

## Modelo en OWL para el sistema clasificador computacional 2012 de la ACM

Everardo Carlos Guevara Hernández.  
 María Auxilio Medina.  
 Jorge de la Calleja Mora.  
 Antonio Benítez Ruíz.  
*carlosgh@ymail.com*  
*mauxmedina@gmail.com*  
*jorgedelacalleja@gmail.com*  
*antonio.benitezruiz@gmail.com*

### Abstract

Este trabajo trata acerca del modelo codificado en OWL de la ontología desarrollada por ACM del sistema clasificador computacional versión 2012.

Las ontologías son una herramienta para inferir conocimiento, además organiza el vocabulario de un dominio, siendo funcionales para resolver problemas de clasificación. Como el problema que se pretende resolver en este trabajo, cuando un nuevo documento es agregado a una colección de documentos y este tiene que ser clasificado de manera adecuado e incluido en el lugar correcto, para identificar el lugar adecuado que tomará el nuevo documento se requiere asistir de la Taxonomía y concretamente el sistema clasificador computacional de ACM.

Para comprobar su funcionalidad de la ontología se realizaron algunas consultas sobre un servidor http Fuseki y el lenguaje de consulta SPARQL para la recuperación de información, en la parte de inferencia se utilizó en razonador Jena.

### 1. Introducción

Hoy en día estamos viviendo una revolución en las comunicaciones como nunca antes había existido; saber qué información proviene de una fuente confiable, es difícil. Una manera de asegurarlo es consultando una biblioteca digital, donde sus colecciones se encuentren validadas y almacenadas en forma electrónica. Las bibliotecas digitales utilizan sistemas de clasificación para organizar las colecciones.

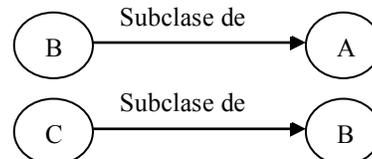
Estos sistemas son propuestos por organizaciones especializadas en un área, un ejemplo es *Association for Computing Machinery* (ACM), fundada en 1947 como la primera sociedad científica y educativa de Computación [1]. Algunas actividades que lleva a cabo esta organización es la edición de varias revistas y periódicos, patrocinar conferencias y organizar el sistema clasificador computacional (CCS).

CCS sirve para organizar documentos de las ciencias computacionales desde 1964. A diferencia de la versión 1998, CCS 2012 se organiza con una estructura poli-jerárquica.

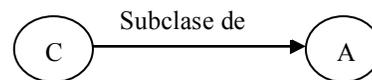
Este documento trata como CCS 2012, se le agrega un nuevo miembro a una clase, hay que entender que cuando se hace referencia a un nuevo miembro, se refiere al momento en que un nuevo documento le es añadido a la colección de documentos ya existentes, para conseguir este objetivo se codificó el CCS en OWL con la ayuda de Protégé.

Como ya se ha mencionado las ontologías pueden inferir información apoyada de las reglas semánticas que no están explícitamente indicadas en la ternas, Por ejemplo, si la clase C es una subclase de la clase B, y B una subclase de A, entonces por implicación C es una subclase A.

Ternas de información almacenadas



Terna de información inferida



En los siguientes secciones trataremos los trabajos que se han realizados teniendo como herramientas a las ontologías, además se describirá brevemente que es una ontología y sus partes, abordaremos como fue codificada la ontología basada en el CCS de ACM versión 2012 y como le fue agregado un miembro con sus respectivas notaciones heredadas, además se explica que herramientas fueron empleadas para realizar estas consultas.

## 2. Trabajos relacionados.

Algunos trabajos relevantes que han utilizado ontologías para clasificar colecciones de documentos son:

- El proyecto GALEN, financiado por la Unión Europea, cuyo objetivo es desarrollar herramientas y métodos para la elaboración y mantenimiento de clasificaciones de procedimientos quirúrgicos [2].
- La iniciativa mundial de clasificación (IMT), promovida por la ONU, que tiene como meta desarrollar un estrategia para unificar clasificadores diversos y apoyar así la implementación del Convenio sobre la Diversidad Biológica [3].

Algunos de los trabajos relevantes que utilizan CCS de ACM son los siguientes:

- La biblioteca digital de ACM (ACM DL) es una colección de artículos y registros bibliográficos que existen hoy en día, cubren los campos de la tecnología informática y de información. La base de datos incluye revistas, actas de congresos y boletines [4]. La ACM Digital Library está indexada con el sistema de clasificación computacional (CCS) versión 2012 de ACM, cuenta con servicios de alertas, tiene formato de exportación como BibTex.
- El consorcio AKT agrupa 5 universidades del Reino Unido su fin es ayudar a desarrollar la próxima generación de tecnologías [5]; unos de los proyectos AKT trata la codificación en OWL del sistema clasificador computacional de ACM en su versión 1998, esta codificación sólo se realizó para la lengua inglesa.

## 3. Ontología

Las definiciones de qué es una ontología varían, dependiendo cada autor. Según Weigand [6], “una ontología es una base de datos que describe los conceptos generales sobre un dominio, algunas de sus propiedades y cómo los conceptos se relacionan unos con otros”. Una definición propia de ontología dice que: “Especifica un vocabulario relativo a cierto dominio, formado por entidades, clases propiedades, predicados y funciones, además de la relaciones entre estos componentes.”

En Ciencias de la Computación se pueden utilizar ontologías para una multiplicidad de propósitos, incluyendo el razonamiento inductivo, problemas de clasificación, y una variedad de técnicas de resolución de problemas.

Los componentes de una ontología varían según el dominio de interés y las necesidades de los desarrolladores. Según Gruber [7], los componentes son:

**Conceptos:** son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias o procesos de razonamiento.

**Relaciones:** representan la interacción y enlace entre los conceptos de un dominio. Suelen formar la taxonomía del dominio. Algunas relaciones son: subclase de, parte de, parte exhaustiva.

**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.

**Instancias:** se utilizan para representar objetos determinados de un concepto.

**Reglas de restricción o axiomas:** son teoremas que se declaran sobre las relaciones que deben cumplir los elementos de la ontología.

Los axiomas, junto con la herencia de conceptos, permiten inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos.

## 4. Codificación de la ontología

El objetivo principal de la CCS de ACM es el clasificar documentos, para lograr este fin se propone organizarlos en las clases que más lo representen y agregarlos a esta ontología adicionando cada documento por su ISBN o una clave única como en la figura 1 se muestra.

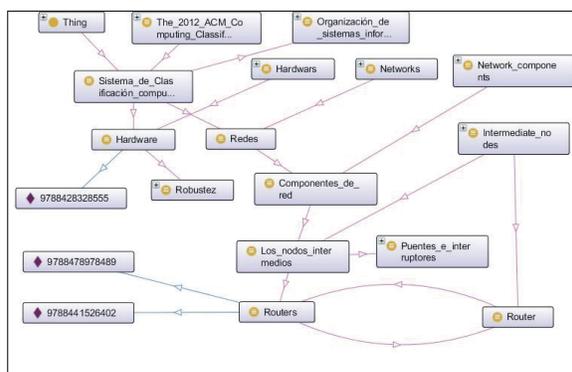


Figura 1. Miembros de una clase.

La jerarquización y la organización de las clases y sub clases se basan en el CCS de ACM la versión 2012, fue necesario codificar esta ontología ya que ACM solo proporciona el diseño estructural, pero no proporciona la ontología en ningún lenguaje con el que podamos realizar inferencia, para la codificación de la ontología se empleó Protégé desarrollado por la

Universidad de Stanford, tiene interfaz para la codificación de RFD y OWL, puede ser empleado por algún razonador lógico para realizar tareas de inferencia.

La ontología basada en CCS ver 2012, se codificó en inglés y español, a través de clases de equivalencia por ejemplo las ternas ver figura 2, <Sistema de Clasificación computacional 2012 de ACM> es equivalente a <The 2012 ACM Computing Classification System> con esto se logra usar los dos idiomas en el CCS.

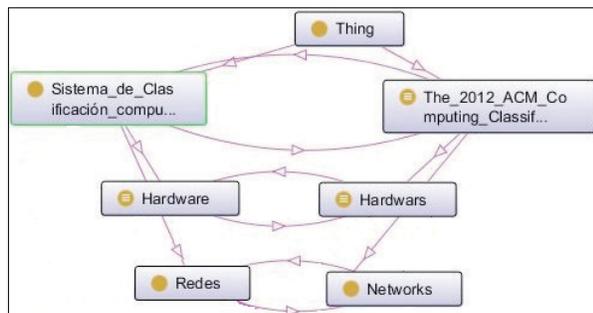


Figura 2. Equivalencias de clases

Se definieron clases disjuntas, por ejemplo, Hardware es distinta a la clase Redes con la siguiente ternas se cumple este objetivo: <Hardware> <disjointWith> <Redes>.

Para colocar una nuevo miembro en una clase, se siguen los siguiente pasos, primero se le asigna como nombre de este miembro el ISBN que es un código único como se ve en la línea 1, además se le incorporan los metadatos del libro que se moldean como anotaciones heredas como se ve en la líneas 3 a la 7, además se indica a que clases de la ontología pertenece, en este caso pertenece a Hardware como se muestra en la línea 2.

```

1. <owl:NamedIndividual rdf:about="&ont;9788428328555">
2.   <rdf:type rdf:resource="&ont;Hardware"/>
3.   <ont:DC.Title>MONTAJE, CONFIGURACIÓN Y REPARACIÓN DEL PC </ont:DC.Title>
4.   <ont:DC.Publisher>PARANINFO</ont:DC.Publisher>
5.   <ace_lexicon:PN_sg>9788428328555</ace_lexicon:PN_sg>
6.   <ont:DC.Language>Español</ont:DC.Language>
7.   <ont:DC.Creator>David Zurdo Saiz</ont:DC.Creator>
8. </owl:NamedIndividual>

```

### 5. Consultas en la ontología.

SPARQL es un acrónimo del inglés “Protocol and RDF Query Language”. Es un lenguaje estandarizado para la consulta de grafo RDF, normalizado por el

RDF Data Access Working Group (DAWG) del World Wide Web Consortium (W3C) [8].

Al igual que sucede con SQL, es necesario distinguir entre el lenguaje de consulta y el motor para el almacenamiento y recuperación de los datos. Existen múltiples implementaciones de SPARQL, generalmente ligados a entornos de desarrollo y plataforma tecnológica, en un principio SPARQL únicamente incorpora funciones para la recuperación sentencias RDF, sin embargo, algunas propuestas también incluyen operaciones como: creación, modificación y borrado de datos.

En este trabajo describen consultas de ejemplo para mostrar la utilidad de la ontología construida:

**Consulta 1:** información de un libro que tiene como código de ISBN 9788428328555.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prop: <http://www.co-ode.org/ontologies/ont.owl#>

```

Este grupo de prefijos, encabezado, cabeceras o URIs son importante porque señalan la ubicación donde se encuentran la entidad de a consultar, con la palabra PREFIX nos permiten abreviar el lugar donde se encuentra la entidad, en este trabajo usaremos esta cabeceras para todas las consultas.

```

SELECT ?Propiedades
WHERE {prop:9788428328555 ?p ?Propiedades.
FILTER (?p != rdf:type)}

```

Resultados:

Propiedades
"MONTAJE, CONFIGURACIÓN Y REPARACIÓN DEL PC"
"PARANINFO"
"9788428328555"
"Español"
"David Zurdo Saiz"

**Consulta 2:** Recupera todos los libros que pertenecen a la clase Routers.

```

Cabecera de prefijos
SELECT ?ISBN ?Propiedades
WHERE {{?ISBN rdf:type prop:Routers.}
optional {?ISBN ?p ?Propiedades.}
FILTER (?p != rdf:type && ?p != lexico:PN_sg) }
}

```

En los resultados de esta consulta se agregó el campo metadatos sólo para poder dar una visión de lo que entrega la búsqueda, el servidor no entregará esa columna y el filtro que incluye esta búsqueda quita

ternas que no son importantes al ser presentada en la pantalla.

Resultados:

ISBN	Propiedades	Metadatos
9788441526402	"ROUTERS CISCO: EDICION REVISADA Y ACTUALIZADA 2010 (GUIA PRACTICA)"	Título
9788441526402	"ANTONIO GALLEGO DE TORRES"	Autor
9788441526402	"ANAYA MULTIMEDIA"	Editorial
9788441526402	"Español"	Idioma
9788478978489	"TÉCNICAS DE CONFIGURACIÓN DE ROUTERS CISCO"	Título
9788478978489	"ERNESTO ARIGANELLO"	Autor
9788478978489	"AG Canarias"	Editorial
9788478978489	"Español"	Idioma

**Consulta 3:** Mostrar los nombre de las clases a las que pertenece un libro que tiene como ISBN 9788441526402.

Cabecera de prefijos

SELECT \*

WHERE {?ISBN rdf:type ?Clase.

FILTER (?ISBN = prop:9788441526402)

?Clase rdf:type owl:Class.

optional{?Clase rdfs:subClassOf ?Ancestro1.}

optional{?Ancestro1 rdfs:subClassOf ?Ancestro2.}

optional{?Ancestro2 rdfs:subClassOf ?Ancestro3.}

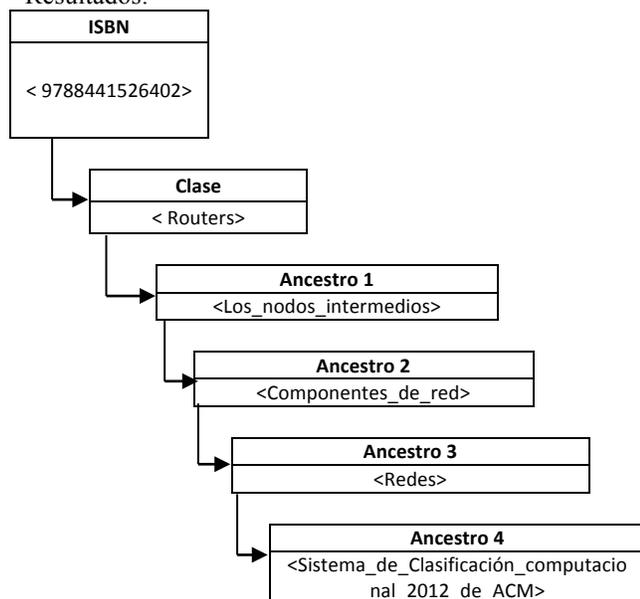
optional{?Ancestro3 rdfs:subClassOf ?Ancestro4.}

optional{?Ancestro4 rdfs:subClassOf ?Ancestro5.}

}

Cabe señalar que las consultas se realizan por ternas es por esta razón que es necesario la clase hijo el conector para obtener la clase ancestro y de esta manera inferir los antecesores.

Resultados:



## 6. Herramientas de consulta.

Existen herramientas que ayudan a representar lo que existe en el mundo real, cimentándose en la sintaxis de la lógica descriptiva; mediante estas herramientas se establecen las reglas y los predicados que son las bases para el razonamiento y la inferencia para la representación del conocimiento.

Las herramientas que se utilizaron es este proyecto son: Protégé editor para RDF y OWL. OWL es un lenguaje de definición de diccionarios semánticos creado por el W3C, OWL o “Lenguaje de Ontologías Web” fue desarrollado para facilitar el procesamiento en aplicaciones que necesitan interpretar e inferir el contenido de la información de algún documento, ya que esta habilidad sólo es propia de los seres humanos. OWL interpreta el contenido web con mayor precisión XML, RDF, y esquema RDF (RDF-S), proporciona vocabulario adicional junto con una semántica formal. OWL se puede formular en RDF, incluye toda la capacidad expresiva de RDF y RDF-S y la extiende con la posibilidad de utilizar expresiones lógicas.

Jena soporta los formatos RDF/XML, N3, N-Triples y OWL. En Jena se almacena la información como tripletas o ternas de RDF en grafos dirigidos esto permite que su código pueda agregar, eliminar, manipular, almacenar y publicar esa información [8]. Jena proporciona los medios para que estas ternas que contienen información sean inferidas de manera transparente al usuario, Fuseki es el servidor el que se realizan las consultas utilizando SPARQL, otra herramienta es Protégé mediante el cual se codifica la ontología en lenguaje RDF y OWL.

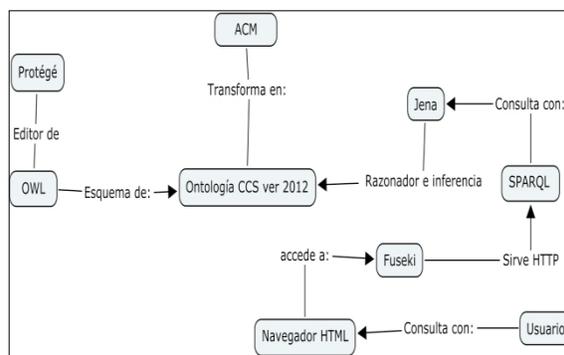


Figura 3. Mapa de transferencia de información

Fuseki es un servidor de SPARQL HTTP que proporciona la facilidad de realizar alta, consulta y

actualizaciones, utilizando el protocolo SPARQL a través de HTTP.

## 7. Conclusiones

En este trabajo se concluye, que la forma correcta de agregar un nuevo documento ya clasificado a la ontología, es agregando un miembro de una clase, este miembro puede pertenecer a varias clases, además con las herramientas de inferencia podemos obtener información que no se encuentra explícitamente capturada en las ternas, pero puede ser inferida de la información que ya se tiene almacenada, además las búsquedas pueden ser diseñadas por medio de SPARQL.

Jena es una buena herramienta de inferencia ya que esta aplicación es transparente en su uso, en combinación con Fuseki, ambas aplicaciones proporcionan una plataforma de trabajo amigable la cuales facilita las tareas de recuperación de información.

## 8. Referencias

- [1] ACM. 2012. "Sistema de clasificación computacional de 2012". En línea [http://delivery.acm.org/10.1145/2380000/2371137/ACMCCSTaxonomy.html?ip=187.186.146.173&ac\\_c=OPEN&CFID=198433416&CFTOKEN=82154632&\\_\\_ac\\_m\\_\\_=1352663893\\_7854295f3b5bb1d8e02051c84075ca0.12](http://delivery.acm.org/10.1145/2380000/2371137/ACMCCSTaxonomy.html?ip=187.186.146.173&ac_c=OPEN&CFID=198433416&CFTOKEN=82154632&__ac_m__=1352663893_7854295f3b5bb1d8e02051c84075ca0.12) septiembre 2012.
- [2] Nardi, D. & Brachman, R. J. 2002. *An Introduction to Description Logics*. Cambridge University Press, Cambridge, Reino Unido :
- [3] Naciones Unidas, 2012, "Iniciativa Mundial sobre Taxonomía". <http://www.cbd.int/gti/> Naciones Unidad, Montreal Canadá
- [4] ACM.2012. "Digital library". <http://dl.acm.org/dl.cfm> septiembre 2012.
- [5] ATK. 2012. "Historia de la organización" En línea <http://www.aktors.org/akt/>. Advanced Knowledge Technologies. Departamento de Electrónica y Ciencias de la Computación de la Universidad de Southampton, Reino Unido, Agosto 2012.
- [6] Weigand, H., 1998. "Multilingual Ontology-based Lexicon For News Filtering the TREVI Project". Reporte técnico pp.138-159, Infolab, Tilburg University,. Tilburg, Netherlands.
- [7] Gruber, T. R. 1993. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". En línea <http://www.google.com.mx/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCQQFjAA&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.91.6025%26rep%3Drep1%26type%3Dpdf&ei=qIVyUN6pLcSfqwGYq4HoDw&usg=AFQjCNH6E4XzTsjjmWIRRePKNA2vGSY0Dw&s>, 7 Octubre 2012.
- [8] Prud'hommeaux E, Seaborne A, 2008 "Lenguaje de Consulta SPARQL para RDF" en línea: <http://www.w3.org/TR/rdf-sparql-query/>, W3C, 7 febrero 2013.
- [9] Apache Software Foundation, 2011. "Jena arquitectura visión general". En línea [http://jena.apache.org/about\\_jena/architecture.html](http://jena.apache.org/about_jena/architecture.html). 7 febrero 2013.