

# OntoCatMedia: ontología para la búsqueda y clasificación automática de medias audiovisuales

Yuniel Eliades Proenza Arias  
Alianis Pérez Sosa

*En la actualidad realizar búsquedas en un lugar donde existe gran cantidad de información es un reto para los desarrolladores de software. A partir de la introducción y utilización de las ontologías en los procesos que conforman la fase de organización y representación de la información, los resultados de las búsquedas serán mejores y el usuario podrá obtener de una manera más fácil y automática lo que necesita. Las búsquedas ontológicas, basadas en esquemas conceptuales, generan resultados coherentes con el criterio del usuario sobre uno o varios dominios de información. El objetivo de este trabajo es crear una ontología para la clasificación y búsqueda de materiales audiovisuales sobre un catálogo de medias, para tal propósito se nombró como OntoCatMedia. Se especifican los lenguajes y herramientas usadas en su construcción, la metodología de desarrollo utilizada y por último se prueba el razonamiento de la ontología desarrollada, demostrando su importancia y utilidad.*

**Palabras clave:** software, ontología, búsquedas ontológicas, representación del conocimiento, metadatos

## RESUMEN

## ABSTRACT

*Currently, searching over big amounts of information is to challenge for software developers. With the introduction and use of ontologies in the processes that organize and represent information, search results will be better and users will get what they need in an easier and expedite way. The ontological search, based on conceptual schemes, generates results consistent with the user's criteria on one or different information domains. The aim of this work is creating ontologies for classifying and searching audio-visual materials over a media catalog. This ontology has been named OntoCatMedia. The languages and tools used for its development and the design methodology used are described and finally, the ontology concept is tested demonstrating its significance and utility.*

**Keywords:** Software, ontology, ontological search, knowledge representation, metadata

## Introducción

**E**l desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC's) han permitido que la gestión de información se realice de manera más eficiente ya que la inclusión de la informática en los diferentes sectores de la sociedad ha desencadenado la digitalización de grandes volúmenes de datos o información.

En la actualidad realizar búsquedas en un lugar donde existe gran cantidad de información es un reto para los desarrolladores de software, por esta razón se pretende perfeccionar este proceso para lograr su eficiencia. Existen fuentes de información en la Web que poseen buscadores que le devuelven al usuario la información solicitada,

estos motores de búsqueda que funcionan sintácticamente devolviendo solo la información en la cual aparecen elementos relacionados con lo consultado, tal y como ha sido especificado, provocando en muchas ocasiones que las respuestas arrojadas no sean las esperadas.

Para solucionar estos problemas se transforma Internet en una Web semántica, ante su función de hacer bajo este paradigma, un gran flujo de congruencias de mecanismos y reglas que proporcionen secuencias correctas a los elementos de un lenguaje de programación, contribuyendo desde sus aristas a potenciar a Internet como una potente fuente de conocimiento.

Basándose en estos necesarios cambios, se hace imprescindible para los esquemas de intercambio que se producen en la Web semántica, unificar contenidos a través de un lenguaje común, la ontología, cumple esta función.

Una ontología es una especificación de una conceptualización o, como diría Gruber, (1993) es una representación conceptual compartida que proporciona una comprensión común de un dominio. En inteligencia artificial, donde además surgió el término, una ontología es una descripción de los conceptos y relaciones que existen para un agente o una comunidad de agentes, así su función principal consiste en facilitar el intercambio de información entre los agentes que hasta principio de los años 90 solo era conocido en el campo de la inteligencia artificial. Una ontología es una herramienta conceptual que define un vocabulario común para quien necesita compartir información dentro de un determinado dominio, esto incluye definiciones de los conceptos básicos del dominio, así como sus relaciones que tienen que ser interpretadas por máquinas (Gruber, 1993).

Las búsquedas ontológicas se caracterizan por el uso de semántica explícita, se realizan inferencias basadas en lógica, a diferencia de la búsqueda tradicional, donde se trata de encontrar los resultados sin tratar de comprender el término de búsqueda especificado. La búsqueda semántica le permite al usuario entender qué significa el término establecido en la consulta, en base al conocimiento semántico explícito dado en la ontología. Estas búsquedas permiten relacionar el término de búsqueda con un concepto del dominio cuyo significado está definido y con ello conocer las propiedades y las relaciones que dicho concepto posee.

En el Centro de Desarrollo Geoinformática y Señales Digitales (GeySED) de la Universidad de las Ciencias Informáticas de La Habana, Cuba, se está desarrollando el proyecto Sistema de Captura y Catalogación de Medias (SCCM).

Este proyecto se encuentra enfrascado actualmente en el desarrollo de un sistema que

permita gestionar materiales audiovisuales. Uno de los procesos que realiza el proyecto es capturar las medias que han sido digitalizadas previamente y catalogarlas. Para ello le asigna a cada media una clasificación (video musical, documental, cine, u otros) y le inserta además todos los datos descriptivos pertenecientes a esta clasificación (año, título, director, estos dependen de la clasificación) y sus valoraciones sobre el mismo. Una vez que el material posee todos estos datos está catalogado. Las medias al ser catalogadas son almacenadas en un servidor especializado, donde se acceden a través de un proceso de búsqueda por un usuario determinado.

Actualmente existe una versión del sistema donde las búsquedas de las medias se realizan a través de consultas a la base de datos, donde se toman como criterio las descripciones que se le adjuntan a las medias, las cuales son funciones descritas manualmente por una persona. La ambigüedad del lenguaje y los diferentes términos existentes para referirse a una misma información provocan que un documentalista audiovisual defina un término para una media y otro documentalista audiovisual defina otro término con igual significado para el mismo tipo de media. Por tanto, debido a la clasificación manual de las medias, el término utilizado para realizar la búsqueda posee disímiles significados y los resultados arrojados no son los deseados.

Con el uso de una ontología se pretende garantizar el tratamiento de tres elementos documentales y descriptores importantes:

**Polisemia:** Se realizan consultas de un término. Los resultados devueltos incluyen documentos con el término solicitado, pero significados distintos.

**Sinonimia:** Se realiza una consulta que devuelve muchos documentos que en ocasiones no son los correctos, esto es debido a que en los textos aparece un sinónimo del término de la consulta y no el propio término, ante la propiedad de este término de relacionar de forma semejante los significados entre determinadas palabras.

**Multilingüismo:** Se realiza una consulta con términos en un determinado idioma que no está presente en los textos. Sin embargo, el término está presente en otro lenguaje. Este término describe el hecho de forma tal que sea capaz de expresarse en varias lenguas, reflejando diversidad cultural y lingüística.

## Metodología

Para la construcción de OntoCatMedia se usaron varios métodos de investigación. El

Análisis Histórico – Lógico se utilizó para el estudio de los antecedentes y uso de las ontologías, tendencias actuales en su desarrollo, conceptos, metodologías y lenguajes de diseño más utilizados en el desarrollo de ontologías, la identificación de ontologías existentes asociadas al dominio en cuestión y las herramientas de desarrollo.

El método Analítico – Sintético fue utilizado en el estudio de las ontologías existentes, analizando sus componentes (clases, relaciones, instancias, funciones, taxonomía, axiomas). Este proceso de análisis permitió definir los requisitos y elementos necesarios para el desarrollo de la ontología e identificar si alguna de las ontologías existentes pudiera ser reutilizada. Finalmente, se utilizó la Modelación en la representación de los elementos relacionados con el dominio, en la construcción de la ontología como modelo semántico, teniendo en cuenta conceptos, relaciones entre conceptos, jerarquía de clases, taxonomía, entre otros elementos.

Para el caso específico del desarrollo de la ontología, se utilizó la metodología Methontology (Fernandez-Lopez, Gomez-Perez, & Juristo, 1997), recomendada por FIPA (*Fundación de Agentes Físicos Inteligentes*) la cual tiene sus raíces en las actividades similares al proceso estándar de desarrollo de software. Esta metodología está basada en *prototipos evolutivos, permitiendo añadir, cambiar y eliminar términos en cada nueva versión de la ontología*. Además incluye tareas de mantenimiento de la ontología y gestión constante de cambios.

## Búsqueda sobre catálogos basada en ontologías

A través de una ontología se pueden realizar búsquedas sobre una información clasificada y ordenada. La posibilidad de seleccionar una palabra o frase y poder realizar búsquedas en catálogos abre una vía para ofrecer una mayor interactividad con el usuario. Una ontología basada en catálogos permitirá describir completamente el esquema de la base de datos que se desee representar mediante la instanciación de la misma. Para realizar las búsquedas los usuarios especifican su criterio de búsqueda en función de conceptos y atributos que se establecen inteligentemente. Una vez establecida la consulta se navega por la estructura jerárquica que define la ontología, recuperando los términos que tengan relación con la petición realizada, de esta forma las respuestas son más precisas y completas.

Existen muchas aplicaciones que están basadas en ontologías para búsqueda sobre catálogos, algunas de ellas son a través de *NextLib* (Vidal, 2008), este sistema utiliza una ontología para realizar búsquedas de términos seleccionados en una página Web en el catálogo de una biblioteca. El **Módulo Ontológico** (Bustamante, 2006) está integrado a la biblioteca digital de la Escuela de Ingeniería de Sistemas y Computación de la Universidad del Valle de Colombia (EISC), para dar soporte a las consultas avanzadas y a la recuperación de documentos. **SABIOS** (Giraldo, Guzmán, & Ceballos, 2006) es otra aplicación que contiene una ontología especializada en Hábitat y Artes Plásticas. Además de estas también se pueden citar (Mauricio Espinoza J. G., 2006), (Mauricio Espinoza R. T., 2006) y (Kannan, 2010).

Todas estas aplicaciones que utilizan ontologías han proporcionado beneficios en el trabajo de muchas personas en el proceso de recuperación de información. El uso de ontologías pretende que las máquinas entiendan por sí solas qué es lo que el usuario necesita, es decir, que entienda el significado de la palabra clave que fue introducida como criterio de búsqueda. De esta forma se hace una búsqueda entre los conceptos definidos en el dominio, atravesando por la herencia y las relaciones que existen entre ellos, permitiéndole al software llegar a un razonamiento.

Luego de realizar un análisis de varias ontologías y aplicaciones, con el objetivo de reutilizar el conocimiento en el desarrollo de la ontología que se propone en esta investigación, se llegó a la conclusión que no se puede reutilizar ninguna de las ontologías debido a que no pertenecen al dominio de las medias audiovisuales. Además, antes de reutilizar una ontología, se debe analizar el contenido que la misma proporciona, los aspectos metodológicos y los tecnológicos.

## Desarrollo de OntoCatMedia

A partir del estudio realizado, se propone el diseño de una ontología para la catalogación de medias (OntoCatMedia), que contenga los principales conceptos descriptivos de materiales audiovisuales y otros asociados a clasificaciones utilizadas en el propio proceso de catalogación.

Además, en la ontología se incluirían las principales relaciones conceptuales para la construcción total del vocabulario utilizado en la clasificación.

A partir de la conceptualización del dominio en cuestión, el sistema de catalogación usaría la ontología en la sugerencia de términos descriptivos para medias, como principal utilidad; pero además en la búsqueda de medias a partir de términos descriptivos, basando ambos análisis en un procesamiento semántico de la información.

Para definir los términos que conforman OntoCatMedia se hicieron investigaciones sobre el dominio de los materiales audiovisuales. Existen diferentes temáticas para clasificar los materiales audiovisuales. Para comenzar a definir los términos se solicitó ayuda en el Centro de Televisión de la UCI, con el objetivo de conocer la forma en que se clasifican actualmente los materiales audiovisuales, además de algunos bibliotecarios. Esta investigación aportó una primera versión para un posible glosario de términos a través de un vocabulario controlado.

Después de construir el glosario de términos se dio paso a la construcción de la taxonomía de conceptos de la ontología. Hay varios enfoques para desarrollar una jerarquía (taxonomía) de clases (conceptos) (McGuinness, 2005). Para el caso de OntoCatMedia se utilizó un enfoque **top-down**.

Para el desarrollo de OntoCatMedia se utilizó la herramienta Protégé<sup>1</sup> y utilizando el lenguaje OWL (Horrocks, Patel-Schneider, & Van Harmelen, 2003). Utilizando OWL se definieron clases principales, de las cuales se derivan otras clases más específicas denominadas subclases (ver Fig. 1) pertenecientes al segundo nivel. Una parte de esa taxonomía se estructura de la siguiente forma: una clase *Media Audiovisual* de la cual se derivan otras clases que son subtipos de ellas, *Imagen*,



Figura 1. Taxonomía de conceptos.

*Promocion*, *Originales de Camara*, *Programa Variado*, y demás. Esta última clase da origen a otras definiciones de conceptos, como son, *Documental* que es un tipo de programa variado al igual que un *Informativo* que a su vez es un *Noticiero*. Es decir una clase más general que engloba clases más específicas con características similares.

Basándose en la definición de la taxonomía de conceptos y de la descripción de cada concepto por separado, se establecieron algunas propiedades que le permiten a la ontología la inferencia de nuevo conocimiento. Entre las propiedades de objeto definidas (Object properties), en la Figura 2 se muestra el ejemplo de la relación entre *Noticiero* y *Noticia*, esta propiedad específica que las medias de tipo noticiero están conformadas por noticias.

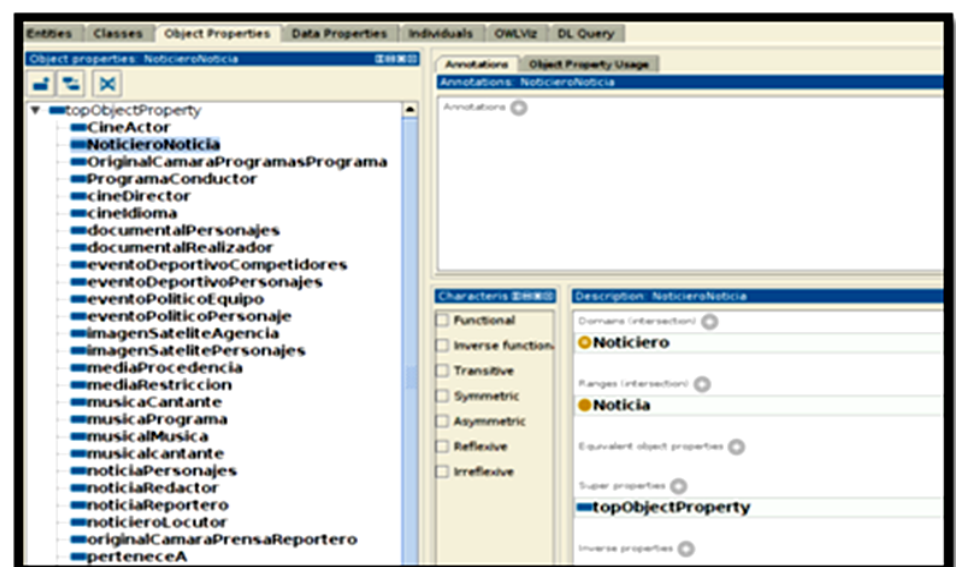


Figura 2. Ejemplo de propiedades de clases.

<sup>1</sup> <http://protege.stanford.edu/>

Por otro lado, existen los axiomas formales, los cuales son expresiones lógicas que son siempre verdaderas y utilizadas normalmente para especificar restricciones en la ontología. Estas restricciones se establecen entre conceptos y propiedades para lograr una inferencia en el conocimiento. Para esta primera versión de OntoCatMedia se definieron algunos axiomas

La Fig. 3 muestra la creación de una clase equivalente de la clase **Cortometraje**, implicando dos clases: **Cine** y **Corto**. La propiedad *poseeUn* especifica que una media audiovisual posee un tiempo de duración determinado. Con esta restricción se infiere que se clasifican como cortometrajes solo aquellas medias que sean cine y que posean un tiempo de duración corto.

### Pruebas de razonamiento

En este epígrafe se describen pruebas de razonamiento para verificar la consistencia de la ontología en cuanto a las relaciones establecidas entre los conceptos (clase equivalentes, propiedades, axiomas, y otros específicos). Una vez realizadas las pruebas se podrá verificar si la ontología soluciona el problema planteado, satisfaciendo las necesidades del proyecto SCCM en los procesos de búsqueda y clasificación de medias sobre el catálogo. Para la realización de las pruebas de razonamiento a OntoCatMedia y comprobar el conocimiento que es capaz de inferir, se utilizan los razonadores semánticos *Pellet* y *FaCT++*, ambos integrados al Protégé.

### Tratamiento de la sinonimia

Con la creación de OntoCatMedia se garantizan mejores resultados en las búsquedas de medias en el proyecto SCCM. Las clases equivalentes en una ontología se establecen con el objetivo de considerar la sinonimia, de esta forma cuando se realiza una consulta se obtienen las respuestas del término deseado y de sus sinónimos. Por ejemplo, cuando se realiza una consulta utilizando como palabra clave el término *plano*, se obtienen entre los resultados las medias clasificadas como *plano* y además las que estén clasificadas como *dibujo* y *esquema*. Para esto se definieron las clases *dibujo* y *esquema* como clases equivalentes de la clase *plano*.

### Clasificación y búsqueda

Los razonadores semánticos se incluyen como plug-ins adicionales en Protégé para hacer

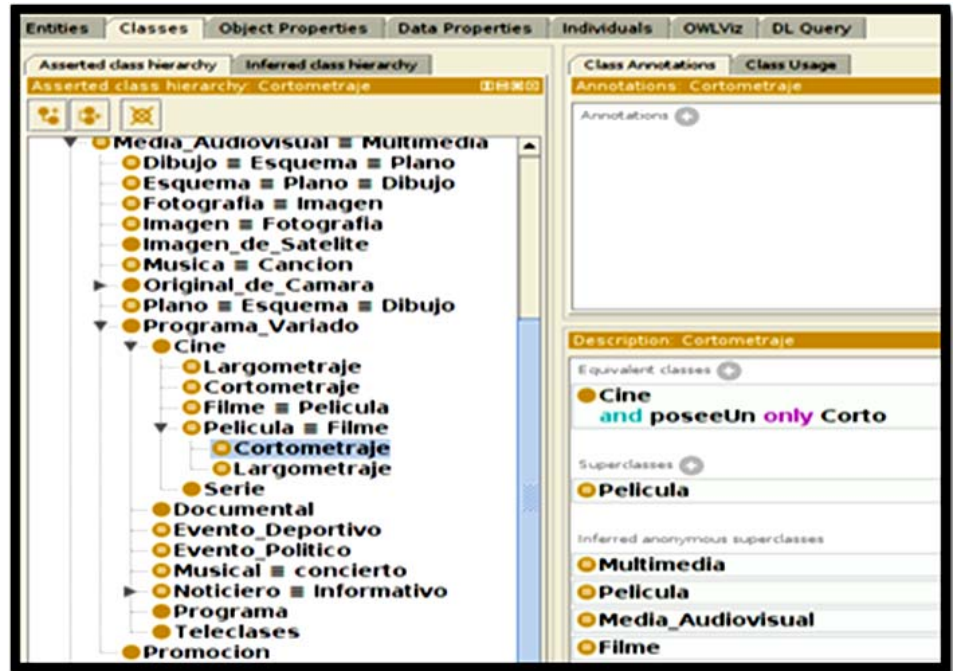


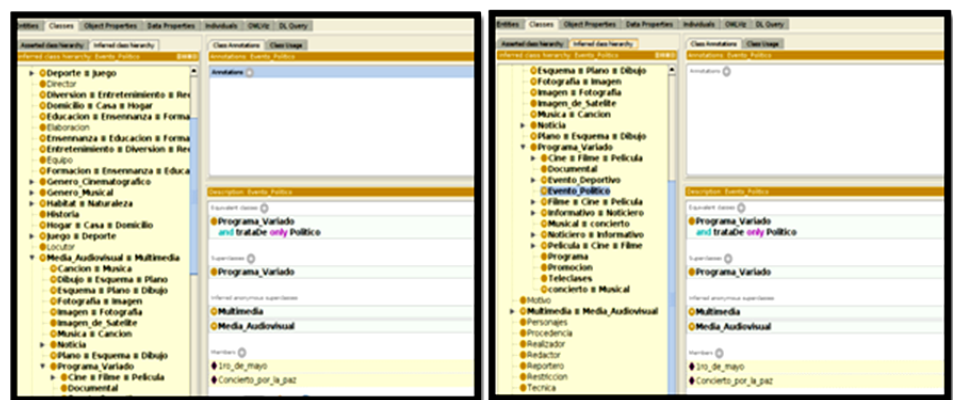
Figura 3. Ejemplo de axiomas formales.

inferencia sobre los datos representados en la ontología. Al aplicarle un razonador a una ontología los datos que se infieren se muestran sombreados en amarillo, de esta forma se comprueba la consistencia de la ontología en cuanto a los conceptos y las relaciones definidas entre ellos. Para comprobar el razonamiento que realiza OntoCatMedia al inferir los datos, se crearon 30 instancias de la clase **Thing**. A estas instancias se les asociaron un conjunto de propiedades que le permiten a la ontología clasificarlas de acuerdo a sus descripciones.

En la Fig. 4 se muestra la primera prueba de razonamiento realizada a OntoCatMedia, al aplicársele ambos razonadores por separado se infirieron los mismos resultados. La comparación de los resultados obtenidos demuestra la consistencia de la ontología y la

capacidad de razonamiento que posee la misma. En el ejemplo se muestra la herencia de clases inferida y los datos que se obtienen sobre la clase seleccionada. Por tanto, se infiere que un **Evento\_Politico** es un programa que trata de política, también es una media audiovisual o multimedia y se especifica que las instancias **Concierto\_por\_la\_paz** y **Marcha\_1ro\_de\_mayo** son medias del tipo evento político, de acuerdo a las características que se establecieron en su descripción.

En otra de las pruebas (ver Fig. 5) se hace inferencia sobre la instancia **Remember\_me**. A esta instancia se le asocian algunas propiedades que infieren que la instancia es un filme, pero además se le especifica en el tipo de descripción a través de la propiedad *poseeUn*, que el tiempo de duración de la instancia es largo. Al aplicar los razonadores

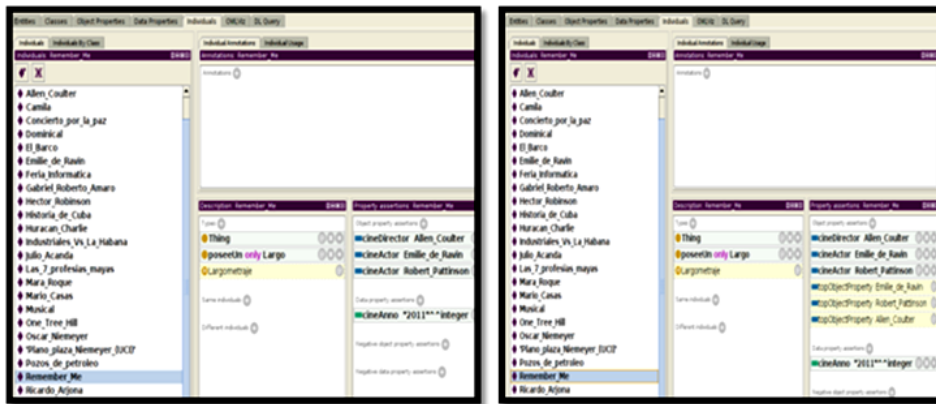


### Razonador Pellet

### Razonador FaCT++

Figura 4. Inferencia de datos con razonadores semánticos (1).

se obtiene que la instancia sea un largometraje, puesto que se declaró un axioma formal que restringe como largometrajes solo aquellos filmes que posean un tiempo de duración largo. En esta prueba realizada también se muestra como se apoya al proceso de búsqueda de medias a través de la ontología, obteniéndose que la instancia es un largometraje, o lo que es lo mismo, es un filme con un tiempo de duración largo.



Razonador Pellet

Razonador FaCT ++

Figura 5. Inferencia de datos con razonadores semánticos (2).

Las pruebas realizadas demuestran que la ontología es capaz de brindarle al documentalista audiovisual una serie de opciones para clasificar la media que está catalogando, también devolver los sinónimos de la palabra clave que se introdujo como criterio de búsqueda. La ontología cumple con los objetivos planteados debido a que es capaz de reconocer las clases equivalentes y clasificar objetos para esas clases, por otra parte permite clasificar los objetos en temáticas, así como definir todos los tipos de datos pertenecientes a una misma temática. En la siguiente tabla (ver Tabla 1) se muestra una comparación de los resultados obtenidos por ambos razonadores al realizar la pruebas de razonamiento.

Tabla 1. Comparación de razonadores en cuanto a resultados.

Aspectos	Razonadores semánticos	
	Pellet	FaCT ++
Clasificación temática de un objeto	X	X
Agrupamiento de objetos por temática	X	X
Reconocimiento de clases equivalentes	X	X
Clasificación temática de objetos a clases equivalentes	X	X
Inconsistencias en la ontología	Ninguna	Ninguna
Reconocimiento de axiomas formales	X	X

## Conclusiones

Con el desarrollo de la investigación se ha arribado a las siguientes conclusiones:

- Se consideró necesario el desarrollo de la ontología denominada OntoCatMedia para el dominio de los materiales audiovisuales, sin reutilizar ninguna de las ontologías existentes debido a que ninguna puede ser reutilizada, puesto que no se adaptan al dominio en cuestión.

- El uso de OntoCatMedia en el SCCM permite realizar la clasificación temática de medias de manera semi-automática, brindando una serie de opciones al documentalista durante el proceso de catalogación, evitando que se definan diferentes términos descriptivos para un mismo tipo de información.

- Las pruebas realizadas a través de los razonadores semánticos integrados a Protégé, demuestran que la ontología presenta gran consistencia y un alto nivel de razonamiento para inferir información y mejorar de esta manera la búsqueda sobre la información almacenada en el catálogo de medias.

## Recomendaciones

- Continuar enriqueciendo semánticamente la ontología, incorporando nuevas clases para inferir más conocimiento.

- Realizarle pruebas de razonamiento a la ontología a través de consultas SPARQL.

- Integrar la ontología al SCCM.

## Referencias

Bustamante, D. P. (2006). Desarrollo de un módulo ontológico para las consultas avanzadas de información de la biblioteca digital de la escuela de Ingeniería de Sistemas y Computación de la Universidad del Valle. Colombia.

Fernandez Lopez, M., Gomez Perez, A., & Juristo, N. (1997). Methontology: from Ontological Art towards Ontological Engineering. Proceedings of the AAAI97 Spring Symposium, (págs. 33-40).

Giraldo, J. D., Guzmán, J. A., & Ceballos, J. C. (2006). Desarrollo de un Módulo de Visualización y Navegación de Información Basada en Ontologías para el Sistema de Recuperación Semántica SABIOS. Avances en Sistemas e Informática, 3 (1), 57-61.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5 (2), 199—220.

Horrocks, I., Patel-Schneider, P. F., & Van Harmelen, F. (2003). From SHIQ and RDF to OWL: The Making of a Web Ontology Language. Journal of Web Semantics, 1 (1), 7-26.

Kannan, R. (2010). Topic Map: An Ontology Framework for Information Retrieval. India.

Mauricio Espinoza, J. G. (2006). Discovering the Semantics of Keywords: An Ontology-based Approach. In The 2006 International Conference on Semantic Web and Web Services (SWWS'06), Las Vegas (págs. 193—201). CSREA Press.

Mauricio Espinoza, R. T. (2006). Discovering and Merging

Keyword Senses using Ontology Matching. In 1st International Workshop on Ontology Matching (OM-2006) at ISWC-2006, (págs. 1—5).

McGuinness, N. F. (2005). Desarrollo de Ontologías 101: Guía para crear tu primera ontología. Stanford.

Vidal, L. F. (2008). NextLib: un sistema de software basado en ontologías para la consulta automática de OPACs desde sedes web especializadas (20).

Recibido: 6 de junio de 2012.  
Aprobado en su forma definitiva:  
6 de noviembre de 2012

---

**Ms.C. Yuniel Eliades Proenza Arias**  
Universidad de las Ciencias Informáticas (UCI)  
País: Cuba  
Correo electrónico: <yproenza@uci.cu>

**Ing. Alianis Pérez Sosa**  
Universidad de las Ciencias Informáticas (UCI)  
País: Cuba  
Correo electrónico: <alianis.perez@etecsa.cu>

---