

Reingeniería y Enriquecimiento de Ontologías Generadas Automáticamente a partir de una Base de Datos

Lozano, Adriana¹; Caliusco, Ma. Laura²; Rico, Mariela²

¹ Sectorial Informática de RRHH, Secretaría de RRHH y Función Pública, Prov. Santa Fe

² Centro de Investigación y Desarrollo de Ing. en Sistemas de Información – UTN – FRSF

Abstract

En este trabajo se presenta un método para obtener una ontología de dominio a partir de una ontología generada automáticamente a partir de una base de datos utilizando una técnica de ontology learning. La automatización de la construcción de la ontología a través de este tipo de técnicas no sólo reduce costos, sino también da por resultado una ontología que se ajusta mejor a su aplicación. Sin embargo, se genera una ontología ligera que sólo representa la estructura de la base de datos. Por lo tanto, la calidad de la ontología depende de la calidad de la base de datos. Para obtener una ontología más rica con el propósito que pueda ser utilizada para, por ejemplo, búsqueda de información, se requiere de un método de reestructuración y enriquecimiento de la misma.

Palabras Clave

ontología, ontology learning, base de datos

Introducción

Con el surgimiento de la Web Semántica [1] y las técnicas y herramientas para su realización, se han desarrollado en estos últimos años muchos trabajos basados en el uso de ontologías [2]. En ingeniería ontológica puede realizarse una distinción entre *ontologías pesadas* y *ontologías ligeras*, que no incluyen axiomas [3].

Existen diferentes metodologías para la construcción de ontologías desde cero [4]. Estas metodologías proponen actividades para cumplir con los principios de diseño, contribuyendo a la calidad de una ontología. Sin embargo, la construcción manual consume mucho tiempo, está sujeta a errores, es costosa y puede estar desviada por los intereses del desarrollador [5]. Para salvar estos problemas se han desarrollado métodos, conocidos como *ontology learning*, para la construcción en forma automática o semi-automática de ontologías a partir de fuentes de datos estructuradas,

semi-estructuradas o no-estructuradas [6].

La automatización de la construcción de la ontología no sólo reduce costos, sino también da por resultado una ontología que se ajusta mejor a su aplicación [7]. Sin embargo, las técnicas de *ontology learning* a partir de bases de datos (BD) generan ontologías ligeras que no son ricas semánticamente, ya que sólo representan la estructura de la BD. A este tipo de ontologías se las conoce como ontologías de información [8], las cuales en general no cumplen con los principios de diseño aceptados por la comunidad ontológica.

Con el objetivo de obtener ontologías más ricas a partir de ontologías de información, en este trabajo se propone un método de reingeniería y enriquecimiento. El propósito de dicho método es recibir como entrada una ontología de información y obtener como salida una ontología de dominio. Es decir, una ontología pesada que defina las conceptualizaciones específicas de un dominio dado [9]. Este método aporta guías claras de cómo llevar adelante los procesos siguiendo principios de diseño.

El trabajo se organiza de la siguiente manera. A continuación se discute la metodología empleada para obtener el método propuesto. Luego se presentan los conceptos relativos a ontologías. Después se presenta el método y un ejemplo de aplicación. Luego se discuten los resultados obtenidos y se finaliza con las conclusiones.

Elementos y Metodología de Investigación

El método de investigación utilizado para el desarrollo del método de reingeniería y enriquecimiento se conoce como Investigación en Acción [10]. Este es un

método cualitativo utilizado para la validación de trabajos de investigación mediante su aplicación en proyectos reales, reforzando la interacción entre los investigadores y los participantes de dichos proyectos. Así, se diferencia de los experimentos de laboratorio, ya que el “laboratorio” de Investigación en Acción es el mundo real, evitando la separación entre la investigación y la práctica [11]. Difiere también de la investigación de casos de estudio, ya que el investigador en acción interviene realizando cambios y estudiando el impacto de dichos cambios [12].

El método de Investigación en Acción se puede ver como un proceso cíclico con cinco etapas [13]:

- Diagnóstico: es la etapa inicial de cada ciclo; en ella se deben identificar los aspectos a mejorar y los problemas a resolver propios de cada ciclo.
- Planificación de la acción: en esta etapa se consideran los diferentes cursos de acción a tomar para resolver los problemas detectados en la etapa anterior. El resultado de esta actividad es la identificación de una serie de acciones a ejecutar sobre una determinada ontología de información.
- Ejecución de la acción: implica la implementación del curso de acción elegido en la etapa anterior sobre la ontología de información seleccionada.
- Evaluación: se debe determinar si se han alcanzado los efectos esperados de la acción, y si esos efectos mitigaron los problemas identificados previamente. Si los resultados obtenidos fueran no satisfactorios, los problemas detectados se trasladan a las siguientes iteraciones.
- Especificación del aprendizaje: esta actividad se lleva a cabo durante todo el ciclo. En ella, los conocimientos adquiridos durante el ciclo se comparten con las personas involucradas, para que dicho conocimiento pueda ser asimilado en las tareas que realizan y para que se puedan planificar nuevos ciclos.

El método propuesto en este trabajo se aplicó iterativamente a cinco ontologías de

información obtenidas a partir de siete grupos de tablas de los sistemas de sueldo y de administración de recursos humanos de la Provincia de Santa Fe. Estos grupos representan la información de: organismos, agentes, cargos, sueldos, sanciones, formación académica y lugares. A continuación se presenta el método resultante de dichas iteraciones y su aplicación a las tablas con información de los organismos de la Provincia de Santa Fe.

Definición de Ontología

Desde un punto de vista conceptual, una ontología se puede definir como $O = \{T, R, A, I\}$, donde T es un conjunto no vacío de términos, R de relaciones, A de axiomas e I de instancias [14]. $t_i \in T = [\eta, p]$; η es una cadena de caracteres que especifica el nombre del término que representa una entidad del mundo real y p es el conjunto de propiedades que caracterizan dicha entidad. Una propiedad p se puede ver como una relación cuyo dominio es un término y su rango un tipo de dato. $R = \{H, S, N\}$ donde H representa relaciones jerárquicas, S relaciones de sinonimia y N las restantes. Los axiomas son expresiones que son siempre ciertas; se pueden incluir con el propósito de definir restricciones complejas sobre los valores de los atributos o argumentos de las relaciones (axiomas de integridad) [15], verificando la corrección de la información especificada en la ontología o deduciendo nueva información. Desde un punto de vista de implementación, para que una ontología pueda ser procesable por una máquina la misma debe estar implementada en un lenguaje adecuado. Actualmente se utiliza para la implementación de ontologías para la Web Semántica la familia de lenguajes Web Ontology Language (OWL) estandarizados por la W3C [16, 17]. En estos lenguajes, los términos de una ontología se representan con el elemento `class`, las propiedades con `data properties` y las relaciones con `object properties`.

Método de Reingeniería y Enriquecimiento de una Ontología

El método propuesto recibe como entrada una ontología de información generada con alguna herramienta de *ontology learning* a partir de una BD y el Documento de Especificación de Requerimientos de la Ontología (DERO) [18]. El resultado es una ontología de dominio enriquecida.

Los roles involucrados en la ejecución del método son dos: experto del dominio e ingeniero ontológico. Cada rol puede ser llevado a cabo por una o más personas.

Proceso 1: Reingeniería de la ontología

Al proceso de reingeniería de una ontología de información lo lleva a cabo el ingeniero ontológico junto con el experto del dominio y se compone de las siguientes actividades.

Actividad 1.1: Eliminar elementos innecesarios

Las técnicas de *ontology learning* aplicadas a BD producen ontologías que modelan los esquemas de las BD consideradas. De esta forma, las ontologías resultantes pueden no representar de manera adecuada el dominio (o la porción del mundo real de interés). Además, se debe tener en cuenta el propósito para el cual se crean las ontologías. Si la BD no será actualizada a partir de la ontología, no es necesario mantener elementos en la ontología con dicho fin. Es así que, siguiendo el criterio de concisión [19], se eliminan de la ontología todos aquellos elementos creados por la herramienta de *ontology learning* que no representan entidades del mundo real de interés. Esta actividad consiste de las siguientes tareas:

Tarea 1.1.1: Eliminar propiedades innecesarias. Se eliminan las propiedades con dominio vacío.

Tarea 1.1.2: Eliminar relaciones innecesarias. Se eliminan las relaciones con dominio y rango vacío.

Tarea 1.1.3: Eliminar términos innecesarios. Se deben identificar los términos creados por la herramienta de *ontology learning* que no representan

entidades del dominio. Para ello, se deben analizar las preguntas de competencia y las entidades más relevantes del dominio definidas en el DERO. Todo aquel término que no es necesario para responder alguna pregunta de competencia o no está representando a una entidad del dominio, es candidato a ser eliminado. Luego se realizan las siguientes subtareas:

Subtarea 1.1.3.1: Eliminar propiedades asociadas. Se deben eliminar las propiedades que tengan sólo a los términos seleccionados como dominio.

Subtarea 1.1.3.2: Eliminar relaciones asociadas. Se deben eliminar las relaciones que tengan sólo a los términos seleccionados como dominio y rango.

Subtarea 1.1.3.3: Eliminar los términos seleccionados. Luego de haber eliminado todos sus elementos asociados, se eliminan los términos seleccionados.

Actividad 1.2: Identificar términos principales de la ontología

Esta actividad es importante para poder comenzar a analizar la ontología; consta de las siguientes tareas:

Tarea 1.2.1: Identificar términos principales del dominio. Identificar en la lista de entidades que se encuentra en el DERO, los términos relacionados a este dominio que las representan y tienen mayor frecuencia.

Tarea 1.2.2: Identificar los términos de la ontología que representan las entidades principales. Con la ayuda del experto del dominio se deben identificar los términos que representan los términos del dominio seleccionados en la tarea anterior.

Actividad 1.3: Analizar términos, sus propiedades y relaciones

Esta actividad se realiza en ciclos. En el primer ciclo se consideran los términos identificados en la Tarea 1.2.2. Por cada término deben realizarse las primeras tres tareas que se describen a continuación. La cuarta tarea tiene por objetivo identificar nuevos términos para comenzar otro ciclo o dar por concluida la actividad.

Tarea 1.3.1: Renombrar el término. Siguiendo el criterio de estandarización de nombres [20], el nombre de un término se debe cambiar si el que posee no es apropiado para el dominio y/o no respeta la convención de nombres. Para determinar esto se consulta al experto del dominio y el glosario de términos, si existiera. Si el término aportado por el experto no está en el glosario, se agrega. Puede asignarse más de un nombre al mismo término, cuando se lo conoce con más de un nombre.

Tarea 1.3.2: Analizar las relaciones que tienen al término considerado en su dominio. Para cada una de ellas se realizan las siguientes subtareas:

Subtarea 1.3.2.1: Analizar relaciones. Si la relación es de interés en el dominio, se deja en la ontología haciendo las modificaciones de nombre que correspondan siguiendo la convención de nombres. En caso contrario, la relación se elimina, siguiendo el criterio de concisión. También puede eliminarse para ser creada en forma correcta.

Subtarea 1.3.2.2: Analizar propiedades redundantes. Se deben eliminar las propiedades que resulten redundantes considerando las relaciones analizadas en la Subtarea 1.3.2.1. Esto puede ocurrir porque se parte de una BD donde las tablas están relacionadas por *foreign key*, y estas propiedades surgen de las columnas involucradas en esas *foreign key*; en la BD se necesitan las columnas a relacionar y la relación; en cambio en una ontología, si existe la relación no es necesaria la propiedad asociada al término origen. Si a pesar de resultar redundantes las propiedades, se determina que alguna puede originar una relación con otro término, se crea esta relación y luego se elimina la propiedad.

Tarea 1.3.3: Analizar el resto de las propiedades asociadas al término considerado. Esta tarea se lleva a cabo con la ayuda del experto del dominio y las preguntas de competencia del DERO. Se deben realizar las siguientes subtareas:

Subtarea 1.3.3.1: Transformar propiedades en relaciones. Si la propiedad representa una relación con otro término, se elimina y se crea la relación correspondiente. Esta relación puede faltar por varias razones: puede haber faltado la *foreign key* en la BD; también puede ser que al elegir las fuentes de información no se haya seleccionado alguna tabla intermedia; o bien, la herramienta de *ontology learning* no creó la relación.

Subtarea 1.3.3.2: Eliminar las propiedades innecesarias. Siguiendo los criterios de concisión y claridad [21], deben eliminarse aquellas propiedades que no son útiles para la ontología, excepto las que pudieran servir para usos futuros.

Subtarea 1.3.3.3: Controlar el rango de las propiedades. Se controla si el tipo de datos es el adecuado; si no lo es, se cambia. Si la propiedad surgió de una columna o campo codificado, cuyo significado es conocido sólo por el experto del dominio, o bien se encuentra codificado en otra fuente no accesible, se cambia el tipo de datos por un tipo enumerado con los valores posibles para esa propiedad.

Subtarea 1.3.3.4: Renombrar las propiedades. Si el nombre de la propiedad no es adecuado y/o no cumple la convención de nombres, se la renombra. Si al renombrar la propiedad, el nombre ya existe, se borra la propiedad a renombrar y se agrega el o los términos asociados a la propiedad borrada como dominio de la ya existente.

Tarea 1.3.4: Identificar términos para el próximo ciclo. El objetivo de esta tarea es identificar términos sin analizar a partir de los rangos de las relaciones analizadas en la Tarea 1.3.2 y las agregadas en la Tarea 1.3.3.1. Si todos ya se analizaron, se pasa a la Actividad 1.4; sino, se comienza un nuevo ciclo retomando la Tarea 1.3.1.

Actividad 1.4: Identificar términos sin analizar

Se deben comparar los términos de la ontología con los ya analizados en la Actividad 1.3. Si aún quedan algunos sin analizar, puede ser que no se necesitan y no

se eliminaron en la Actividad 1.1; también puede suceder que no estén relacionados con otros términos. En todos los casos, una vez identificados, se debe decidir si se eliminan o no; en este último caso se deben analizar comenzando un nuevo ciclo siguiendo las tareas a partir de la Actividad 1.3.1. Si ya se analizaron todos los términos, se pasa a la Actividad 1.5.

Actividad 1.5: Identificar relaciones sin analizar

Se debe identificar si quedan relaciones sin analizar. Esto no debería suceder ya que si todas las relaciones tienen dominio y todos los términos se analizaron, sus relaciones también se analizaron al realizar la Tarea 1.3.2; si las relaciones no tienen dominio podrían haber sido eliminadas en la Tarea 1.1.2. Si aún así quedan relaciones, se deben analizar para renombrarlas o eliminarlas realizando la Tarea 1.3.2; de lo contrario, se continúa con la Actividad 1.6.

Actividad 1.6: Identificar propiedades sin analizar

Se debe determinar si quedan propiedades sin analizar. Si esto es así, deben analizarse según la Tarea 1.3.3; sino se continúa con la Actividad 1.7.

Actividad 1.7: Eliminar anotaciones innecesarias

Se deben eliminar las anotaciones que pudiera haber creado la herramienta de *ontology learning* y que no sean necesarias, para brindar mayor claridad a la ontología.

Actividad 1.8: Reestructurar la ontología

Esta actividad modifica la estructura de la ontología con el fin de cumplir criterios de diseño como el de minimización de la distancia semántica entre términos hermanos [20]. Esta actividad está compuesta por dos tareas:

Tarea 1.8.1: Transformar términos en propiedades. Si un término no representa una entidad del dominio sino que representa una propiedad, entonces debe eliminarse y crearse la propiedad, indicando como dominio el término que representa la entidad a la cual corresponde esa propiedad.

Tarea 1.8.2: Crear o reestructurar la jerarquía de términos. Esta modificación consiste en crear una jerarquía a partir de un determinado término creando subtérminos. Puede que no sea necesario crear términos sino ordenarlos creando la clasificación; esto responde al criterio de minimización de la distancia semántica entre términos hermanos el cual indica que los términos similares se agrupan y representan como subtérminos de un término y deberían ser definidos con las mismas primitivas [20].

Proceso 2: Enriquecimiento de la ontología

Este proceso recibe como entrada la ontología obtenida en el Proceso 1 y genera como salida una ontología pesada de dominio enriquecida. El proceso de enriquecimiento lo lleva a cabo principalmente el ingeniero ontológico y consta de las actividades siguientes.

Actividad 2.1: Agregar axiomas

Esta actividad consiste en revisar cada una de las relaciones definidas en la ontología y agregar las restricciones de cardinalidad y existencialidad donde correspondiera. Además, hay que agregar los axiomas relacionales, es decir, si una relación es funcional, simétrica, reflexiva o transitiva. Por último, para contribuir a la consistencia de la ontología y considerando que la misma se implementa en lógica descriptiva [22], se debe agregar la restricción de universalidad donde correspondiera.

Actividad 2.2: Agregar términos sinónimos

En esta actividad se debe revisar el DERO y, junto con el experto del dominio, determinar si existen términos sinónimos que se deban agregar a la ontología para hacerla más rica. Esta actividad se debe llevar a cabo principalmente si el propósito de la ontología es realizar búsquedas.

Actividad 2.3: Definir nuevos términos

En esta actividad se deben revisar las preguntas de competencia definidas en el DERO y determinar si existen términos que no están representados en la ontología, que se derivan de términos sí considerados en la

ontología y que debieran representarse. Si existieran dichos términos, deben agregarse a la ontología definiéndolos con reglas condicionales *si...entonces*. Para ello se puede utilizar el lenguaje Semantic Web Rule Language (SWRL) [23].

Actividad 2.4: Hacer explícitas las disyunciones

Dos términos son disjuntos si no pueden tener instancias en común. Esta actividad consiste en analizar cada término y definir si correspondiera la disyunción con los demás términos de la ontología. Esto ayuda a comprobar la consistencia de la ontología.

Proceso 3: Verificación y validación de la ontología

El objetivo de este proceso es determinar si la ontología obtenida en el Proceso 2 cumple con los principios de diseño y los requerimientos descritos en el DERO, y responde a las preguntas de competencia [3]. De no cumplirse estas condiciones, se debe revisar lo actuado desde el Proceso 1.

Aplicación del Método a un Caso Real

El objetivo de la ontología de organismos es que la misma sea una fuente que provea información a la ciudadanía en el marco de un proyecto de Gobierno Abierto.

La ontología de información (Figura 1) que se obtuvo a partir de las tablas de

organismos fue producto de aplicar el *plugin* Datamaster 1.3.2¹ del editor de ontologías Protégé 3.4.8. Datamaster genera una ontología en el lenguaje de implementación OWL 1.0. Se utilizó la siguiente configuración: Importar tablas como clases (esto crea una clase por cada tabla y una clase para las *foreign key*), incluir el nombre de la tabla en el nombre de la columna (ya que hay columnas en diferentes tablas con el mismo nombre y esto genera inconvenientes) y no importar el contenido de las tablas.

Para iniciar con la aplicación del método, además de la ontología de información de organismos, se recibió como entrada el DERO de la misma (Tabla 1).

Para aplicar el método se utilizó el editor de ontologías Protégé 4.3 ya que el mismo se basa en el lenguaje de implementación OWL 2. Por esta razón, la descripción del método se realiza utilizando la terminología correspondiente a dicho lenguaje. Los autores de este trabajo cumplieron el rol de ingenieros ontológicos y quienes trabajan en el área de desarrollo del SARH y de los sistemas de sueldos del Gobierno de la Provincia de Santa Fe, cumplieron el rol de expertos del dominio.

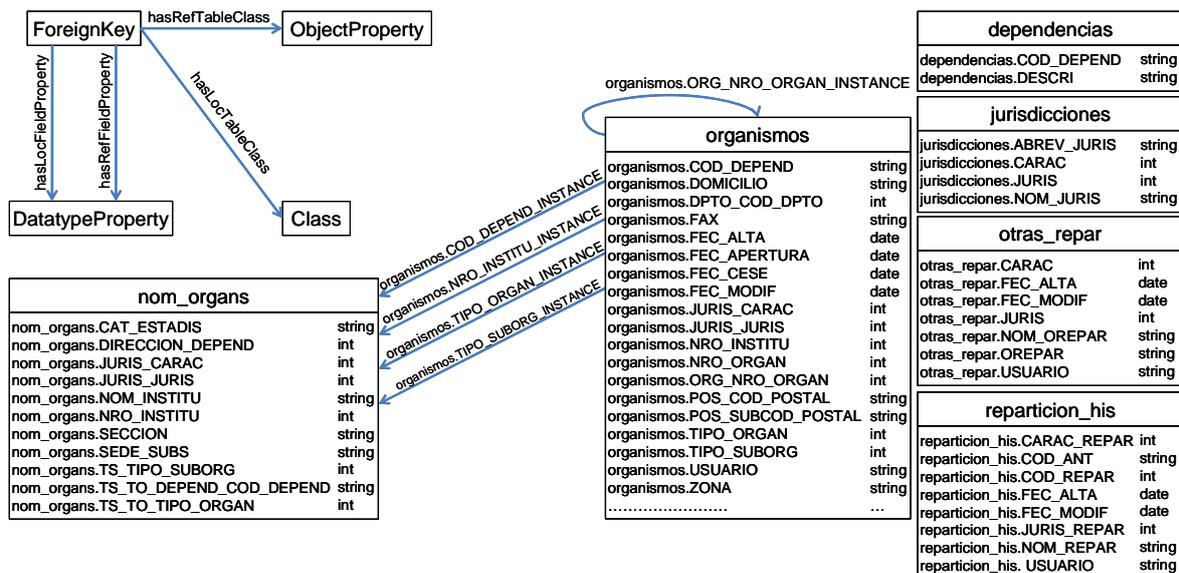


Figura 1. Ontología de información de organismos

¹ [http:// http://protegewiki.stanford.edu/wiki/DataMaster](http://http://protegewiki.stanford.edu/wiki/DataMaster)

Proceso 1: Reingeniería de la ontología

Actividad 1.1: Eliminar elementos innecesarios

Se eliminaron de la ontología todos los elementos creados por la herramienta de *ontology learning* que no representan entidades del dominio de organismos.

DERO		
Objetivo		
El objetivo de la ontología es la descripción del significado de la información relativa a los organismos del Gobierno de la Provincia de Santa Fe para mejorar la búsqueda sobre dicha información, atendiendo a la filosofía de transparencia en el Estado.		
Alcance		
La ontología tiene como alcance los datos de los organismos de la Provincia de Santa Fe que puedan ser publicados por tratarse de datos no sensibles.		
Lenguaje de Implementación		
OWL 2.0		
Usuarios finales previstos		
Usuario 1:	Un ciudadano que busca datos acerca de algún organismo de Santa Fe.	
Usuario 2:	Un agente de la Provincia de Santa Fe que necesita información acerca de algún organismo para el desarrollo de su trabajo.	
Usuario 3:	Un funcionario de la Provincia de Santa Fe que consulta, analiza, compara datos acerca de los organismos y en base a esto toma decisiones.	
Usuario 4:	Otra Administración Pública que puede necesitar información relativa a los organismos de la Provincia de Santa Fe.	
Usos previstos		
Uso 1: Publicación datos de organismos. El Gobierno publica en forma accesible datos de sus organismos. Uso 2: Búsqueda de datos de organismos. Cualquier agente o funcionario de la APP, ciudadano o empresa, busca datos relacionados a organismos de la APP.		
Requerimientos		
<i>Preguntas de Competencia</i>	<i>Resultados Esperados</i>	
¿Cuáles son los organismos pertenecientes al ministerio x?	Listado de organismos	
¿A qué ministerio pertenece el organismo x?	Ministerio de Economía	
¿A qué ámbito pertenece el organismo x?	Provincial Oficial o Provincial Privada	
¿Qué organismos se encuentran en el domicilio x?	Listado de organismos	
¿Qué organismos se encuentran en la localidad x?	Listado de organismos	
Entidades principales del dominio y su frecuencia		
Organismo 5	Ministerio 2	Ámbito 1
Domicilio 1	Localidad 1	

Tabla 1. DERO de la ontología de organismos

Tarea 1.1.1: Eliminar propiedades innecesarias. Se eliminaron las `data properties` `hasColumnName`, `hasPrimaryKeyFields` y `isBridgeTable`, ya que no tienen dominio y por lo tanto no están asociadas a ningún elemento `class`.

Tarea 1.1.2: Eliminar relaciones innecesarias. No se eliminaron `object properties`, ya que no existen de éstas con dominio y rango vacío.

Tarea 1.1.3: Eliminar términos innecesarios. Se identificaron como `classes` innecesarias en la ontología las siguientes: `Class`, `DatatypeProperty`, `ForeignKey` y `ObjectProperty`.

Subtarea 1.1.3.1: Eliminar propiedades asociadas. Se eliminaron las siguientes `data properties` que tenían como dominio la `class ForeignKey`: `hasFKName`, `hasLocalField`, `hasReferenceTable` y `hasReferenceField`.

Subtarea 1.1.3.2: Eliminar relaciones asociadas. Se eliminaron las `object properties` `hasLocFieldProperty` y `hasRefFieldProperty` que tenían como dominio la `class ForeignKey` y como rango la `class DatatypeProperty`.

Se eliminó la `object property` `hasForeignKeys`, sin dominio y con rango la `class ForeignKey`.

Se eliminó la `object property` `hasLocalTableClass` que tenía como dominio `ForeignKey` y como rango `Class`.

Se eliminó la `object property` `hasRefTableClass` que tenía como dominio la `class ForeignKey` y como rango la `class ObjectProperty`.

Subtarea 1.1.3.3: Eliminar los términos seleccionados. Se eliminaron las `classes` identificadas en la Tarea 1.1.3.

Actividad 1.2: Identificar términos principales de la ontología

Tarea 1.2.1: Identificar términos principales del dominio. El término de mayor frecuencia relacionado al dominio de esta ontología es “organismo”.

Tarea 1.2.2: Identificar los términos de la ontología que representan las entidades principales. Las `classes` `organismos`, `otras_repar` y `reparticion_his` representan el término “organismo”.

Actividad 1.3: Analizar términos, sus propiedades y relaciones

En el primer ciclo se consideraron las

classes organismos, otras_repar y reparticion_his.

Primer ciclo:

Tarea 1.3.1: Renombrar el término. Se renombraron las classes: organismos por OrganismoEnSARH, otras_repar por OtroOrganismo y reparticion_his por OrganismoAntesDelSARH.

Tarea 1.3.2: Analizar las relaciones que tienen al término considerado en su dominio.

Subtarea 1.3.2.1: Analizar relaciones. Se renombró la object property organismos.ORG_NRO_ORGAN_INSTANCE como esAnexoDe, con dominio OrganismoEnSARH y rango OrganismoEnSARH. Se borraron las siguientes object properties, con dominio OrganismoEnSARH y rango nom_organs:

organismos.COD_DEPEND_INSTANCE, organismos.NRO_INSTITU_INSTANCE, organismos.TIPO_ORGAN_INSTANCE y organismos.TIPO_SUBORG_INSTANCE. Se creó en forma correcta la object property tieneNombre, con dominio OrganismoEnSARH y rango nom_organs.

Subtarea 1.3.2.2: Analizar propiedades redundantes. Se analizaron las data properties organismos.COD_DEPEND, organismos.NRO_INSTITU, organismos.TIPO_ORGAN y organismos.TIPO_SUBORG, que surgieron de las columnas orígenes de una *foreign key*. Como organismos.COD_DEPEND representa una relación con dependencias, se creó la object property correspondeAlAmbito con dominio OrganismoEnSARH y rango dependencias, y se borró la data property organismos.COD_DEPEND. También, se eliminaron las data properties organismos.TIPO_ORGAN y organismos.TIPO_SUBORG. A pesar de existir una data property similar (nom_organs.NRO_INSTITU con dominio nom_organs), se dejó la data property organismos.NRO_INSTITU con dominio OrganismoEnSARH, porque representa un número significativo de organismos.

Tarea 1.3.3: Analizar el resto de las propiedades asociadas al término considerado.

Subtarea 1.3.3.1: Transformar propiedades en relaciones. Se eliminaron las data properties organismos.JURIS_CARAC y organismos.JURIS_JURIS y se creó la object property perteneceA, con dominio OrganismoEnSARH y rango jurisdicciones. Se eliminaron las data properties repartición_his.CARAC_REPAR y repartición_his.JURIS_REPAR. Como la object property perteneceA ya existe, se agrega a ésta como dominio OrganismoAntesDelSARH. Se eliminaron las data properties otras_repar.CARAC y otras_repar.JURIS. En este último caso no se creó una object property, a pesar de existir la *foreign key*, porque no representa la jurisdicción a la cual corresponde ese otro organismo.

Subtarea 1.3.3.2: Eliminar las propiedades innecesarias. Se eliminaron las siguientes data properties: organismos.AMBITO, organismos.ANEXO, organismos.AREA_PROG, organismos.CATEGORIA, organismos.CUE, organismos.DIRECCION_DEPEND, organismos.DOMIC_REFERENCIA, organismos.DPTO_COD_DPTO, organismos.FEC_ALTA, organismos.FEC_MODIF, organismos.LOT_ANIO, organismos.LOT_BOCA_PAGO, organismos.LOT_DEPENDENCIA, organismos.LOT_LOTE, organismos.LOT_MES, organismos.NRO_ORGAN, organismos.SEDE, organismos.TELEF_PROPIO, organismos.USUARIO, organismos.ZONA, otras_repar.FEC_ALTA, otras_repar.FEC_MODIF, otras_repar.OREPAR, otras_repar.USUARIO, repartición_his.COD_ANT, reparticion_his.COD_REPAR, repartición_his.FEC_ALTA, repartición_his.FEC_MODIF y repartición_his.USUARIO.

Subtarea 1.3.3.3: Controlar el rango de las propiedades. Las data properties

organismos.CLAUSURA y organismos.TIPO_JORNADA surgieron de columnas codificadas, cuyos significados se encuentran en la tabla RECURSOS_HUMAN_REF_CODES, no incluida en esta ontología. El experto del dominio indicó la correspondencia entre los valores y sus significados.

organismos.CLAUSURA: {"S" = Funciona; "T" = Cierre Temporal; NULL = Cerrado}.

organismos.TIPO_JORNADA: {"AA" = Doble y Simple; "CO" = Completa; "SN" = Sin Información; "NC" = No corresponde; "SI" = Simple; "DO" = Doble}.

En ambos casos, el tipo de datos es string; se borraron y se crearon los enumerados tipoFuncionamiento con los valores {"En Funcionamiento", "Cerrado Temporalmente", "Cerrado"} y valorJornada con los valores {"Doble y Simple", "Completa", "Sin Información", "No corresponde", "Simple", "Doble"}.

Subtarea 1.3.3.4: Renombrar las propiedades. Se renombran las siguientes data properties (se guarda en comment el nombre anterior): organismos.CLAUSURA por funcionamiento, organismos.DOMICILIO por domicilio, organismos.EMAIL por email, organismos.FAX por fax, organismos.FEC_APERTURA por fechaApertura, organismos.FEC_CESE por fechaCierre, organismos.FEC_IMPLEMENT por fechaImplementación, organismos.NRO_INSTITU por númeroOrganismo, organismos.REGION por región o zona, organismos.TELEFONO por teléfono, organismos.TELEF_INTERNOS por teléfonosInternos, organismos.TIPO_JORNADA por tipoJornada, organismos.POS_COD_POSTAL por códigoPostal, organismos.POS_SUBCOD_POSTAL por subCódigoPostal, otras_repar.NOM_OREPAR por nombre y repartición_his.NOM_REPAR por nombre.

Tarea 1.3.4: Identificar términos para el próximo ciclo. La class OrganismosEnSARH se relaciona con las classes nom_organs (mediante la object property

tieneNombre), dependencias (mediante la object property correspondeAlAmbito) y jurisdicciones (mediante la object property perteneceA). Se comienza el segundo ciclo con estas tres classes.

Segundo ciclo:

Tarea 1.3.1: Renombrar el término. Se renombran las classes jurisdicciones por Jurisdicción o Ministerio, dependencias por Ámbito y nom_organs por NombreOrganismo.

Tarea 1.3.2: Analizar las relaciones que tienen al término considerado en su dominio. Al no existir object properties que posean a estas classes en sus dominios, no se realizaron las subtareas 1.3.2.1 y 1.3.2.2.

Tarea 1.3.3: Analizar el resto de las propiedades asociadas al término considerado.

Subtarea 1.3.3.1: Transformar propiedades en relaciones. No existen data properties que representen relaciones con otras classes.

Subtarea 1.3.3.2: Eliminar las propiedades innecesarias. Se eliminaron las siguientes data properties: nom_organs.CAT_ESTADIS, nom_organs.DIRECCION_DEPEND, nom_organs.JURIS_CARAC, nom_organs.JURIS_JURIS, nom_organs.NRO_INSTITU, nom_organs.SECCION, nom_organs.SEDE_SUBS, nom_organs.TS_TIPO_SUBORG, nom_organs.TS_TO_DEPEND_COD_DEPEND, nom_organs.TS_TO_TIPO_ORGAN y dependencias.COD_DEPEND.

Subtarea 1.3.3.3: Controlar el rango de las propiedades. La data property jurisdicciones.CARAC surgió de una columna codificada, cuyos significados se encuentran en la tabla RECURSOS_HUMAN_REF_CODES, no incluida en esta ontología. El experto del dominio indicó la correspondencia entre los valores y sus significados.

jurisdicciones.CARAC: {1 = Administración Central, 2 = Organismos

Descentralizados, 3 = Instituciones de Seguridad Social, 9 = Instituciones fuera de la APP}.

El tipo de datos era int; se borró y se creó el enumerado valorTipoJurisdicción con los valores {"Administración Central", "Organismos Descentralizados", "Instituciones de Seguridad Social", "Instituciones fuera de la APP"}.

Subtarea 1.3.3.4: Renombrar las propiedades. Se renombraron las siguientes data properties (guardando en comment el nombre anterior): dependencias.DESCRI por nombre, jurisdicciones.ABREV_JURIS por nombreAbreviado, jurisdicciones.CARAC por tipoJurisdicción, jurisdicciones.JURIS por númeroJurisdicción, jurisdicciones.NOM_JURIS por nombre y nom_organs.NOM_INSTITU por nombre.

Tarea 1.3.4: Identificar términos para el próximo ciclo. Como ya se analizaron todas las classes, se pasó a la Actividad 1.4.

Actividad 1.4: Identificar términos sin analizar

Se controlaron las classes de la ontología con las analizadas en la Actividad 1.3 y se observó que todas se habían analizado, por lo que se siguió con la Actividad 1.5.

Actividad 1.5: Identificar relaciones sin analizar

Todas las object properties ya se han analizado. Se continúa con la Actividad 1.6.

Actividad 1.6: Identificar propiedades sin analizar

Todas las data properties se han analizado. Se continúa con la Actividad 1.7.

Actividad 1.7: Eliminar anotaciones innecesarias

Se eliminan las anotaciones creadas por DataMaster que no son necesarias. De cada class se eliminan: hasPrimaryKeyFields y isBridgeTable. De cada data property, hasOrigColumnName.

Actividad 1.8: Reestructurar la ontología

Tarea 4.8.1: Transformar términos en

propiedades. La class NombreOrganismo representa el nombre de un organismo. Una representación más adecuada es a través de una data property de la class OrganismoEnSARH. Por esto, se elimina NombreOrganismo y la object property tieneNombre (que posee a esta class como rango), y se modifica el dominio de la data property nombre, cambiando la class NombreOrganismo por OrganismoEnSARH.

Tarea 1.8.2: Crear o reestructurar la jerarquía de términos. Se creó la class Organismo, y se reubicaron las classes OrganismoEnSARH, organismoAntesSARH y OtroOrganismo como subclasses de la recién creada. Estas subclasses comparten la data property nombre, por lo tanto se modifica su dominio eliminando las classes OrganismoEnSARH, OrganismoAntesSARH y OtroOrganismo y agregando Organismo.

Proceso 2: Enriquecimiento de la ontología

Actividad 2.1: Agregar axiomas

Se revisaron todas las object properties definidas y se agregaron los axiomas relacionales correspondientes. La Figura 2 muestra dichos axiomas para la object property perteneceA. Además, se agregaron las restricciones de existencialidad y universalidad sobre las classes.

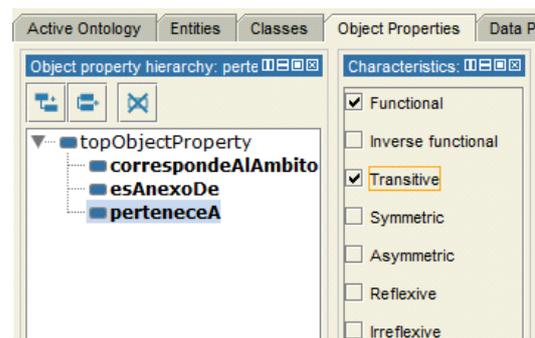


Figura 2. Definición de axiomas relacionales

Actividad 2.2: Agregar términos sinónimos

Se revisó el DERO y en conjunto con el experto del dominio se decidieron agregar como clases sinónimas de la class Organismo a Repartición Y Dependencia.

Actividad 2.3: Definir nuevos términos

Al revisar el DERO junto con el experto del dominio no se detectaron `classes` que debieran ser agregadas.

Actividad 2.4: Hacer explícitas las disyunciones

Para todas las `classes` de la ontología se definieron las disyunciones (Figura 3).



Figura 3. Disyunciones de la `class` Ministerio

La ontología resultante de aplicar el método se muestra en la Figura 4.

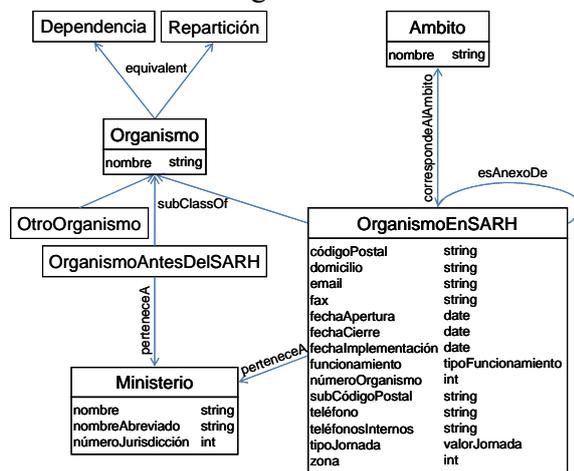


Figura 4. Ontología de dominio enriquecida de organismos

Proceso 3: Verificación y validación de la ontología

A lo largo del desarrollo de los Procesos 1 y 2 se tuvieron en cuenta la aplicación de los principios de diseño. En este proceso se corroboró su aplicación correcta y completa.

Para comprobar que la ontología respondía a las preguntas de competencia, se las implementó en SPARQL [24] y se sometieron las respuestas obtenidas a la evaluación por parte de los expertos del dominio, quienes las consideraron aceptables, dando por finalizado el desarrollo de la ontología. La Figura 5 muestra parte de la respuesta obtenida a la

pregunta: ¿Cuáles son los organismos pertenecientes al Ministerio de Economía?

subject	object
DIR_PROV_CONTROL_AGENTE_FINANCIERO	MinisterioDeEconomía
DCION_GENERAL_DE_ADMINISTRACION	MinisterioDeEconomía
CONTADURIA_GRAL_DE_LA_PROVINCIA	MinisterioDeEconomía

Figura 5. Parte de la respuesta a una pregunta de competencia

Discusión

Para construir el método de reingeniería y enriquecimiento de una ontología de información se utilizó el método Investigación en Acción. Este último permitió, a través de la participación colaborativa de los diferentes actores, aprender de las experiencias e ir mejorando el método que se estaba construyendo a lo largo de las cinco iteraciones realizadas sobre casos reales.

Si bien el método propuesto se construyó considerando como entrada una ontología cuya fuente de información es una BD, éste se podría adaptar a otras fuentes de datos semi-estructuradas o no-estructuradas.

Conclusión

Con el propósito de reducir costos y tiempos en la construcción de una ontología de un dominio dado, se utilizan técnicas de *ontology learning*. Si bien estas técnicas ayudan a construir ontologías que se adaptan a la aplicación, las ontologías resultantes son ontologías livianas y desde un punto de vista de diseño son poco ricas ya que no cumplen con criterios de diseño de una ontología. Con el propósito de obtener una ontología de dominio más rica y pesada, en este trabajo se presentó un método de reingeniería y enriquecimiento de una ontología de información. La aplicación del método se mostró en un caso real.

Agradecimientos

Los autores agradecen a la UTN y a CONICET por su apoyo financiero y a la Provincia de Santa Fe por la colaboración recibida.

Referencias

[1] Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) The Semantic Web Revisited. IEEE Intelligent Systems, 21(3), pp. 96-101.

- [2] Staab, S. and Rudi, S. (Eds.) (2009) Handbook on Ontologies. Series: International Handbooks on Information Systems. 2nd ed., Springer.
- [3] Gómez-Pérez, A., Fernández-López, M. and Corcho, O. (2004) Ontological Engineering. Springer/Heidelberg.
- [4] Suárez-Figueroa, M.C.; Gómez-Pérez, A.; Motta, E.; Gangemi, A. (Eds.) (2012) Ontology Engineering in a Networked World. Springer-Verlag Berlin Heidelberg.
- [5] Hazman, M., El-Beltagy, S. and Rafea, A. (2011) A Survey of Ontology Learning Approaches. International Journal of Computer Applications, 22(8), pp. 36-43.
- [6] Maedche, A. and Staab, S. (2001) Ontology Learning for the Semantic Web. IEEE Intelligent Systems 16 (2), pp. 72-79.
- [7] Sabou, M., Wroe, C., Goble, C., and Mishne, G. (2005) Learning Domain Ontologies for Web Service Descriptions: an Experiment in Bioinformatics. In: Proc. of the 14th International World Wide Web Conference (WWW2005), Chiba, Japan.
- [8] Van Heijst, G., Schreiber, A.T., Wielinga, B.J. (1997) Using Explicit Ontologies in KBS Development. International Journal of Human and Computer Studies, 46(2/3), pp. 293-310.
- [9] Fensel, D (2004) Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. Second Edition, Springer-Verlag, Berlin, Heidelberg.
- [10] Avison, D., Lan, F., Myers, M. y Nielsen, A. Action Research. Communications of the ACM, 42(1), pp. 94-97, 1999.
- [11] Susman, G.I. and Evered, R.D. (1978), An assessment of the scientific merits of action research, Administrative Science Quarterly, Vol. 23, pp. 582-603.
- [12] Baburoglu, Oguz N. and Ravn, Ib "Normative Action Research." Organization Studies 13(1), 1992, pp. 19-34.
- [13] Susman, G.I. and Evered, R.D. (1978), An assessment of the scientific merits of action research, Administrative Science Quarterly, Vol. 23, pp. 582-603.
- [14] Rico, M. (2011) Soporte para Enriquecer la Representación de Entidades en una Ontología. Tesis Doctoral. UTN - Facultad Regional Santa Fe.
- [15] Wagner, G. (2005) Rule Modeling and Markup, in: N. Eisinger, J. Maluszyński (Eds.), Reasoning Web, Springer Berlin Heidelberg, pp. 251-274.
- [16] McGuinness, D.L. and van Harmelen, F. (2004), OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004, <http://www.w3.org/TR/owl-features/>
- [17] W3C OWL Working Group (2012), OWL 2 Web Ontology Language, Document Overview (Second Edition), W3C Recommendation 11 December 2012, <http://www.w3.org/TR/owl2-overview/>
- [18] Suárez-Figueroa, M. C., Gómez-Pérez, A. and Villazón-Terrazas, B. (2009) How to write and use the Ontology Requirements Specification Document. In: The 8th International Conference on Ontologies, DataBases, and Applications of Semantics, ODBASE 2009, Vilamoura, Algarve-Portugal.
- [19] Gómez-Pérez, A. (1999) Evaluation of taxonomic knowledge in ontologies and knowledge bases. En: B.R. Gaines, R. Kremer and M.A. Musen (Eds.), Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management, KAW'99. Banff, Alberta, Canada, October 16-21, volumen 2, pp. 6.8.1-6.8.18. SRDG Publications, University of Calgary.
- [20] Arpírez Vega, J.C., Gómez-Pérez, A., Lozano Telo, A. and Andrade N.P. Pinto, H.S. (1998) ONTO²Agent: An ontology-based WWW broker to select ontologies. En A. Gómez-Pérez and V.R. Benjamins (Eds.), Proceedings of the Workshop on Applications of Ontologies and Problem-Solving Methods, held in conjunction with ECAI'98. Brighton, UK, August 24-25, pp. 16-24.
- [21] Grüber, T (1995). Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43(5-6), pp. 907-928.
- [22] Nardi, D., Brachman, R. J., Baader, F., Nutt, W., Donini, F. M., Sattler, U., Calvanese, D., Mlitor, R., Giacomo, G. D., Ksters, R., Wolter, F., McGuinness, D. L., Patel-Schneider, P. F., Miller, R., Haarslev, V., Horrocks, I., Borgida, A., Welty, C., Rector, A., Franconi, E., Lenzerini, M., & Rosati, R. (2003). The Description Logic Handbook - Theory, Implementation and Applications. Cambridge University Press.
- [23] O'Connor, M. Knublauch, H., Tu, S. & Musen M.: Writing rules for the semantic web using SWRL and Jess. In: Proceedings 8th International Protégé Conference, Protégé with rules Workshop, 2005.
- [24] Harris, S. & Seaborne, A. (2013), SPARQL 1.1 query language, W3C Recommendation 21 March 2013, <http://www.w3.org/TR/2013/REC-sparql11-query-20130321/>

Datos de Contacto:

Ing. Adriana Lozano. Sectorial de Informática de RRHH, Secretaría de RRHH y Función Pública. Arturo Illia 1151, S3000DWC Santa Fe. alozano@santafe.gov.ar