

Diseño de una ontología para la gestión de datos heterogéneos en universidades: marco metodológico

Design of an ontology for the management of heterogeneous data at the universities: methodological framework

Yorbelis Rosell León,^I José Antonio Senso Ruiz,^{II} Amed Abel Leiva Mederos^{III}

^I Universidad de La Habana. La Habana, Cuba.

^{II} Universidad de Granada, España.

^{III} Universidad Central de Las Villas "Martha Abreu". Villa Clara, Cuba.

RESUMEN

Una ontología constituye una herramienta útil para trabajar la heterogeneidad de los datos y su semantización, para elevar la calidad en los procesos de organización, búsqueda y recuperación de la información en los sistemas de gestión institucionales, especialmente de aquellos implementados en la Web de una universidad. Por lo general, las universidades cuentan con un acervo de conocimientos pluridisciplinarios, registrados en diferentes formatos, por lo que se propone UH-Ontology para el manejo de sus datos, tomando a la Universidad de La Habana como ejemplo de su implementación. En el artículo se presentan las características y condiciones de los datos heterogéneos que actualmente se gestionan desde los diferentes sistemas de gestión de información en la universidad, lo que ha motivado la presentación de una propuesta de marco metodológico para el diseño de una ontología para el manejo de datos heterogéneos en esta institución. Se presenta la propuesta de diseño de la ontología, sus clases, anotaciones, lenguajes ontológicos y esquema de anotación semántica siguiendo la metodología de Noy and McGuinness y Protegè como herramienta para su construcción.

Palabras clave: Universidad; Ontología; lenguajes ontológicos; datos heterogéneos.

ABSTRACT

An ontology is a useful tool to deal with the heterogeneity of data and their semantization, with a view to raising the quality of processes of organization, search and retrieval of information in institutional management systems, particularly those implemented in a university web. In view of the wealth of multidisciplinary knowledge treasured by the Universities, a UH Ontology is proposed for the management of data. The paper describes the characteristics and conditions of the heterogeneous data currently managed by various information management systems at the Universities, and proposes a methodological framework for the design of an ontology for the management of heterogeneous data at the institution. The proposal includes the design of the ontology, its classes, annotations, ontological languages and semantic annotation scheme, based on the methodology developed by Noy and McGuinness and the software Protégé as a construction tool.

Key words: University; Ontology; ontological languages; heterogeneous data.

INTRODUCCIÓN

Las ontologías son un tema de investigación en varias comunidades y áreas de estudio: la ingeniería de software, las matemáticas, la informática y más recientemente en el campo de las Ciencias de la Información, como herramientas en la representación de información (como esquema conceptual), en búsqueda y recuperación (como herramienta), como sistemas de información cooperativos y su aplicación a bibliotecas digitales y herramientas para la gestión del conocimiento. Generalmente, las universidades cuentan con un acervo de conocimientos multifacéticos. Paradójicamente, este logro genera dificultades para la gestión de los datos, debido a que la amplia gama de formatos, genera un desbalance de su tratamiento. Una ontología puede ser la herramienta que trabaje sobre la heterogeneidad de los datos y su posterior semantización, de tal forma que eleve la calidad en los procesos de organización, búsqueda y recuperación de la información en los sistemas de gestión institucionales, especialmente de aquellos implementados en la Web, y facilite además la interoperabilidad entre los sitios web.

En el presente trabajo se propone la descripción del marco metodológico para el diseño de una ontología para la gestión de datos heterogéneos en universidades, tomando como marco lógico a la Universidad de La Habana (dado su liderazgo académico en la formación académica cubana y su prestigio nacional e internacional).

MÉTODOS DE INVESTIGACIÓN

Los métodos utilizados en la investigación deben estar en correspondencia con el objeto de estudio. A continuación se presentan de manera resumida la metodología y herramienta empleada en la construcción de la ontología.

METODOLOGÍA APLICADA

El desarrollo de la ontología parte de la metodología expuesta en "Desarrollo de Ontologías-101: guía para crear tu primera ontología"¹ por la simplicidad, claridad y objetividad de los pasos establecidos. Consta de 7 pasos:

- 1) Determinar el dominio y el alcance de la ontología.
- 2) Reutilizar ontologías existentes.
- 3) Enumerar términos importantes.
- 4) Definir las clases y la jerarquía de clases.
- 5) Definir propiedades de las clases: SLOTS.
- 6) Definir facetas de los slots.
- 7) Crear instancias.

Herramientas empleadas

En la propuesta se hizo uso de Protegè 4.3 para la modelación e implementación de la ontología. Protegè es un software; un editor de ontologías de código abierto creado por la Universidad de Stanford. Como herramienta para crear ontologías en un formato estandarizado, como OWL o RDF, permite compartir y utilizar otras aplicaciones y plataformas.

MARCO TEÓRICO: DEFINICIÓN DE ONTOLOGÍA

El término ontología proviene de la filosofía. La definición declarativa más consolidada, a decir de *Lozano*² es la propuesta por *Gruber*^{3,4} y extendida por *Studer*⁵ que la describe como una especificación formal y explícita sobre una conceptualización compartida. Se asume la ontología como una descripción de los conceptos y relaciones en un dominio de aplicación, descrito en un lenguaje equipado con una semántica formal compartida y consensuada, legible y utilizable por los ordenadores. Para lograr sus objetivos, las ontologías requieren de varios componentes que permiten representar el conocimiento de algún dominio:⁴

- *Conceptos*: son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento.
- *Relaciones*: representan la interacción entre los conceptos del dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a.
- *Funciones*: son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como categorizar-clase, asignar fecha.
- *Instancias*: se utilizan para representar objetos determinados de un concepto, o ejemplos de este.
- *Axiomas*: son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología.

Las ontologías han sido consideradas la columna vertebral de la Web Semántica. La capa de la ontología "describe e identifica lo que nuestros conceptos significan para que podamos utilizar los metadatos para crear servicios de integración, federar, compartir y procesar los datos. La ontología da un entendimiento compartido y que permite la reutilización de conocimiento del dominio".⁶ Para las organizaciones, una ontología es útil para visualizar y comunicar el dominio. Esto hace que sea más fácil para los administradores, editores, diseñadores, de un sistema de información implementado en la Web, asegurar que la estructura de un sitio Web sea estable, y que permita modificar el modelo sin cambiar la base de datos o volver a escribir ningún código, con total independencia en el manejo del diseño, el contenido y la semántica.

Aunque no todo el mundo tiene que compartirlas, para las organizaciones una ontología es útil para visualizar y comunicar el dominio. Esto hace que sea más fácil para los administradores, editores, diseñadores, etc., de un sistema de información implementado en la Web, asegurar que la estructura de un sitio Web sea estable, y que permita modificar el modelo sin cambiar la base de datos o volver a escribir ningún código, con total independencia en el manejo del diseño, el contenido y la semántica.

MARCO LÓGICO: UNIVERSIDAD DE LA HABANA

La Universidad de La Habana (UH) forma profesionales en 32 carreras universitarias de ciencias naturales y exactas, sociales y económicas. El campus universitario capitalino está desplegado por toda la ciudad. Posee una estructura compleja, tanto desde el punto de vista estructural, como temático y profesional.⁷

Las intranets se caracterizan por el uso de un gran número de documentos en versión digital, de calidad variable y temática multidisciplinar. La naturaleza de los documentos no siempre es textual. Hay gran cantidad de imágenes, y en menor proporción videos y documentos de audio. El sistema tiene fallos relevantes, especialmente con los documentos no textuales, pues la indización se basa en la indización sintáctica de palabras que forman parte del texto del documento y no existe un mecanismo alternativo en la UH de hoy. El grado de interoperabilidad entre las intranet es escaso, lo cual provoca que el nivel de reutilización de la información es casi nulo.

Aunque con el diseño, implementación y uso de los diferentes sistemas la Universidad ha ganado mucho en la gestión de los procesos administrativos*, la información aún se encuentra fragmentada y no hay manera de recuperarla o visualizarla de manera integrada y coherente. Hay un extenso volumen de documentos en lenguaje natural expresado en lenguaje HTML y con escasa semántica, que provoca que las aplicaciones informáticas no siempre accedan de manera correcta a los contenidos, pues la mayor parte de estos documentos carecen de estructura, con excepción de elementos como título o etiquetas.

Puede valorarse que existe falta de normalización en las estructuras y los documentos, problemas de polisemia y sinonimia heredados del lenguaje natural. Por otra parte, se reconoce que en la UH existen otros sistemas que no se limitan a la gestión de información administrativa, sino que van hacia la gestión del conocimiento gris, como tesis, investigaciones de todo tipo y a diversas escalas y alcances. Para estos casos, un esquema de construcción de sitios web y parámetros basados en la indización sintáctica resulta insuficiente. Para conseguir mayores funcionalidades se necesitaría crear un esquema que permita el procesamiento del lenguaje natural de manera que sea entendido por el sistema en un entorno tan

heterogéneo como la UH con una adecuada y coherente estructuración de los documentos y desambiguación semántica, utilización de estadísticas y minería de datos, creación de sistemas de organización del conocimiento. Todo esto unido a la visión de crear y mejorar continuamente la interoperabilidad entre bases de datos o para los mecanismos de procesamiento por aplicaciones informáticas. Esto hace que se pueda especular sobre la factibilidad de proponer una ontología como herramienta base para la gestión del conocimiento en la UH, de manera que su dominio pueda intervenir en la construcción y comprensión semántica de la información y cuya estructura pueda ser entendida tanto por humanos como por los agentes de software.

PROPUESTA DE LA ONTOLOGÍA: UH-ONTOLOGY

UH-Ontology ha sido pensada como herramienta de referencia en la construcción de la base de conocimiento, contribuir y aportar consistencia, fiabilidad, para combatir el exceso de ambigüedad en la recuperación de información en los sitios Web. La ontología va a determinar la estructura de la información; la arquitectura de los objetos, sus dependencias y su representación gráfica; incluyendo el almacenamiento de información en una base de datos semántica en forma persistente.

Las funciones fundamentales de la ontología que se propone son las siguientes:

- Normalizar los atributos de los metadatos aplicables a los documentos.
- Crear una red de relaciones que aporte especificación y fiabilidad.
- Compartir conocimiento en pos de hacer explícitos diferentes criterios, así como la integración de diferentes perspectivas de los usuarios.
- Posibilitar el trabajo cooperativo al funcionar como soporte común de conocimiento entre las distintas áreas y estructuras.
- Tratamiento ponderado del conocimiento para recuperar información de forma automatizada.
- La reutilización del conocimiento existente en nuevos sistemas.
- La interoperabilidad entre los distintos sistemas existentes.
- Establecer modelos normativos que permitan la creación de la semántica de un sistema y un modelo para poder extenderlo y transformarlo entre diferentes contextos.

La ontología UH-Ontology responderá a:

- *Preguntas sobre el dominio:* permitirá modelar (representar) los conceptos manejados con respecto a personas, documentos y espacios geográficos.
- *Preguntas sobre el uso:* esta ontología facilitará la búsqueda de recursos de información.
- *Uso:* los usuarios potenciales de la ontología son los estudiantes, profesores, personal administrativo y directivos.
- *Mantenimiento:* los administradores de la ontología serán las personas encargadas de su mantenimiento. Se prevé que potencialmente sean especialistas de la Dirección de Informatización, la Dirección de Información y el personal colaborador de la Facultad de Matemática y Computación de la propia Universidad de La Habana.

Al finalizar la implementación de UH-Ontology, será capaz de dar respuestas a preguntas de competencia, que pueden servir como base para la elaboración de un test de calidad, como por ejemplo:

- ¿Cuál es el nombre del Decano de la Facultad X?
- ¿Cuál es el correo electrónico del administrador del sitio de la facultad X?
- ¿Quiénes son los miembros del Consejo Científico de la facultad X?
- ¿En qué lugar se encuentra el centro de estudio X de la Facultad Y?
- ¿Cuál es la temática del documental X dirigido por Y?
- Obtener listado de participantes del evento X auspiciado por el organismo Y.
- ¿Cuáles son los cursos de posgrado aprobados en el año X por la persona Y?
- Obtener las temáticas en las que investiga el profesor X.

REUTILIZACIÓN DE ONTOLOGÍAS

Se toma como base para el desarrollo de UH-Ontology la ontología VIVO. Dentro de esta última se crearon clases en función de las particularidades de la UH. Entre ellas Learning, tomada de la IEEE. En el proyecto UH-Ontology se han reutilizado:

- SKOS,⁸ con funciones de tesoro.
- FOAF,⁹ que sirve de herramienta de control de autoridades.
- BIBO,¹⁰ el cual se utiliza para describir objetos bibliográficos en la web semántica en RDF. Esta ontología se puede utilizar como una ontología citación, como una clasificación ontología documento, o simplemente como una forma de describir cualquier tipo de documento en RDF.
- VIVO,¹¹ que es una herramienta para la representación de información sobre la investigación y los investigadores (sus trabajos académicos, líneas de investigación y las relaciones de organización).
- Geontology, utilizados en la gestión de datos geográficos.

Definición de las clases

Las clases principales establecidas para UH-Ontology son: Agent, Collection, Document, Resource, Concept, Credential, EducationalTraining, Event, Project, Area, Seq y Software.

Las clases y subclases que componen la ontología serán abordadas a continuación:

Agent

La clase Agent se utiliza para describir cualquier "agente" en relación con los artículos bibliográficos. Dichos agentes pueden ser personas, organizaciones o grupos de cualquier tipo. Por tanto, se le definen dos subclases: Organization (dedicada a la identificación de instituciones, organizaciones o grupos, sus diferentes estructuras y sus subordinaciones) y Person (centrada en establecer los

roles o funciones que puedan tener las personas en su relación con la UH y los diferentes recursos de información gestionados). Además, se establecen relaciones entre Agent como rango y otras clases y subclases como dominio: con la clase Event (relación organizer) y su subclase Performance (relación performer) y con AudioVisualDocument (relación director) y PersonalCommunicationDocument (relación recipient), ambas subordinadas a la clase Document. Se establecen relaciones dentro de la misma clase Agent de tipo interviewee e interviewer (Fig. 1).

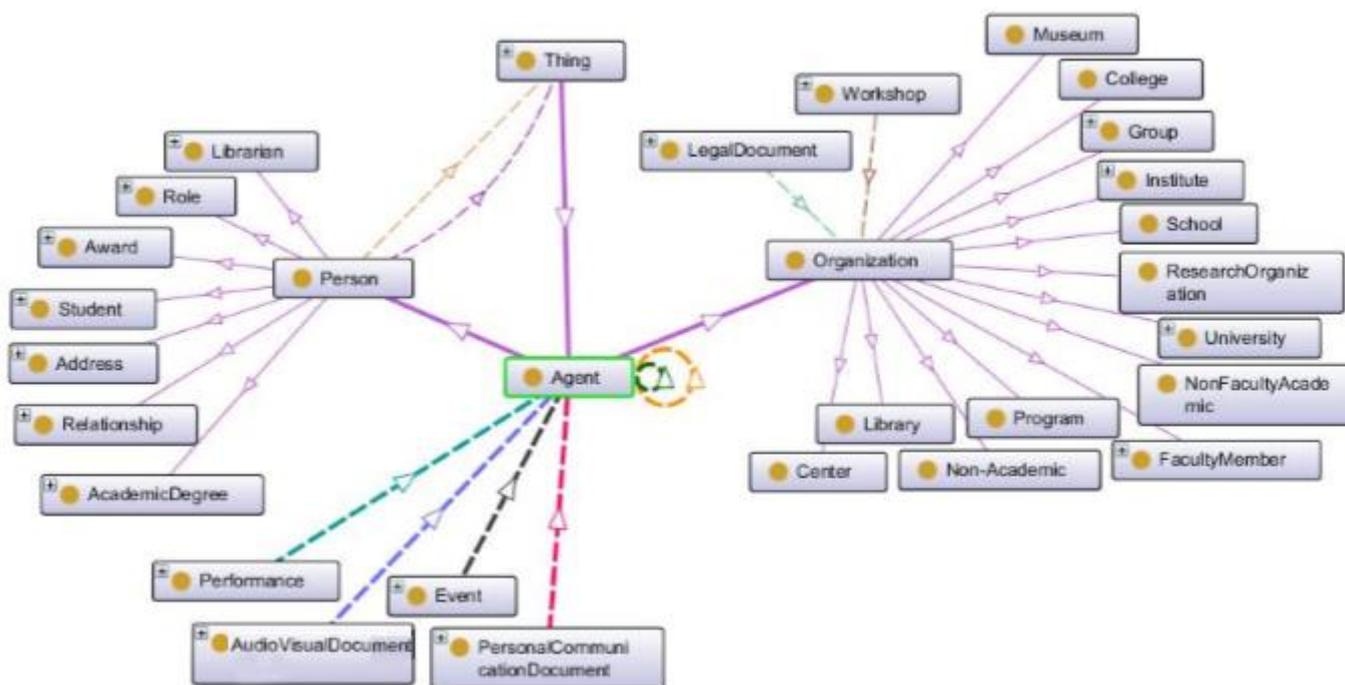


Fig. 1. UH ontology: Clase Agente y su relación con otras clases y subclases de la ontología. Visualización desde Protegé.

En la figura 1 se aprecia que dentro de la clase Agent se encuentra la subclase Organization, la cual permite presentar en su jerarquía de clases la complejidad organizacional de la UH, dada la variedad de estructuras representadas en la ontología. Además de las relaciones de subordinación y los enlaces mediante las líneas continuas, la subclase Organization establece otros vínculos (como dominio) con las subclases LegalDocument (court) de la clase Documenty con Workshop (assignedby) de la clase Event.

La subclase Person, igualmente contenida en la clase Agent, presenta en su jerarquía de clases la estructura y escalabilidad necesaria para describir al usuario desde su posición de trabajador y/o estudiante. Además de las relaciones de subordinación de la subclase Person (enlaces mediante las líneas continuas), se establecen otros vínculos como dominio con Thingen calidad de rango, para las relaciones IsReferencedBy y hasPrincipalInvestigatorRole. Collection, Document, Resource, InformationResource.

Se definen a sí mismas de manera independiente, pero siempre bajo el principio de declarar su similitud conceptual. Dadas las implicaciones de los términos en el ámbito de la Bibliotecología y la Documentación en el reconocimiento y tratamiento de la información, se realiza la declaración de subclases para las clases de Collection y Document. Los nexos conceptuales establecidos para Document, Resource, Collection e InformationResource son claramente perceptibles a través de la figura 2.

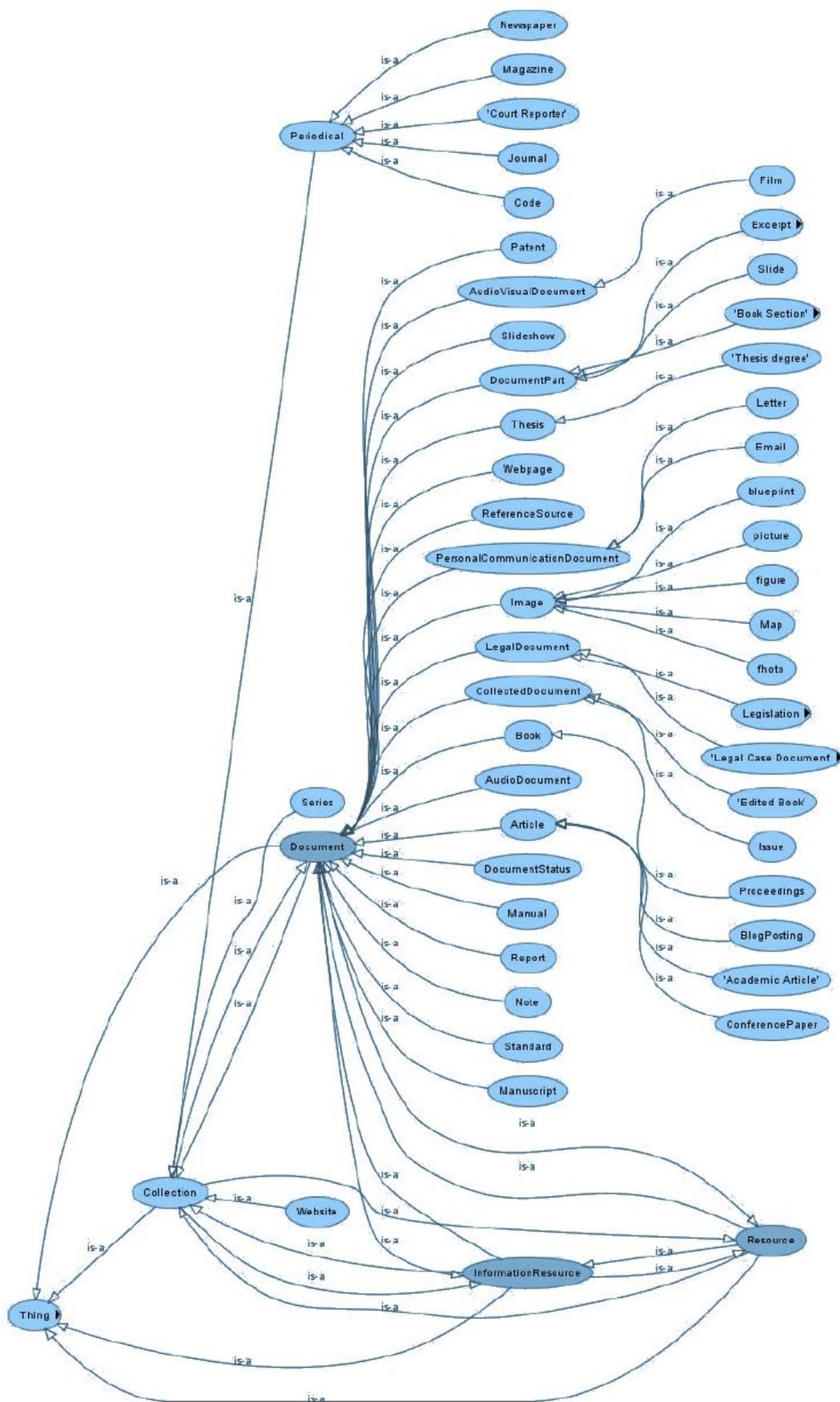


Fig. 2. UH ontology: relaciones entre las clases Document, Resource, Collection e Information Resource. Visualización desde Protegé.

La clase Document presenta varias subclases que, a su vez, determinan otras relaciones subordinadas (Fig. 3). Además de las relaciones de subordinación de la clase Document (Fig. 4), se establecen otros vínculos en calidad de rango y dominio sobre ella misma para las relaciones de citedBy, reproducedIn, translation on y cites (representadas en la figura 4 en forma circular), con Thingen calidad de dominio para las relaciones "presented at" e IsReferencedBy y en calidad de rango para las relaciones: isVersionOf, "list of authors", "list of contributors", "presented at" y "list of editors". Se establecen relaciones de tipo "review of" y "transcript of" con la clase Resource (esta última en calidad de rango). CollectedDocument y Series se definen como hasPart (para toda subclase de documento).

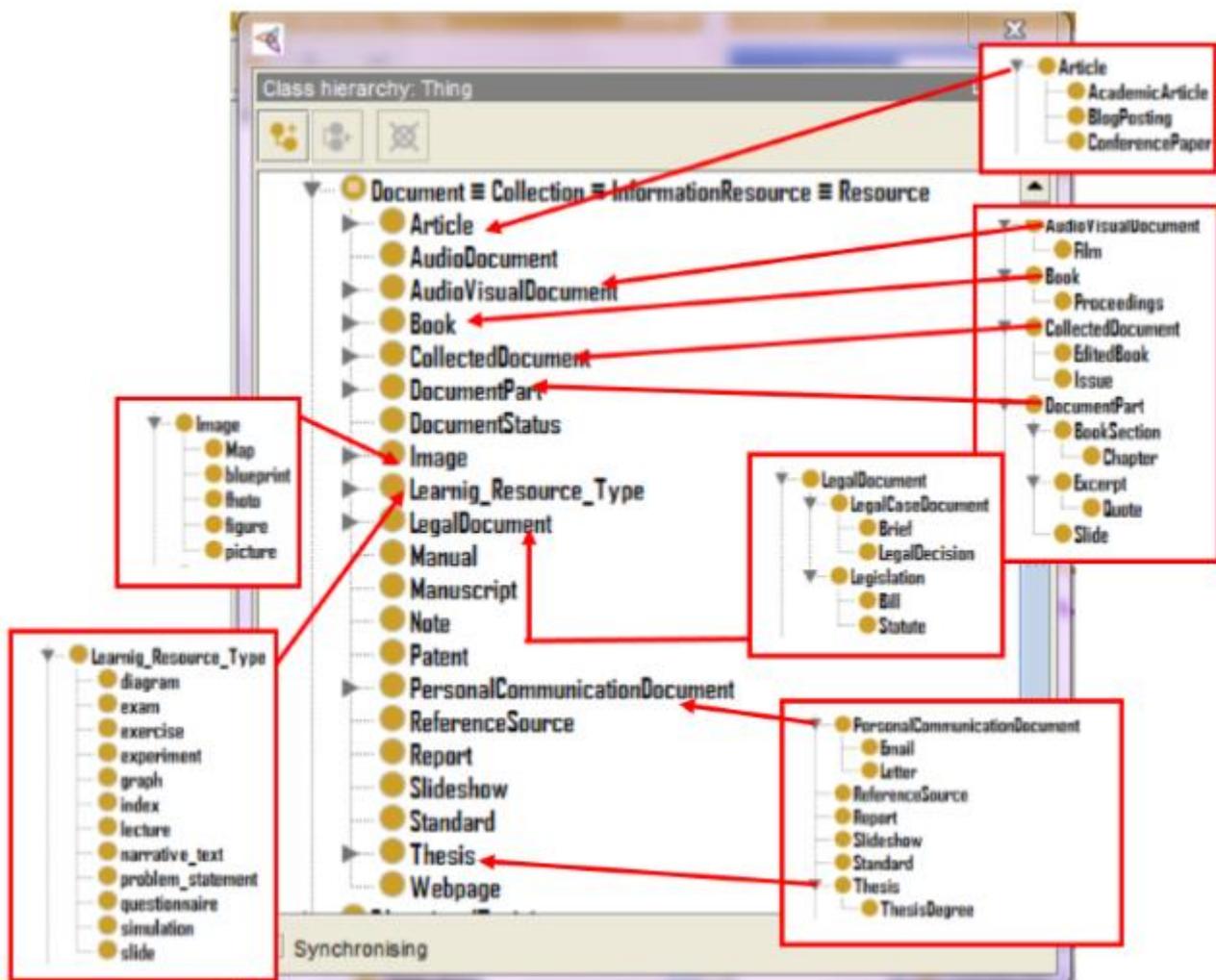


Fig. 3. UH ontology: Subclases definidas dentro de Document. Visualización desde Protegé.

Concept

Clase asumida como unidad básica de la forma de conocimiento humana. Solo tiene relación de subordinación con Thing, al ser definida como clase principal de la ontología.

Credential

Definida para el registro de certificados y números de licencia. Clase vinculada a personas, organizaciones o grupos como propietarios, responsables o beneficiarios de esta. Contiene las subclases Certificate y License.

EducationalTraining

Clase referida a los estudios universitarios posteriores al título de grado, que comprende los estudios de maestría, doctorado, postdoctorado y cursos de especialización dentro de este ámbito académico. Tienen como antecedente obligatorio los estudios de pregrado, y solo se puede acceder a ellos tras la obtención del grado.

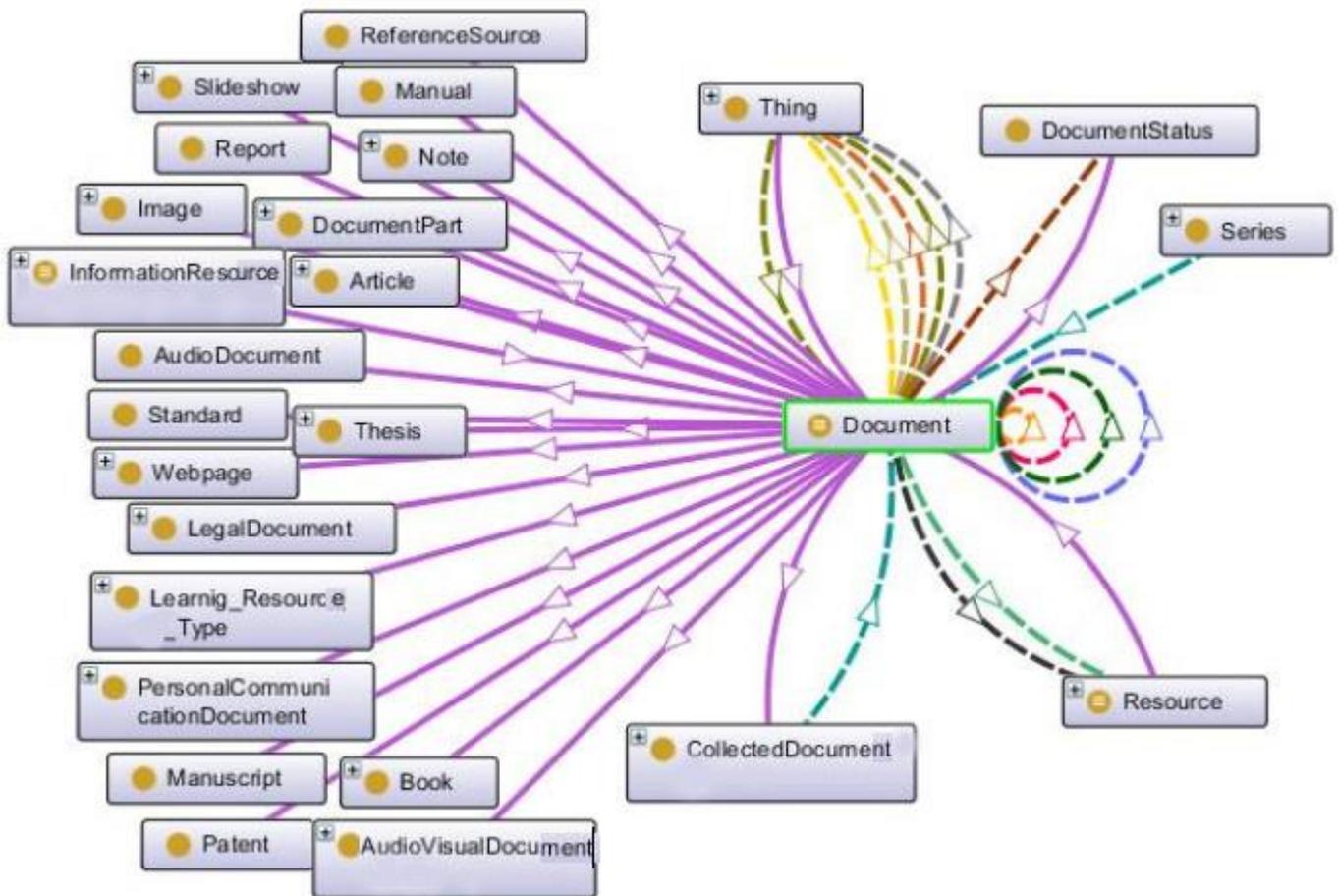


Fig. 4. UH ontology: Relaciones de la clase Documento. Visualización desde Protegé.

Event

Clase definida para identificar los diferentes eventos o actividades planificadas de manera periódica y que convoca a gran número de personas u organizaciones a los que pueda estar vinculada la UH como anfitrión o participante. Event contiene varias subclases (Fig. 5), enlaces con líneas continuas). Establece relación de tipo organizer en calidad de dominio con la clase Agent y en calidad de rango con el núcleo de la ontología Thing de tipo hasProceedings.

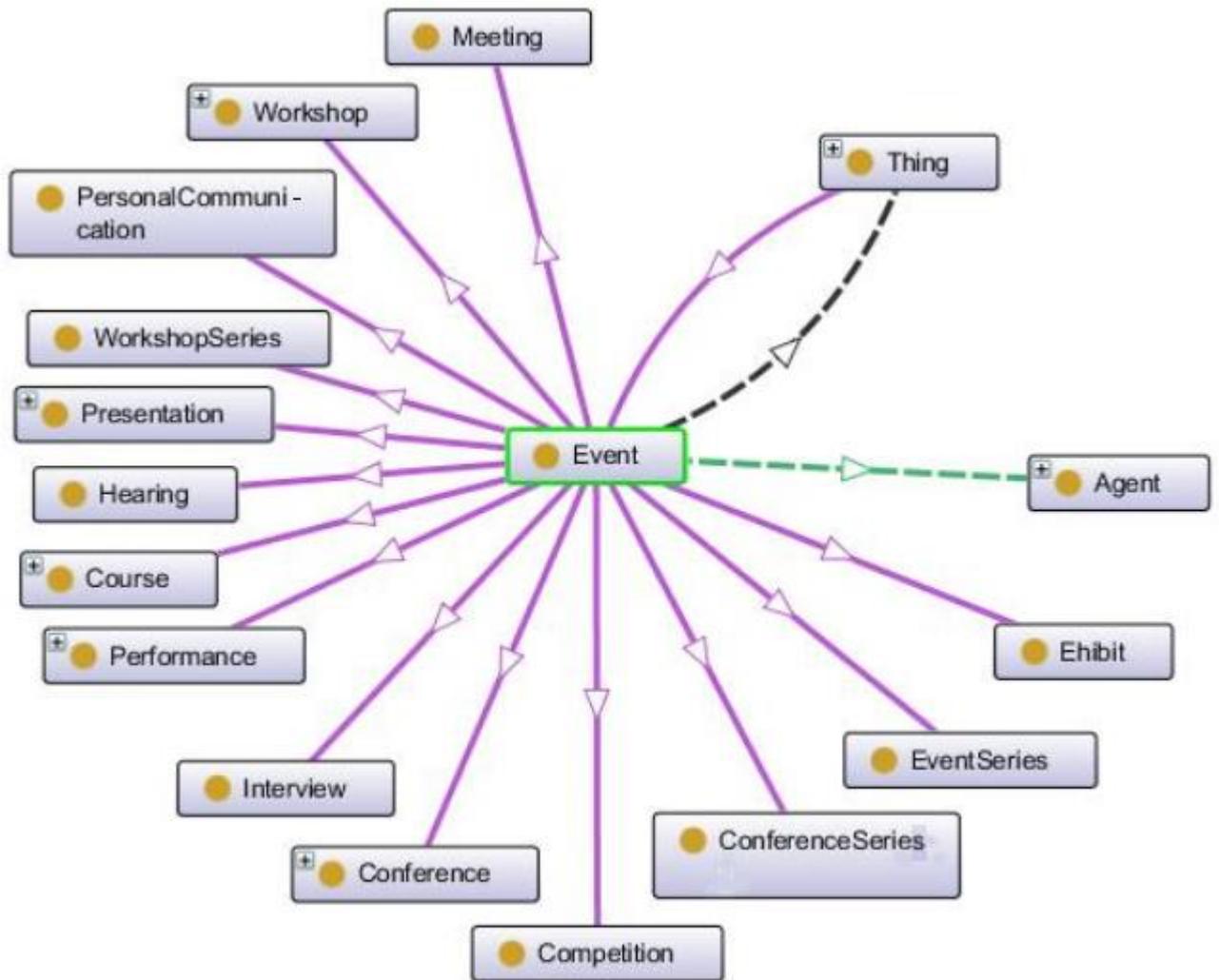


Fig. 5. UH-Ontology: Clase Evento. Visualización desde Protegè.

Project

Clase definida para identificar el desarrollo de proyectos vinculados a la UH. Solo tiene relación de subordinación con Thing, al ser definida como clase principal de la ontología.

Software

Clase definida para identificar el desarrollo y uso de softwares vinculados a la UH. Solo tiene relación de subordinación con Thing, al ser definida como clase principal de la ontología.

Area

Clase definida para identificar áreas geográficas. Contiene a las subclases: territory y group.

Propiedades de objetos

Las propiedades OWL representan relaciones entre dos objetos o individuos y, aunque están pensadas para gestionar las propiedades de los objetos, favorecen el establecimiento de relaciones entre un individuo y otro.^{12,13} Las propiedades de los datos definidas para UH-Ontology son las siguientes:

- *accessprovidedby*: "acceso ofrecido por", establecido como propiedad para designar mención de responsabilidad sobre el acceso a la información.
 - *administers*: administradores es la propiedad establecida para clarificar dicha función.
 - *AdvisingRelationship*: tutor se establece como propiedad para referirse al que realiza el asesoramiento, seguimiento, guarda o custodia de personas o bienes.
 - *agent*: agente se utiliza para enlazar una persona (un agente) a un evento (una conferencia, un conversatorio).
 - *assignedby*: "asignado por" establecido como propiedad para designar mención de responsabilidad sobre asignación del agente.
 - *assigneeforpatent*: asignado por patente.
 - *assigns*: asignados.
 - *awardConferred*: premio conferido.
 - *base_near*: con sede cerca se utiliza para enlazar un agente, en relación con los recursos bibliográficos, a un lugar donde tiene su sede cerca de: puede ser una ciudad, un monumento, un edificio.
 - *citedBy*: citado por se utiliza para relacionar un documento a otro documento que cita al primer documento. Por tanto tiene definido como rango y dominio de la relación a la clase Document.
 - *contributor*: contribuidor se utiliza para vincular un material bibliográfico a un colaborador: puede ser un autor, un editor, una editorial. Contiene las siguientes subpropiedades:
 - a) *director*: referido a director de cine. Relación establecida entre la subclase AudioVisualDocument (de Document) como dominio y la clase Agent como rango.
 - b) *editor*: es una persona que tiene la responsabilidad de gestión y, a veces la formulación de políticas para la parte de redacción de una editorial o de un periódico, una revista u otro tipo de publicación. Relación establecida entre la clase Document o Collection como dominio y Agent como rango.
 - c) *interviewee*: el entrevistado es un agente que es entrevistado por otro agente. Por tanto la clase Agentes dominio y rango a la vez.
 - d) *Interviewer*: un agente que realiza una entrevista a otro agente. Por tanto la clase Agentes se considera dominio y rango a la vez.
 - e) *performer*: ejecutante es el agente que ejecuta. Por tanto se establece una relación entre la subclase Performance (de la clase Event) como dominio y la clase Agent como rango.
 - f) *translator*: el traductor es una persona que traduce el documento escrito de un idioma a otro. Relación establecida entre la clase Document o Collection como dominio y Agent como rango.
-

- *courseOfferedBy*: "curso ofrecido por" propiedad concerniente a un agente y vinculado a event.
- *court*: una institución asociada con la expedición de un documento legal; por ejemplo, el que haya expedido una decisión. Relación establecida entre la subclase LegalDocument (de la clase Document o Collection) como dominio y la subclase Organization (de la clase Agent) como rango.
- *CredentialOf*: credencial de.
- *CurrentlyHeadOf*: "actualmente jefe de" propiedad consignada a un agente (person) en relación con otro agente (person, organization o group).
- *degree*: no se define mediante una enumeración, sino como se define el rango de bibo: degree en la lista definida de Bibo: ThesisDegree. Se espera que el personal que interactúa con los datos de la ontología sea capaz de definir nuevos Degree si es necesario por algunos casos de uso especiales. La creación de dicha enumeración restringiría que esto suceda. Es la relación establecida entre la subclase Thesis (de la clase Document o Collection) como dominio y Thesis degree como rango.
- *degreeEarned*: título obtenido.
- *distributor*: distribuidor de un documento o una colección de documentos. Relación establecida entre la clase Document o Collection como dominio y Agent como rango.
- *DomesticGeographyFocus*: enfoque geográfico nacional.
- *editor*: la propiedad "editor" referida a la clase Agent se utiliza para identificar a la persona que tiene por oficio editar libros, periódicos, películas, discos, entre otros. Se utiliza para relacionar a Agent con Document.
- *EducationalTrainingOf*: referido a las acciones de superación posgraduadas
- *EventWithin*: definida para establecer actividades y acciones realizadas en el marco de un evento.
- *Format*: se utiliza para describir el formato de un recurso bibliográfico.
- *grantAwardedBy*: subvención concedida por.
- *hasAcademicDegree*: para designar el grado académico de una persona.
- *hasCredential*: tiene credencial o autorización para. Es una propiedad consignada para la clase Agent.
- *hasCurrentMember*: para consignar los miembros de la subclase Organización (grupo, organización o institución) o la clase Event.
- *hasEditorRole*: tiene rol de editor.
- *hasInvestigatorRole*: tiene rol de investigador.
- *haspart*: es parte de.
- *hasPrerequisite*: tiene prerequisite.
- *hasPrincipalInvestigatorRole*: es el primer investigador (para el dominio de la clase Person).
- *hasProceedings*: tiene procedimientos (para el dominio de la clase Event).
- *hasResearchArea*: tiene área de investigación.
- *hasReviewerRole*: tiene rol de revisión.

- *hasServiceProviderRole*: tiene rol de proveedor de servicios.
- *hasTeacherRole*: tiene rol docente.
- *hisTopicIS*: su tema es.
- *homepage*: se utiliza para unir un agente a su página de inicio (que es una página web accesible a través de una URL).
- *IncludesEvent*: incluye eventos.
- *inEventSeries*: en serie de evento
- *isPartOf*: es parte de, contiene la subpropiedad reproducedI
- *reproducedIn*: se utiliza para representar la relación del recurso en el que se produce otro recurso. Tiene dominio y rango en la clase Document.
- *isReferencedBy*: se utiliza para relacionar una cita de referencia a un recurso bibliográfico (para el dominio de la clase Person). Contiene la subpropiedad subsequentLegalDecision (para registrar la decisión legal en la apelación que actúa en un caso (afirmarlo, darle la vuelta). Con dominio y rango de la subclase Decision de la clase LegalDocument.
- *isVersionOf*: es versión de (para el dominio de la clase Document); contiene la subpropiedad translation of (referido a un documento traducido del documento original. Rango y dominio de la clase Document.
- *lenguaje*: se utiliza para vincular un recurso bibliográfico a la lengua utilizada para expresarlo.
- *list of contributors*: es una lista ordenada de los contribuyentes, generalmente tienen en cuenta el orden de importancia. Definido para la clase Document. Contiene las siguientes subpropiedades:
 - a) *list of authors*: lista ordenada de los autores.
 - b) *list of editors*: lista ordenada de los editores.
- *offersDegree*: ofertas de Grado.
- *organizer*: el organizador de un evento; incluye organizadores de la conferencia, agencias gubernamentales u otros organismos que son responsables de la realización de las audiencias. Tiene por dominio a la clase Event y por rango a la clase Agent.
- *owner*: propietario de un documento o una colección de documentos. Tiene por dominio a la clase Document o la clase Collection.
- *PersonInPosition*: situación de la persona.
- *place*: se utiliza para relacionar un evento como una conferencia con el lugar geográfico donde ocurre; por ejemplo, de París.
- *prerequisiteFor*: prerequisite para.
- *produced_in*: producido en contiene la subpropiedad presented at. Dicha subpropiedad se refiere al documento de un evento; por ejemplo, la ponencia para una conferencia.
- *producer*: productor de un documento o una colección de documentos. Tipo de relación establecida entre la clase Document o Collection como dominio y Agent como rango.
- *product*: se utiliza para enlazar un evento (como una conferencia) para un resultado (un producto) de ese evento: por ejemplo, un artículo, un

procedimiento. Contiene la subpropiedad `presented at` que relaciona un evento a los documentos asociados.

- *publisher*: se utiliza para vincular un material bibliográfico a su editor. Contiene la subpropiedad `issuer` como encargada de emitir los documentos publicados, a menudo informalmente, como comunicados de prensa, informes. La subpropiedad `issuer` (o emisor) establece una relación entre la clase `Document` o `Collection` como dominio y `Agent` como rango.
- *receipt*: recibido.
- *recipient*: referido a un agente que recibe un documento de comunicación. Por tanto, establece una relación entre la subclase `PersonalCommunicationDocument` (perteneciente a la clase `Document`) como dominio y la clase `Agent` como rango.
- *references*: referencia contiene la subpropiedad `cites`, que relaciona un documento a otro que se cita por el primero como referencia, comentario, opinión, cita o para otro propósito. Por tanto, establece una relación como dominio y rango con la clase `Document`.
- *relation*: la propiedad `relation` (relación) contiene la siguiente subpropiedad:
 - a) *annotates*: es la subpropiedad para establecer una nota crítica o explicativa para un documento. Establece una relación entre la subclase `Note` (perteneciente a la clase `Document`) como dominio y la clase `Resource` como rango.
 - b) *review of*: se relaciona con un documento de opinión sobre una cosa (de recursos, artículo). Establece una relación entre la clase `Document` como dominio y la clase `Resource` como rango.
 - c) *transcript of*: se refiere a un documento de un original transcripto. Establece una relación entre la clase `Document` como dominio y la clase `Resource` como rango.
- *rights*: se utiliza para describir los derechos relacionados con un recurso bibliográfico.
- *sponsoredBy*: patrocinado por; es una propiedad consignada para la clase `Agent`.
- *status*: se emplea igual que el rango de `bibo:status`, definido en la lista de `bibo:DocumentStatus`. Se espera que el personal que interactúa con los datos de la ontología sea capaz de definir nueva estatus si es necesario por algunos casos de uso especiales como el estado de la publicación de contenidos (normalmente académica).
- *sub_event*: se utiliza para vincular los grandes eventos con los eventos más pequeños, como talleres que se producen en el contexto de una conferencia.
- *SubGrantOf*: subtítulo de un `Document`.
- *Subject*: se utiliza para describir la temática de un recurso bibliográfico.
- *time*: se utiliza para describir el tiempo de duración un evento o acción.
- *title*: se utiliza para describir el título de un recurso bibliográfico.
- *value*: se utiliza para describir el contenido de un "Bibo:Document" y de otros recursos bibliográficos. Se aconseja utilizar esta propiedad en lugar de la "Bibo:content".

Propiedades de datos

Las propiedades OWL representan relaciones entre dos objetos o individuos. Dichas relaciones se pueden establecer mediante un XML Schema Data-type value o un literal RDF.^{12,13} Las propiedades de los datos, definidas para UH-Ontology, son las siguientes:

- *abstract*: con el propósito de acceder al resumen de los recursos.
- *changeNote*: para consignar cambio en una nota.
- *countryArea*: área o país.
- *date*: se utiliza para vincular un material bibliográfico a la fecha de un evento. Compruebe *dcterms:created* y *other* para especializaciones, adecuados para esta propiedad. Esta propiedad contiene las subpropiedades: *created* (representa la creación de un ítem bibliográfico) e *issued* (representa el dato numérico en un recurso bibliográfico).
- *dateargued*: la fecha en la cual un caso legal se alega ante un tribunal. El formato de la fecha es *xsd: date*.
- *description*: usado para describir un recurso bibliográfico.
- *doi*: representa la ubicación física de un documento antes de ser publicado.
- *edition*: edición.
- *editorialNote*: nota editorial.
- *elssn*: es equivalente a *eissn*.
- *endingPage*"pageend".
- *family_name*: equivalente al apellido de familia de una persona cuando guarda relación con un recurso bibliográfico.
- *givenname*: equivalente al nombre de pila de una persona relacionada con un recurso bibliográfico.
- *identifier*: es un identificador. Esta propiedad contiene a las subpropiedades: *asin*, *coden*, *doi*, *eanucc*¹², *eissn*, *isbn*, *issn*, *lccn*, *oclcnum*, *pmid*, *sici*, *upcy uri*.
- *isbn*: para el número de identificación internacional asignado a los libros.
- *issn*: para el número de identificación internacional asignado a las publicaciones periódicas.
- *issue*: se atribuye a la unidad de trabajo para realizar una mejora en un sistema informático. Un *issue* puede ser el arreglo de un fallo, una característica pedida, una tarea, un pedido de documentación específico y todo tipo de solicitud al equipo de desarrollo.
- *localityName*: se utiliza como número local para nombrar a la localidad de un editor y autor.
- *locator*: una descripción (para este caso numérica) que localiza un elemento dentro de un documento o colección que contiene. Es equivalente a la propiedad *number*. Incluye como subpropiedades *chapter*, *issue*, *page end*, *page start*, *pages*, *section*, *volume*.

- *number*: equivalente a locator.
- *numberID*: se utiliza para registrar número de carnet de identidad.
- *numberof pages*: se utiliza para representar el número de páginas contenidas en un documento.
- *number of volumes*: representa el número de volúmenes contenidos en una colección de documentos (por lo general una serie, publicación periódica).
- *prefixname*: referido al prefijo de un nombre.
- *PublishingGroup*: para registrar el grupo de publicación.
- *short title*: título abreviado de un documento.
- *shortdescription*: para registrar el resumen (especialmente el indicativo sobre el contenido del documento).
- *startingPage*: se utiliza para puntualizar la página de inicio.
- *suffixname*: sufijo del nombre.
- *volume*: volume.

Lenguajes Ontológicos utilizados

Con los lenguajes ontológicos que se utilizan se pretende mejorar la calidad de las descripciones y anotaciones semánticas. A continuación se muestran los lenguajes ontológicos empleados en UH-Ontology y las ontologías reutilizadas para hacer más rápido el proceso de diseño y hacerlo más interoperable con otras ontologías:

- *SKOS*:¹⁴ organiza las materias a partir de la clase SKOS:Concept e incluye las propiedades skos:prefLabel y Skos:altLabel, con las que se logran las relaciones sindéticas entre términos. En el diseño se suprimieron las relaciones de mapeo, ya que pueden ser asumidas por el sistema de información que emplee esta ontología. Se ha decidido usar el esquema del tesoro de la Unesco para normalizar la actividad de la universidad.
- *Bibo*:¹⁵ es una ontología que se ajusta a la descripción que se precisaría para los documentos que existen en la UH y se conecta con las clases de foaf a través de otras clases entre las que se encuentran: Colecciones, Recursos de Información, Documentos. Ellas han sido unidas en el modelo ontológico formando una restricción que las conecta como clases equivalentes que describen artículos, documentos audiovisuales, libros, parte de documentos, imágenes, documentos legales, manuales, manuscritos, notas, patentes, comunicaciones personales y sus referencias.
- *Vivo*:¹⁶ es una ontología que se nutre de relaciones con otros vocabularios ontológicos. Su objetivo fundamental está en crear un CRIS (Sistema de Gestión de la Investigación) con los datos de investigación de los docentes y estudiantes de la UH. Las clases en Vivo¹⁷ se relacionan con diversas instancias identificadas con una URI base única. Las instancias de la clase son individuales y se identifican de forma única por cada grado, persona o actividad. Para lograr relaciones eficientes en Vivo se utiliza la clase ISF, que tiene la intención de relacionar, mediante restricciones, aspectos de la vida académica de un docente: relaciona un grado académico único con la organización que lo emite, la persona que lo recibe, los valores de tiempo, ubicación (se trata de una representación

de una relación n-aria). Además las instancias de esta clase también pueden especificar el tipo de título recibido. La opción desarrollada en este trabajo ha sido crear clases adicionales en su modelo ontológico para separar y definir coherentemente los premios, distinciones y reconocimientos de la UH incluyendo clases propias. En el diseño de la ontología se incluyó una clase, *Event*, que viene del vocabulario para agrupar los eventos descritos por vivo-ontology. Sirve de clase de agrupamiento a las subclases *Exhibición*, *Performance*, *Eventos seriadados*, *Conferencias*, *Reuniones*, *Cursos*, *Presentaciones*, *Comunicaciones Personales*, *Talleres*, *Competencias*, *Entrevistas*. Las propiedades y sentencias describiendo conceptos jerárquicos de la clase *Event* de la ontología de eventos es la siguiente: *factor*, *agent*, *time*, *factor*, *sub-event*.

- *Geo-ontología*: se desarrolla a partir de una metodología para la representación de datos denominada ORD (Ontología de Representación de Datos), propuesta por *Larin-Fonseca* y *Garea Llano*, en la cual la integración de datos semánticos es la clave.¹⁸ Este tipo de ontología provee un mayor enriquecimiento semántico en el proceso de integración, pues permite una representación profunda de los contenidos. La estructura flexible y dinámica de la ORD permite capturar las características de los objetos geoespaciales descritas en los datos y las relaciones que se generan entre ellos. Esta propuesta es realizada sobre la premisa de que la integración de los datos geoespaciales ha de realizarse tomando en cuenta la naturaleza semántica que subyace en este tipo de datos en combinación con las Ontologías de Niveles Superiores (ONS). La fuente de datos utilizada es el tesoro *Agrovoc*, de la FAO, que posee especificaciones para datos geográficos explícitos, lo que no entra en contradicción con el modelo de datos que se propone. Este objetivo puede lograrse con el uso de la ORD para la representación explícita de la naturaleza semántica de los geo-datos y las relaciones entre ellos y posteriormente su vinculación con las diferentes ONS. Esto conlleva varios beneficios:
 - a) Se habilita la integración tomando en cuenta la semántica de bajo nivel.
 - b) Se incorpora el conocimiento subyacente en los geo-datos y las relaciones entre ellos en el proceso de integración.
 - c) Permite el uso simultáneo de diferentes enfoques de integración semántica a alto nivel.
 - d) Habilita la integración basada en representaciones semánticas multidimensionales de los objetos.
 - e) Habilita el trabajo con fuentes de datos dinámicas y heterogéneas.
 - f) Habilita la posible generación de nuevos niveles de abstracción.
- *DublinCore*: es un sistema de metadatos que centra su extensión en el manejo de 15 definiciones semánticas descriptivas que pretenden transmitir un significado *semántico* y *físico* de los documentos. Su extensión y aplicabilidad se basa en la capacidad de proporcionar la información descriptiva básica sobre cualquier recurso, sin que importe el formato de origen, el área de especialización o el origen cultural. Sus elementos se clasifican teniendo en cuenta el contenido, la propiedad intelectual y las características intrínsecas y extrínsecas del recurso o documento descrito. En el sistema de anotación se han usado los siguientes elementos:
 - a) *Título*: el nombre dado a un recurso, habitualmente por el autor (DC.Title).
 - b) *Claves*: las materias que indizan el recurso. Se fomentará el uso de vocabularios controlados y de sistemas de clasificación formales (DC.Subject).

- c) *Descripción*: una descripción textual del recurso. Puede ser un resumen en el caso de un documento o una descripción del contenido en el caso de un documento visual (DC.Description).
- d) *Fuente*: secuencia de caracteres usados para identificar unívocamente un trabajo a partir del cual proviene el recurso actual (DC.Source).
- e) *Relación*: es un identificador de un segundo recurso y su relación con el recurso actual (DC.Relation).
- f) *Autor o Creador*: la persona u organización responsable de la creación del contenido intelectual del recurso (DC.Creator).
- g) *Editor*: la entidad responsable de hacer que el recurso se encuentre disponible en la red en su formato actual (DC.Publisher).
- h) *Otros Colaboradores*: una persona u organización que haya tenido una contribución intelectual significativa, pero que esta sea secundaria en comparación con las de las personas u organizaciones especificadas en el elemento Creator (DC.Contributor).
- i) *Derechos*: son una referencia (por ejemplo, una URL) para una nota sobre derechos de autor, para un servicio de gestión de derechos o para un servicio que dará información sobre términos y condiciones de acceso a un recurso (DC.Rights).
- j) *Fecha*: una fecha en la cual el recurso se puso a disposición del usuario en su forma actual (DC.Date).
- k) *Formato*: es el formato de datos de un recurso, usado para identificar el software y, posiblemente, el hardware que se necesitaría para mostrar el recurso (DC.Format).
- l) *Identificador del Recurso*: secuencia de caracteres utilizados para identificar unívocamente un recurso. Ejemplos para recursos en línea pueden ser URLs y URNs. Para otros recursos pueden ser usados formatos de identificadores como el ISBN ("International Standard Book Number") (DC.Identifier).
- m) *FOAF*: es el acrónimo de la ontología Friend of a Friend utilizada para la definición de 3 entidades básicas: Group, Organization y Person. Aquí se emplean para definir la organización de todos los miembros de la comunidad universitaria, las organizaciones que contiene, los grupos de trabajo. Se han empleado elementos como foaf:mbox, foaf:homepage, foaf:surname, foaf:givenname, foaf:birthaday , foaf:knows, foaf:title...
- n) *Lom*: el uso de objetos de aprendizaje ha adquirido importancia en los últimos años, especialmente en las universidades. El desarrollo de iniciativas de estandarizar la manera de describirlos lo demuestra y ejemplos de esto es LOM (Learning Object Metadata). Este se refiere a la estructura de organización de los objetos de aprendizaje. Este valor puede ser especificado por varios de los términos definido en una ontología con lenguaje OWL. En el caso de la UH-Ontology se ha definido dentro de la clase Document y contiene varias especificaciones. El objetivo es clasificar los tipos de recursos de aprendizaje que pudiesen gestionarse desde la ontología (Learning_Resource_Type).

Anotaciones

Las anotaciones constituyen otro tipo de propiedad. Se definen y utilizan para agregar comentarios y descripciones en lenguaje natural. Las anotaciones previstas desde UH-Ontology permitirán realizar anotaciones a las clases, propiedades de objetos y propiedades de datos sobre los siguientes aspectos: *changeNote*, *comment*, *creator*, *deprecated*, *description*, *editorialNote*, *example*, *historyNote*, *identifier*, *incompatibleWith*, *isDefinedBy*, *isPartOf*, *label*, *note*, *priorVersion*, *scopeNote*, *seeAlso*, *term_status*, *title* *versionInfo*. Además, permite anotarlos identificando los siguientes idiomas: inglés (en), español (es) y francés (fr).

Esquema de anotación semántica

La anotación semántica será el proceso mediante el cual se describan los contenidos que se asocian a un sitio web. Se basa en alguna de las herramientas al uso para estos fines, como Annotea. Desde la óptica de esta propuesta, la anotación semántica requiere el reconocimiento de las limitaciones de PHP 5 para conseguir descripciones de datos de forma profunda y coherente de acuerdo a las necesidades de algunas entidades. Por tanto, se ha determinado trabajar tomando como base el esquema de anotación de Annotea, al que se le hacen algunas modificaciones (que denominamos agregaciones). Basándonos en la propuesta de *Navarro Galindo*, se tienen en cuenta tres requisitos que deben estar presentes en el proceso de anotación:¹⁹

- El uso de UH-Ontology tanto a nivel de infraestructura durante el proceso de creación de anotaciones semánticas, como a nivel de referencia durante el proceso de asociación de significado a los textos marcados.
- El uso de algún esquema de anotación propuesto por W3C. Para este caso particular se parte de Annotea y se enriquece en función de las necesidades de la propuesta.
- El uso de estándares de la W3C para el marcado semántico. Para este caso se implementan RDF, RDFa y OWL.

El modelo de anotación semántica de UH-Ontology se emplea para las siguientes tareas:

- Almacenar el conocimiento generado en la UH en el cumplimiento de su función social en el ámbito docente, investigativo y extensionista y de su ejercicio administrativo.
- Estructurar el conocimiento de manera que propicie la comprensión de los significados por los autómatas y por humanos en los procesos de búsqueda de información.
- Garantizar la escalabilidad del sistema utilizando formatos de metadatos aprobados por el consorcio de la W3C.
- Facilitar la búsqueda compartida y la identificación unívoca de cada recurso de información.
- La reutilización de los acervos temáticos de la comunidad universitaria en forma de concepto.

No queda descartada la posibilidad de vincular, en su futura implementación, ontologías de dominio que favorezcan el uso de un vocabulario consensuado, al aportar una taxonomía de conceptos, dada la diversidad de temas que se trabajan en la UH desde la docencia y la investigación en sus áreas. Las ontologías tienen la finalidad de facilitar la gestión eficiente de las anotaciones; por eso, en este modelo se necesita declarar cómo las diversas ontologías que sirven de punto de apoyo al sistema funcionarán para la realización de anotaciones especiales en el texto HTML.

CONCLUSIONES

UH-Ontology ofrecerá a la UH la posibilidad de normalizar los atributos de los metadatos aplicables a los documentos; crear una red de relaciones que aporte especificación y fiabilidad; la reutilización del conocimiento existente; compartir conocimiento en pos de hacer explícito los diferentes criterios, así como la integración de diferentes perspectivas de los usuarios, fomentando el trabajo cooperativo al funcionar como soporte común de conocimiento entre las distintas áreas y estructuras de la UH. Como consecuencia, la interoperabilidad entre los distintos sistemas existentes en la UH y el establecimiento de modelos normativos, contribuirán a la creación de la semántica de un sistema y de un modelo para poder extenderlo y transformarlo en los diferentes contextos de la Universidad.

La propuesta de la ontología para homogenizar y gestionar los datos online de la UH promete significativos avances en los procesos sustantivos de gestión de la información institucional, tanto de sus recursos de información como de datos personales y organizacionales. Bajo estas expectativas, se pretende que UH-Ontology se implemente como núcleo de un sistema gestor de contenido semántico (S-CMS) cuyo objetivo será facilitar la recuperación y descripción semántica de los contenidos. Este sistema debe permitir la anotación semántica para la realización de búsquedas federadas y semánticas en diferentes aplicaciones e integrar un protocolo de cosecha de metadatos para manejar las aplicaciones que lo nutren. Tal proyecto pondrá a la UH en consonancia con las nuevas tendencias de la Web semántica y las consecuentes ventajas para la gestión de información institucional, la visibilidad como centro de educación superior en las redes y el fomento del trabajo colaborativo.

DECLARACIÓN DE AUTORÍA

Yorbelis Rosell, José Senso y Amed Leiva diseñaron el estudio. Yorbelis Rosell y Amed Leiva diseñaron la ontología. Yorbelis Rosell y José Senso redactaron la primera versión del manuscrito. Todos los autores revisaron la redacción del manuscrito y aprueban la versión finalmente remitida.

CONFLICTO DE INTERESES

Los autores no declaran conflicto de intereses.

REFERENCIAS BIBLIOGRÁFICAS

1. Noy NF, McGuinness DL. Desarrollo de Ontologías-101: guía para crear tu primera ontología. 2005 [citado 16 de agosto de 2016]. Disponible en: http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf

2. Lozano-Tello A, Gómez-Pérez A, Sosa E. Selection of ontologies for the semantic Web. In: Lovelle JMC, Rodríguez BMG, Aguilar LJ, Gayo JEL, Ruiz MDP (eds.). Web Engineering, Proceedings. Berlin: Springer-Verlag Berlin. 2003.
3. Gruber T. Toward principles for the design of ontologies used for knowledge sharing. Technical Report KSL-93-04.CA: Knowledge Systems Laboratory Stanford University; 1993.
4. Gruber T. A translation approach to portable ontologies. Knowledge Acquisition. 1993;5:199-220.
5. Studer R, Benjamins R, Fensel D. Knowledge Engineering: Principles and Methods. Data and Knowledge Engineering. 1998;25:161-97.
6. Antoniou G, Franconi E, Van Harmelen F. Introduction to semantic web ontology languages. Reasoning web. Springer; 2005.
7. Universidad de La Habana. Intranet de la Universidad de La Habana. La Habana: Producciones IRIS. 2010 [citado 18 de diciembre de 2015]. Disponible en: <http://intranet.uh.cu/universidad-de-la-habana/mision-vision-y-objetivos>
8. Solomou G, Papatheodorou T. The use of SKOS vocabularies in digital repositories: the DSpace case. Pittsburgh, Pennsylvania: IEEE Fourth International Conference; 2010.
9. Brickley D, Miller L. FOAF vocabulary specification 0.98. 2010 [citado 18 de diciembre de 2015]. Disponible en: <http://xmlns.com/foaf/spec>.
10. Dimi Surla B, Segedinac M, Ivanović D. A BIBO ontology extension for evaluation of scientific research results. In: Proceedings of the Fifth Balkan Conference in Informatics; 2012:275-8.
11. Corson-Rikert J. BIBO 1.3 Ontology overview diagram. Universidad de La Habana; 2016.
12. Florez Fernández HA. Construcción de ontologías OWL. Vínculos. 2013;4:19-34.
13. Horridge M, Knublauch H, Rector A, Stevens R, Wroe C. A Practical Guide to building OWL ontologies using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. University of Manchester; 2004.
14. W3C. SKOS Simple Knowledge Organization System. 2012 [citado 10 de diciembre de 2015]. Disponible en: <http://www.w3.org/2004/02/skos/>
15. D'Arcus B, Giasson F. Bibliographic Ontology Specification. Madrid: Biblioteca Nacional Española; 2009.
16. Cornell UD. VIVO: ISF. 2009 [citado 17 de septiembre de 2015]. Disponible en: <https://wiki.duraspace.org/display/VIVO/VIVO-ISF+Ontology>
17. Corson-Rikert J, Mitchell S, Lowe B, Rejack N, Ding Y, Guo C. The VIVO ontology. Synthesis Lectures on Semantic Web: Theory and Technology; 2012.

18. Larin-Fonseca R, Garea-Llano E. Automatic Generation of the Data-Representation Ontology for Semantic Integration of Heterogeneous Geographical Data. VII Congreso Internacional Geomática; 2011.

19. Navarro Galindo JL. FLERSA: un sistema semántico de gestión de contenido web (S-CMS). [Tesis doctoral]. Universidad de Granada. 2012.

Recibido: 25 de mayo de 2016.

Aprobado: 11 de julio de 2016.

Yorbelis Rosell León. Universidad de La Habana. La Habana, Cuba.
Correo electrónico: yorbelisr@dict.uh.cu

* Aguilera F. Entrevista a especialista de la Dirección de Formatización; 2016.