res_1$ y $?res_2$ son variables que contienen la respuesta final, que al preguntar por el tiempo, se debe arrojar una fecha conformada por día y mes.

Tabla 6. Relación de variables entre patrón estructural 1 y 7-tupla ontológica.

Variable de patrón estructural 1	Variable en mapeo ontológico
<i>ON</i>	<i>O</i>
<i>CL</i>	<i>C</i>
<i>DP</i>	<i>P</i>
<i>VS</i>	<i>S</i>
<i>DPS</i>	<i>P₂</i>

Tabla 7. Relación de variables entre patrón estructural 2 y 7-tupla ontológica.

Variable de patrón estructural 2	Variable en mapeo ontológico
<i>ON</i>	<i>O</i>
<i>DP</i>	<i>P</i>
<i>VS</i>	<i>S</i>
<i>DPS</i>	<i>P₂</i>

El Patrón 3 es el resultado del análisis detallado de las preguntas de tipo espacio físico-evento, espacio físico-persona y evento-persona, del cual se obtuvo el patrón estructural 3 mostrado a continuación:

$$\begin{aligned} \text{Patrón 3: } & PR_1 : ON_1(?var_1) \wedge PR_1 : DP(?var_1, "VS") \wedge PRG : \\ & DPS(?var_1, ?var_2) \wedge PR_2 : ON_2(?var_2) \\ & \wedge PR_2 : DPS(?var_2, ?res) \rightarrow sqwrl : select(?res), \end{aligned}$$

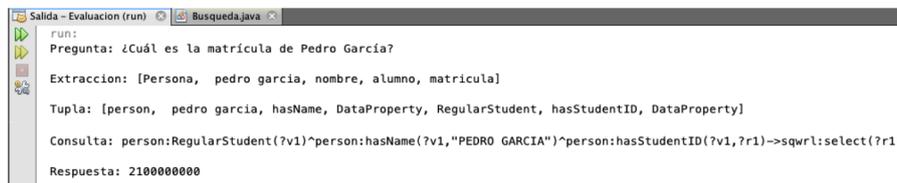
donde PR_1 es el prefijo de ontología 1 y PR_2 es el prefijo de ontología 2. El manejo de prefijos para este tipo de preguntas es relevante ya que se manejan dos ontologías y con ellos se evitan las ambigüedades en los nombres de clases y relaciones, además, en SQWRL ayuda a saber la ubicación de cada valor a preguntar dentro del sistema de ontologías; ON_1 es el nombre de la ontología 1 y ON_2 el nombre de ontología 2; DP es el nombre de la relación de tipo de dato (propiedad); VS es el nombre del sujeto de la relación; DPS es la relación de tipo de dato (propiedad) de la respuesta. La relación entre las variables del patrón estructural 3 y la 7-tupla ontológica de la representación mostrada (2) pueden observarse en la Tabla 8.

Tabla 8. Relación de variables entre patrón estructural 3 y 7-tupla ontológica.

Variable de patrón estructural 3	Variable en mapeo ontológico
PR_1	<i>O</i>
<i>DP</i>	<i>P</i>
PR_2	<i>O</i>
<i>VS</i>	<i>S</i>
ON_1	<i>O</i>
<i>DPS</i>	<i>P₂</i>
ON_2	<i>O</i>

PRG es el prefijo de sistema de ontologías (*IntelligentEnviroment*) es una parte de la consulta que habla de la unión existente entre las dos ontologías dentro del sistema. En el caso de $?var_1$ y $?var_2$ son variables dentro de la consulta, y $?res$ es la variable que contiene la respuesta final.

Un ejemplo de una consulta utilizando uno de estos patrones se muestra en la Figura 2. Donde se observa que la pregunta de entrada es “¿Cuál es la matrícula de Pedro García?”, se lleva a cabo la extracción como en la representación mostrada (1) [15], después se realiza la γ -tupla como en la representación mostrada (2). En este caso se identificó el tipo de pregunta persona por lo que en la parte de consulta se aprecia que se utilizó el patrón estructural 1. Donde *person* es *ON*, *RegularStudent* es *CL*, *hasName* es *DP*, “*PEDRO GARCIA*” es el valor de “*VS*” y *hasStudentID* es *DPS*. La variable $?res$, se encuentra en la respuesta, es decir, el valor extraído del sistema de ontologías en este caso 2100000000.



```
Salida - Evaluacion (run) Busqueda.java
run:
Pregunta: ¿Cuál es la matrícula de Pedro García?
Extracción: [Persona, pedro garcia, nombre, alumno, matricula]
Tupla: [person, pedro garcia, hasName, DataProperty, RegularStudent, hasStudentID, DataProperty]
Consulta: person:RegularStudent(?v1)^person:hasName(?v1,"PEDRO GARCIA")^person:hasStudentID(?v1,?r1)->sql:select(?r1)
Respuesta: 2100000000
```

Fig. 2. Ejemplo de consulta utilizando patrón 1.

4. Experimentación y resultados

El método propuesto se evalúa a partir de los individuos existentes en el sistema de ontologías del dominio académico obtenido de [16]. Este modelo consta de 855 individuos representados con la información adecuada para responder cada tipo de pregunta, 91 clases, 10736 axiomas, 36 relaciones de objeto y 51 relaciones de tipo de dato (propiedades). Para la experimentación se realizaron 258 preguntas que fueron evaluadas por expertos del dominio académico (profesores investigadores, estudiantes y ayudantes). El desarrollo de este método esta disponible en: <https://github.com/UamAISII/AppVoz.git>. La distribución de estas preguntas en cada tipo se observa en la Tabla 9.

Las 258 preguntas son generadas en el idioma español y obtenidas de la comunidad universitaria de la UAM Azcapotzalco, aproximadamente 70 personas entre las cuales se encuentran profesores investigadores del departamento de sistemas, alumnos de ingeniería en computación e ingeniería industrial y ayudantes de profesores, todos con una diversidad de maneras de preguntar. El corpus de cada tipo de pregunta esta disponible en: <https://github.com/UamAISII/AppVoz.git>, este es formado y evaluado de manera manual por un miembro de dicha comunidad. En la Tabla 10 se muestran los resultados de la evaluación realizada con preguntas en español.

Tabla 9. Distribución de preguntas realizadas.

Tipo de pregunta	Individuos	Preguntas realizadas
Persona	559	111
Espacio Físico	96	19
Evento	200	40
Persona - Espacio Físico	42	8
Espacio Físico - Evento	200	40
Evento - Persona	200	40
Total	855	258

Tabla 10. Valores de evaluación.

Tipo de pregunta	Respuesta correcta	Respuesta incorrecta
Persona	110	1
Espacio Físico	18	1
Evento	38	2
Persona - Espacio Físico	8	0
Espacio Físico - Evento	38	2
Evento - Persona	38	2
Total	250	8

Los resultados de la experimentación fueron evaluados utilizando la métrica, bien conocida en la Extracción de Información, denominada precisión y mostrada en la ecuación (3), por un miembro de la comunidad universitaria. Los resultados obtenidos fueron alentadores, logrando un 96.8 % de respuestas correctas y 3.2 % de respuestas incorrectas:

$$P = \frac{\text{resultadosrelevantes}}{\text{totaldeconsultas}}. \quad (3)$$

5. Conclusiones

En este artículo se ha presentado un método para realizar consultas en idioma español de México dentro de un ambiente académico. Además, se describe un enfoque basado en ontologías como modelo de representación de información, para realizar consultas en SQWRL a partir de una consulta semiestructurada utilizando patrones estructurales.

Con respecto a la ejecución de las pruebas realizadas, se obtuvieron resultados prometedores. La arquitectura propuesta conlleva las siguientes contribuciones: (a) la construcción de una *7-tupla* ontológica a partir de técnicas de procesamiento de lenguaje natural en español; (b) la identificación e implementación de tres patrones estructurales con sintaxis en el lenguaje formal SQWRL para la consulta en un sistema de ontologías del dominio académico; (c) la adaptación de un método de extracción de información basada en lexicones (conjunto de palabras relacionadas a un mismo tema) en español a una arquitectura de consultas.

Como trabajo a futuro se pretende: a) adaptar el método propuesto a un reconocimiento de voz como entrada y salida de las preguntas realizadas; b) realizar una interfaz gráfica adaptable a distintos dispositivos móviles; c) generación de una respuesta en lenguaje natural.

Referencias

1. SPARQL Tutorial, Apache Jena, Recuperado de <https://jena.apache.org/tutorials/sparql.html> (2019)
2. Cardoso, J., Lytras, D. M.: *Semantic Web engineering in the knowledge society*. IGI Global, IRM Press (2009)
3. Villatoro, C.: *Desarrollo de una Web API para el Tratamiento Automático de Reglas Contextuales Aplicadas a Servicios Basados en Localización*. Tesis de maestría, Centro Nacional de Investigación y Desarrollo Tecnológico, Departamento de Ciencias Computacionales (2011)
4. Lopez, V., Pasin, M., Motta, E.: *Aqualog: An ontology-portable question answering system for the semantic web*. In: Gómez, A., Euzenat, J. (eds) *Proceedings of the Second European Semantic Web Conference, ESWC 2005, LNCS, vol 3532*, Springer, Berlin, Heidelberg, pp. 546–562 (2005)
5. Muhammad, A., Lipika, D.: *An Ontology-Based Pattern Mining System for Extracting Information from Biological Texts*. In: Dominik, S., JingTao, Y., James, P., Wojciech, Z., Xiaohua H. (eds) *Proceedings of the 10th International Conference, RDFDGrC 2005, LNCS, vol. 3642*, pp. 420–429. Springer-Verlag Berlin Heidelberg (2005)
6. Bernstein, A., Kaufmann, E., Kaiser, C., Kiefer, C.: *Ginseng: A guided input natural language search engine for querying ontologies*. *Jena User Conference, Citeseer* (2006)
7. Tianyong, H., Dawei, H., Liu, W., Qingtian, Z.: *Semantic patterns for user-interactive question answering*. In: W. Walker, D., Chen, J. (eds) *Proceedings of the Second International Conference on Semantics, Knowledge and Grid, SKG2006 2008, LNCS, vol. 20*, pp. 783–799. John Wiley & Sons Ltd (2007)
8. Tablan, V., Damjanovic, D., Bontcheva, K.: *A Natural Language Query Interface to Structured Information*. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds) *Proceedings of The Semantic Web: Research and Applications, Proceedings of the 5th European Semantic Web Conference, ESWC 2008, LNCS, vol. 5021*, pp. 361–375. Springer, Berlin, Heidelberg (2008)
9. Damjanovic, D., Agatonovic, M., Cunningham, H.: *Natural Language Interfaces to Ontologies: Combining Syntactic Analysis and Ontology-based Lookup through the User Interaction*. In: Aroyo, L., Antoniou, G., Hyvönen, E., Teije, A., Stuckenschmidt, H., Cabral, L., Tudorache, T. (eds) *Proceedings of the 7th Extended Semantic Web Conference, ESWC 2010, LNCS, vol. 6088*, pp. 106–120. Springer, Berlin, Heidelberg (2010)
10. Eka, A., Winarko, E., Azhari, A.: *Ontology-based Why-Question Analysis Using Lexico-Syntactic Pattern*. *International Journal of Electrical and Computer Engineering (IJECE)*, 5(2), 318–332 (2015)
11. Bereta, K., Xiao, G., Koubarakis, M.: *Answering GEOSPARQL queries over relational data*. In: *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, LNCS, vol. 42*, pp. 43–50. International Society for Photogrammetry and Remote Sensing (2017)

12. Solís, A., Florencia, R., Acosta, J., López, F.: Interfaz de lenguaje natural para deducción de información almacenada en ontologías. *Research in Computing Science*, 147(6), pp. 189–205 (2018)
13. Setyo, F., Suryana, N., Sanusi, M.: New instances classification framework on Quran ontology applied to question answering system. *TELKOMINIKIA Indonesian Journal of Electrical Engineering* 17(1), pp. 139–146 (2019)
14. Mourad, S., Said, O.E.A.: SemBioNLQA: A semantic biomedical question answering system for retrieving exact and ideal answers to natural language questions. *Artificial Intelligence in Medicine*, 102, pp. 1–16 (2020)
15. García, G., Reyes, J., González, B., Priego, A.: Extracción de información a partir de preguntas en español basada en lexicones. *Academia Journals* 11(8), pp. 1054–1059 (2019)
16. Padilla, J.: Detección y representación de eventos en un ambiente académico inteligente. Tesis de Maestría, Universidad Autónoma Metropolitana, Unidad Azcapotzalco (2019)