



Universidad de las Ciencias Informáticas
Facultad 1
Centro de Informatización Universitaria

“Propuesta de modelo semántico para un gestor de documentos históricos”

Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas

Autora

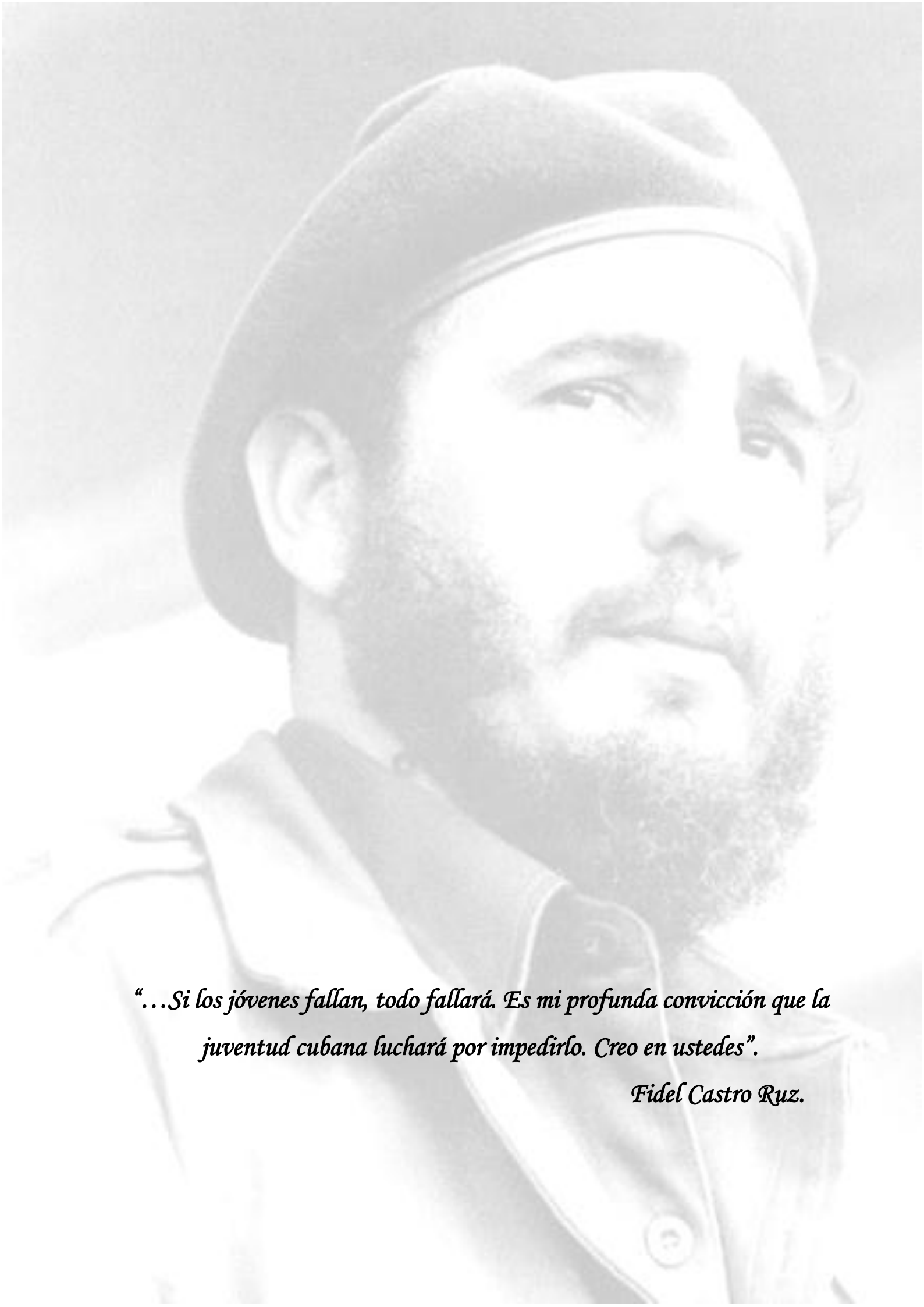
Mavis García Rodríguez

Tutoras

MSc. Laritza Cabrera Barroso

Ing. Arlennys Susana Velázquez Hidalgo

La Habana, Cuba
Junio, 2012



“...Si los jóvenes fallan, todo fallará. Es mi profunda convicción que la juventud cubana luchará por impedirlo. Creo en ustedes”.

Fidel Castro Ruz.

Declaración de autoría

Declaro que soy la única autora del trabajo titulado **“Propuesta de modelo semántico para un gestor de documentos históricos”** y autorizo al Centro de Informatización Universitaria de la Universidad de las Ciencias Informáticas; así como a dicho Centro para que hagan uso del mismo en su beneficio.

Para que así conste firmo la presente a los ____ días del mes de _____ del año ____2012____.

Mavis García Rodríguez

Laritza Cabrera Barroso

Arlennys Susana Velázquez Hidalgo

Firma del autor

Firma del tutor

Firma del tutor

A mi mamá por darme su amor, por estar siempre a mi lado en los momentos buenos y malos, por su apoyo y comprensión durante todos estos años de sacrificio, por darme fuerzas y ánimo para seguir adelante con mi sueño que hoy se hace realidad.

A mis abuelos Eva y Raúl por su apoyo, sus consejos y brindarme todo su cariño.

A mi tía Niana, a mi tío Mansito por su preocupación, por siempre darme fuerzas para triunfar en la vida.

A mis primos Nianet, Vonhan por su cariño.

A una persona muy especial Juan (mi papá adoptivo) que siempre se preocupa por mí, por toda su dedicación y ayuda muchas gracias.

A mi novio Pavel por todo su amor, cariño, paciencia y comprensión todos estos meses, por estar a mi lado y por todo lo que ha hecho por mí, te amo.

A la familia de mi novio por acogerme como uno más de su familia a Ampari y Juan por todo su apoyo durante este tiempo.

A toda mi familia en general por darme su cariño y fuerza para ser una profesional.

A mi tutora Laritza porque sin ella nada hubiese sido posible, por su paciencia, por demostrarme que siempre podía contar con ella, por ser una amiga gracias de corazón.

A todas mis amistades, los que están aquí hoy y los que ya no están por alguna razón, a todos los que de alguna forma u otra me han ayudado en estos cinco años.

A todos y cada uno de ustedes, Muchas Gracias.

A mi mamá porque es mi razón de ser, gracias a su sacrificio, amor y entrega hoy soy alguien en la vida.

A toda mi familia por su cariño, preocupación y ayuda todos estos años.

En la archivística como en otras áreas de la ciencia, resulta de gran importancia la existencia de estándares, tanto para facilitar el intercambio de información entre entidades como para normalizar el acceso a dicha información. Un ejemplo de las normas vigentes para la elaboración de descripciones archivísticas es la Norma Internacional General de Descripción Archivística ISAD (G) y la Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias ISAAR (CPF).

El surgimiento y desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC), ha modificado totalmente la experiencia acumulada durante años de trabajo en los aspectos más importantes de la creación, distribución, almacenamiento, modo de acceso y representación de los recursos de información archivística.

En los últimos años novedosas técnicas han sido desarrolladas con el fin de mejorar los mecanismos de búsquedas y recuperación de la información. La Web Semántica es un fenómeno que propone nuevos paradigmas para la representación de la información y el conocimiento que faciliten la localización, compartición, integración y recuperación de recursos a través de la web.

En el presente trabajo de diploma se realiza un estudio de los principales aportes de las tecnologías de la Web Semántica a la gestión archivística y a partir de ello se propone la aplicación de dichas tecnologías en el desarrollo del sistema de gestión de documentos históricos, deXcriba.

Palabras Clave: archivo, Archivística, documento de archivo, gestión de documentos de archivo, ontología, Web Semántica.

Abstract.

In the archives as in other areas of science, it is very important the existence of standards, both to facilitate the exchange of information between agencies and to standardize access to such information. An example of existing standards for the preparation of archival descriptions is the General International Standard Archival Description ISAD (G) and the International Standard on Archival Authority Records relating to Corporate Bodies, Persons and Families ISAAR (CPF).

The rise and development of Information and Communications Technologies and Communications (TIC) has totally changed the accumulated experience during years of work in the most important aspects of the creation, distribution, storage, access mode and representation of archives information resources. In recent years, new techniques have been developed to improve the mechanisms for searching and retrieving information. The Semantic Web is a phenomenon that proposes new

paradigms for representing information and knowledge to facilitate the location, sharing, integration and recovery of resources through the web.

This diploma work is a study of the major contributions of Semantic Web technologies to archival management and from this we propose the application of these technologies in the development of the management system of historical documents, deXcriba.

Keywords: File, archival, document archiving, records management, ontology, Semantic Web.

Introducción	1
Fundamentación Teórica	6
1.1 Conceptos y definiciones.....	6
1.2 Normas y estándares archivísticos.....	7
1.2.1 Descripciones de las normas y estándares que rigen los procesos de descripción archivística.....	7
1.2.2 Norma ISAD (G).....	7
1.2.3 Norma ISAAR (CPF).....	8
1.2.3 Norma ISDIAH.....	9
1.2.5 Estándar EAD.....	10
1.2.6 Estándar EAC.....	11
1.3 Técnicas de recuperación de la información en la archivística.....	12
1.3.1 Índices.....	12
1.3.2 Palabras clave.....	12
1.3.3 Tesauros.....	12
1.4 Mecanismos de búsqueda y recuperación de la información en los sistemas de gestión archivística.....	13
1.4.1 Archivo3000.....	13
1.4.2 PARES.....	14
1.4.3 ICA-AtoM.....	14
1.4.4 Albalá.....	15
1.4.5 ArchiVenHIS.....	15
Aplicación de las tecnologías de la Web Semántica a la gestión archivística	17
2.1 Tecnologías de la Web Semántica.....	18
2.2 Aplicación de las tecnologías de la Web Semántica a la gestión archivística.....	21
2.2.1 Ontología para la Norma ISAD (G).....	22
2.2.2 Ontología para ISAAR (CPF).....	24
Herramientas para el modelado de ontologías	28
3.1 Redes semánticas.....	28
3.2 Sistemas basados en marcos.....	28
3.3 Sistemas basados en lógicas descriptivas.....	29
3.4 Tareas de razonamiento.....	29
3.4.1 Pellet.....	29
3.4.2 FaCT++.....	30
3.4.3 Racer.....	30
3.5 Herramientas para el diseño de ontologías.....	30
3.5.1 Protégé.....	31
3.5.2 OntoEdit.....	31
3.5.3 SWOOP.....	31
3.6 Validación de la propuesta.....	31
Conclusiones	34
Recomendaciones	35
Referencias bibliográficas	36
Bibliografía consultada	39
Glosario de términos	41
Anexos	42

La historia de la evolución humana ha podido construirse gracias a las salvaguardas de la información en disímiles soportes a través del tiempo. Los espacios físicos destinados al almacenamiento de los documentos, tomaron diferentes nombres en dependencia de las culturas y de las funciones para los cuales fueron creadas. Se les preservaron de diversas maneras y en disímiles lugares con un fin determinado, pero sin una concepción preestablecida. En la era moderna aparece una ciencia capaz de regir la organización, conservación y gestión de los documentos, la Archivística [1].

Los Archivos Históricos son los más conocidos dentro del Sistema Archivístico Institucional (SAI), por su larga vida al servicio de la investigación y la cultura. Constituyen la garantía de la evolución jurídica y administrativa de la sociedad, son la memoria colectiva de una nación, región o localidad. Su finalidad es gestionar, atesorar, conservar y difundir el patrimonio documental, pueden almacenar documentos históricos recibidos por donación, depósito, transferencia y adquisición. Los usuarios por excelencia de este tipo de archivo son los investigadores interesados en el valor secundario que han adquirido los documentos al final de su ciclo de vida [1].

Antes de la aparición de los sistemas modernos, la gestión de documentos de archivo se hacía generalmente de dos formas: mediante agrupaciones temáticas y la creación de fichas descriptivas. Las agrupaciones temáticas ubicaban los documentos por fechas, nombres y tipos. Con el paso de los años traía como resultado la acumulación de una cantidad indiscriminada de documentos, sin contar el permanente deterioro de los mismos por los insectos y los factores ambientales. Con la creación de fichas descriptivas se garantiza que debían existir una por cada documento y contenían una serie de datos específicos del documento en cuestión: autor, año de publicación, título, así como su localización física dentro del local. Este mecanismo simplificaba la búsqueda ya que el investigador podía auxiliarse de una breve descripción del documento sin tener que llegar a él. Pero aún así ambas variantes padecían de los mismos problemas: pérdida de los originales, gasto de tiempo y esfuerzo excesivo durante las búsquedas y recuperación de la información, que no pocas veces eran infructuosas.

Los primeros sistemas automatizados de recuperación de la información fueron presentados durante la década de 1950 a 1960. A partir de 1950 fueron creadas las primeras automatizaciones de los catálogos, se requería de equipos de alta velocidad para imprimir grandes volúmenes de información. En 1958, se creó el primer sistema automático de asignación de índices, índice KWIC¹ [2]. El

¹ **Índice KWIC.** Tipo de índice permutado en el que el contenido temático de una obra se representa mediante palabras clave de su título o de otra fuente de información del documento.

surgimiento y desarrollo de Internet ha modificado totalmente la experiencia acumulada durante años de trabajo en los aspectos más importantes de la creación, distribución, almacenamiento, modo de acceso y representación de los recursos de información archivística. Sin duda, los métodos más rápidos y eficaces de recuperación de documentos de la *World Wide Web* son los que proporciona la propia web a través de los buscadores [3]. Los más habituales son los índices temáticos, que pueden ser catálogos o buscadores por categoría en línea y los motores de búsqueda que son buscadores por contenido.

Un sistema de gestión de documentos de archivo informatizado no será útil si no es capaz de brindar los documentos con su información en el momento que se necesite, por lo que el requisito de búsqueda es uno de los más importantes a tener en consideración. En este sentido, se debe exigir que el sistema cuente con un mecanismo de recuperación de información eficiente.

El departamento Gestión documental del Centro de Informatización Universitaria (CENIA), tiene como misión, el desarrollo de *software* para la informatización de los procesos que define la gestión documental y archivística como ciencia de la información. ArchiVenHIS es un sistema especialmente diseñado para la gestión de documentos históricos en el Archivo General de la Nación (AGN) de la República Bolivariana de Venezuela. A partir de esta experiencia se pretende desarrollar un sistema que permita el tratamiento archivístico en cualquier institución y que posea mecanismos de búsqueda robustos que faciliten el acceso y recuperación de la información.

En los sistemas de gestión documental de archivos estudiados, principalmente ArchiVenHIS que ha sido tomado como punto de referencia, los buscadores o motores de búsqueda, están enfocados principalmente en efectuar búsquedas sintácticas. La búsqueda sintáctica se basa en realizar una búsqueda de las palabras o múltiples frases que aparecen de los documentos archivados, que coinciden sintácticamente con los criterios de búsquedas especificados, esta tiene como desventaja que deja de indexar información estrechamente relacionadas con la palabra o frase que se ha especificado, debido a que no es posible mediante este mecanismo asociar contenidos a partir de su semántica o significado, por lo que no devuelve todas las semejanzas posibles.

Esto conlleva a la formulación del siguiente **problema científico**: ¿cómo mejorar los procesos de acceso y recuperación de información en los sistemas de gestión de documentos históricos? siendo el **objeto de estudio**: los mecanismos de búsqueda y recuperación de la información en la web, enmarcado como **campo de acción**: los mecanismos de búsqueda y recuperación de la información en los sistemas web de gestión archivística.

La web hasta hace poco tiempo atrás procesaba los contenidos de forma sintáctica relacionando y rescatando significados de forma individual mediante palabras claves. Sin embargo las múltiples relaciones que tienen las palabras permiten generar diferentes significados que son mucho más complejos que las palabras por sí solas. La web actual posee una gran capacidad para almacenar datos y puede leer y visualizar los contenidos, pero no es capaz de pensar ni de entender todo lo que contiene. Se precisa, por lo tanto una nueva web la *Web Semántica*² como mecanismo capaz de almacenar no sólo los datos, sino de entender e interpretar el sentido de la información [4].

Por lo cual la presente investigación tiene como **objetivo general**: realizar una propuesta de aplicación de las tecnologías de la Web Semántica en el desarrollo del sistema de gestión de documentos históricos, con el propósito de optimizar la búsqueda y recuperación de la información.

Del objetivo general se derivan los siguientes **objetivos específicos**:

- Realizar un estudio de los principales conceptos y definiciones del área de la gestión documental de archivos así como los principales términos asociados a la Web Semántica.
- Realizar un informe del estado del arte de las principales propuestas de aplicación de tecnologías de la Web Semántica en la gestión documental de archivos.
- Proponer vocabularios semánticos y metodologías para la creación y mantenimiento de ontologías web, aplicadas a la herramienta de gestión de documentos históricos.
- Realizar una propuesta de entorno de trabajo para el diseño del modelo ontológico propuesto.

En aras de dar respuesta a dichos objetivos específicos se proponen las siguientes **tareas de la investigación**:

- Realización de la fundamentación teórica de los principales conceptos y definiciones asociados a la gestión documental de archivos y la Web Semántica.
- Evaluación de las tendencias actuales respecto al uso de tecnologías de la Web Semántica en la gestión documental de archivos.
- Investigación sobre las principales normas y estándares que rigen los procesos de descripción archivística.

² **Web Semántica.** Vendría a ser una extensión de la web actual dotada de significado, de manera que pudiera ser interpretada tanto por agentes humanos como por agentes computarizados.

- Realización de un estudio sobre vocabularios semánticos y metodologías para la creación y mantenimiento de ontologías web.
- Realización de una propuesta de ontología, basada en las normas para la descripción de documentos de archivo y ontologías específicas del dominio de trabajo de los archivos.
- Evaluación de las principales herramientas utilizadas para el diseño del modelo ontológico.
- Realización de una propuesta con la herramienta seleccionada para el diseño de la ontología.

Para dar cumplimiento a las tareas planteadas se utilizarán los siguientes métodos de investigación:

Métodos teóricos.

- ✓ Analítico – sintético:

Este método permite extraer la esencia a partir de la bibliografía consultada acerca de la gestión documental de archivos, los sistemas de gestión archivística más referenciados, así como del acceso y recuperación de información en los mismos, favoreciéndose el entendimiento y la claridad de los conceptos implicados en la investigación y las ideas expresadas en este documento, posibilitando un alto nivel de concreción y objetividad en la información expuesta.

- ✓ Histórico – lógico:

Permite el estudio de los antecedentes de las técnicas de recuperación de información en la gestión documental de archivos, de manera que fuesen extraídos los elementos necesarios para la comprensión de la investigación realizada.

Métodos empíricos.

- ✓ Observación:

A través de este método se realizaron pruebas a los sistemas implementados para la gestión archivística y se observaron las dificultades que estas puedan presentar en el proceso de búsqueda y recuperación de la información.

Justificación de la investigación.

La aplicación de las tecnologías de la Web Semántica en el desarrollo del sistema gestor de documentos históricos, contribuirá a mejorar el acceso y recuperación de la información. El uso de

vocabularios semánticos proporciona una descripción y organización de conceptos que representan el marco de trabajo conceptual para el análisis, discusión y consulta de información en el dominio de la gestión de documentos de archivo. La descripción semántica de los contenidos dotará a la información de un significado bien definido, pudiendo razonar e inferir nuevo conocimiento o resolver problemas análogamente como lo haría un experto en el área. La utilización de esta nueva filosofía de la web actual en los sistemas gestores de documentos históricos, posibilita la implementación de mecanismos de búsqueda más eficientes, que le permiten al investigador el acceso rápido y preciso a la información deseada.

Estructura del documento.

El documento estará estructurado en 3 capítulos, a continuación se presenta una breve descripción del contenido que se abordará en cada uno:

Capítulo1: Fundamentación teórica.

Contiene los conceptos relacionados a la gestión documental de archivos, las normas y los estándares que se utilizan en la descripción archivística, las técnicas de recuperación de la información en la web, así como el estudio de la implementación de dichas técnicas en los sistemas de gestión archivística.

Capítulo2: Propuesta de aplicación de las tecnologías de la Web Semántica a la gestión archivística.

Contiene toda la información relacionada con las tecnologías de la Web Semántica, arquitectura, lenguajes y estándares para la representación del conocimiento. Se describen novedosas propuestas encontradas dirigidas a este dominio y se concluye con la valoración de la utilización de la misma en el sistema de gestión de documentos históricos.

Capítulo3: Herramientas para el modelado de ontologías.

Contiene la información relacionada con los paradigmas de representación del conocimiento, las herramientas para el diseño, implementación, integración y anotación de ontologías, así como los razonadores utilizados por dichas herramientas. Se concluye con la propuesta del entorno de trabajo para el desarrollo del modelo ontológico propuesto.

Capítulo 1

Fundamentación teórica.

En el presente capítulo se exponen los conceptos y definiciones fundamentales asociados a la gestión documental de archivos, principales normas y estándares de la archivística, así como las técnicas empleadas en la recuperación de la información. Se evalúan las tendencias actuales respecto a los mecanismos de búsqueda implementados en los sistemas afines a la gestión archivística.

1.1 Conceptos y definiciones.

La archivera española Antonia Heredia, una de las teóricas de la archivística que más ha influido en los últimos tiempos en Cuba, explica que: **archivo** es uno o más conjuntos de documentos, sea cual sea su fecha, su forma y soporte material acumulados en un proceso natural por una persona o institución pública o privada en el transcurso de su gestión, conservados, para servir como testimonio en formación para la persona o institución que los produce, para los ciudadanos o para servir de fuente de historia [5], a cada ejemplar de este conjunto de documentos se les denomina, documento de archivo.

Un **documento de archivo** presenta particularidades que son las que realmente otorgan las singularidades de los archivos. Se define como “el testimonio material de un hecho o acto realizado en el ejercicio de sus funciones por personas físicas o jurídicas, públicas o privadas, de acuerdo con unas características de tipo material y formal” [5].

La **gestión documental** es “el conjunto de operaciones comprometidas en la búsqueda de la economía y la eficacia en la producción, mantenimiento, uso y destino final de los documentos de archivo” [6], esta no puede circunscribirse solo a la primera etapa de la vida de los documentos, la etapa administrativa. Es preciso abordar, entonces, la gestión documental como una función esencialmente archivística, que cubre todas las etapas de la vida de los documentos y que involucra todas las fases de tratamiento de dichos documentos.

Uno de los procesos fundamentales en la gestión documental es el tratamiento de los documentos mediante la **descripción archivística**. “La descripción archivística es el medio utilizado por el archivero para obtener información contenida en los documentos y ofrecerla a los interesados en ella”

[5]. La descripción de los documentos debe ser concreta, breve y contener las ideas básicas; son realizadas por los archivistas a través de instrumentos descriptivos como guías, inventarios, índices y catálogos. Un instrumento descriptivo que ha ganado gran aceptación en el campo archivístico son los estándares y normas.

1.2 Normas y estándares archivísticos.

Los estándares y normas son descripciones técnicas detalladas, elaboradas con el fin de garantizar la interoperabilidad entre elementos construidos independientemente, así como la capacidad de replicar un mismo elemento de manera sistemática [7]. Las normas contienen, criterios precisos que aseguran que los materiales, productos, procesos y servicios están hechos con la calidad necesaria para alcanzar sus objetivos. Contribuyen a hacer la vida más simple y a incrementar la fiabilidad y efectividad de los bienes y servicios que se utilizan.

El surgimiento de las normas y los estándares archivísticos han permitido que hoy en día los archivos de todo el mundo distribuyan la información sobre sus fondos con facilidad. Estos aseguran una enorme mejora en cuanto a la facilidad de acceso, permiten la interconexión entre sistemas locales ya que se utiliza el mismo formato para la entrada de datos informativos y favorecen el intercambio internacional.

1.2.1 Descripciones de las normas y estándares que rigen los procesos de descripción archivística.

1.2.2 Norma ISAD (G).

La norma ISAD (G) (Norma Internacional General de Descripción Archivística) es una normativa internacional para la descripción de documentos. Contiene reglas generales para la descripción archivística que pueden aplicarse con independencia del tipo documental o del soporte físico de los documentos y forma parte de un conjunto dirigido a:

- ✓ Garantizar la elaboración de descripciones coherentes, pertinentes y explícitas. Facilitar la recuperación y el intercambio de información sobre los documentos de archivo.
- ✓ Compartir los datos de autoridad.
- ✓ Hacer posible la integración de las descripciones procedentes de distintos lugares en un sistema unificado de información [8].

La ISAD (G) se estructura en siete áreas de información y establece veintiséis elementos que pueden ser combinados, para constituir la descripción de cualquier unidad:

- ✓ **Área de identificación:** Contiene la información esencial para identificar la unidad de descripción.
- ✓ **Área de contexto:** Contiene la información relativa al origen y custodia de la unidad de descripción.
- ✓ **Área de contenido y estructura:** Contiene la información relativa al objeto y organización de la unidad de descripción.
- ✓ **Área de acceso y utilización:** Contiene información relativa a la accesibilidad de la unidad de descripción.
- ✓ **Área de documentación asociada:** Contiene información relativa a aquellos documentos que tienen una relación significativa con la unidad de descripción.
- ✓ **Área de notas:** Contiene información especial y aquella que no ha podido incluirse en ninguna de las demás áreas.
- ✓ **Área de control de la descripción:** Contiene información relativa al cómo, cuándo y quién ha elaborado la descripción archivística [8].

Esta norma propone determinadas reglas de descripción multinivel que se basan en realizar la descripción de un fondo y todos los componentes implícitos en él, series, secciones, expedientes y para su correcta aplicación propone cuatro reglas generales: regla de descripción de lo general a lo particular, regla de información pertinente para el nivel de descripción, regla de vinculación de las descripciones y regla de no repetición de la información [8].

Las reglas de la ISAD (G) pretenden ser aplicables del modo más amplio posible a las descripciones de archivos, sin tener en cuenta el volumen y la naturaleza de la unidad de descripción. La norma no define los formatos de edición ni el modo de presentación de estos elementos en los instrumentos de consulta (guías, inventarios, catálogos). La estructura y el contenido de la información de cada uno de los elementos deberán formularse de acuerdo con las normas nacionales aplicables en cada país.

1.2.3 Norma ISAAR (CPF).

La ISAAR (CPF) (Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias) sirve de guía para elaborar registros de autoridad de Archivos que proporcionan descripciones de entidades (instituciones, personas y familias) asociadas a la producción y a la gestión de archivos. Se utiliza para:

- ✓ Describir una institución, persona o familia como unidades dentro de un sistema de descripción archivístico.
- ✓ Controlar la creación y utilización de los puntos de acceso en las descripciones archivísticas.
- ✓ Documentar las relaciones entre diferentes productores de documentos y entre estas entidades y los documentos creados por ellas u otros recursos que les conciernen [9].

El objetivo primordial de esta norma es sentar las reglas generales para la normalización de las descripciones archivísticas relativas a los productores de documentos de archivo y al contexto de su producción.

ISAAR (CPF) define 4 áreas para describir la autoría de un recurso:

- ✓ **Área de identificación:** en la que se incluye la información que identifica unívocamente a la entidad que se está describiendo y que define un punto de acceso normalizado.
- ✓ **Área de descripción:** en la que se proporciona información pertinente sobre la naturaleza, contexto y actividades de la entidad que se está describiendo.
- ✓ **Área de relaciones:** en la que se consignan y describen las relaciones con otras instituciones, personas y/o familias).
- ✓ **Área de control:** en la que se identifica de forma unívoca el registro de autoridad y se incluye la información necesaria sobre cómo, cuándo, por qué agencia se creó y actualizó el registro de autoridad [9].

Dentro de estas 4 áreas se definen 27 elementos para describir los registros de autoría. Los elementos pueden variar dependiendo del autor del recurso: una corporación, una persona o una familia.

Esta norma está concebida para que los registros de autoridad de archivos se puedan compartir, fomentando la elaboración de descripciones consistentes, adecuadas y autoexplicativas de las instituciones, personas y familias que producen los documentos [9].

1.2.4 Norma ISDIAH.

La ISDIAH (Norma internacional para describir instituciones que custodian fondos de Archivo) proporciona las reglas generales para la normalización de la descripción de instituciones que conservan fondos de Archivo. El principal objetivo de esta norma es facilitar la descripción de aquellas instituciones cuya función primordial es la conservación de los documentos de archivo y de su difusión al público en general [10].

Estas descripciones pueden utilizarse para:

- ✓ Describir instituciones como unidades dentro de un sistema de descripción archivística.
- ✓ Servir como punto de acceso normalizado a las instituciones que custodian los fondos de archivo, en el seno de un directorio, en un sistema de información archivística o en la red.
- ✓ Documentar las relaciones entre las instituciones y entre esas entidades y los documentos de archivo que custodian.

Esta norma se organiza en seis áreas de información **[10]**:

- ✓ Área de identificación.
- ✓ Área de contacto.
- ✓ Área de descripción.
- ✓ Área de acceso.
- ✓ Área de servicios.
- ✓ Área de control.

La mayoría de los elementos que figuran en la descripción de una institución que custodia fondos de archivo pueden utilizarse como puntos de acceso a la descripción. Esta norma aborda sólo una parte de los requisitos necesarios para el intercambio de información de instituciones detectoras de fondos de Archivo. El éxito del intercambio automatizado de la información depende de la adopción de un formato de comunicación adecuado para todos los servicios de archivo que participen en dicho intercambio **[10]**.

1.2.5 Estándar EAD.

La EAD (Descripción Archivística Codificada) es una estructura de datos normalizada que reproduce en formato digital los instrumentos de descripción archivística. No es una norma, sino una serie de reglas, que se utilizan para diseñar las partes intelectuales y físicas de instrumentos de descripción archivística, para que la información contenida pueda ser buscada, recuperada, visualizada e intercambiada de manera predecible e independiente de la plataforma **[11]**.

Sin codificación estandarizada, es difícil de manejar la indexación, visualización y manipulación de Archivos. EAD ofrece una gran flexibilidad en cuanto a la codificación de un instrumento de descripción. Es un estándar para codificar instrumentos de descripción archivística por medio de SGML (Estándar de Lenguaje de Mercado Generalizado) y XML (Lenguaje de Marcas Extensible). Hay tres

documentos técnicos asociados a EAD:

- ✓ La DTD (Definición de Tipo de Documento): EAD DTD, adaptable a SGML o XML, con material complementario como archivos de entidades opcionales.
- ✓ La biblioteca de etiquetas: documento de referencia sobre cada uno de los elementos y atributos definidos en EAD, con apéndices útiles como pasarelas (*crosswalks*) a otros estándares (ISAD (G)).
- ✓ Las directrices de aplicación [11].

A un nivel muy básico, un documento "instrumento de descripción" codificado utilizando EAD, consta de tres segmentos: uno que proporciona información sobre el instrumento de descripción en sí mismo (su título, compilador, fecha de compilación); un segundo componente que incluye las cuestiones preliminares necesarias para la publicación formal del instrumento de descripción y un tercero que proporciona la descripción del material archivístico en sí mismo, además de la información contextual y administrativa asociada. Permite incluir, información suplementaria opcional que no describe directamente los registros pero facilita su uso por parte de los investigadores [11].

1.2.6 Estándar EAC.

EAC (Contexto Archivístico Codificado) es un estándar para estructurar e intercambiar información sobre los creadores de materiales Archivísticos [12]. Los registros que pueden ser creados utilizando EAC trabajan de una manera similar a los archivos de autoridad de las bibliotecas, pero con énfasis en la provisión de historias biográficas o administrativas para personas u organizaciones interesadas. Para una interpretación exacta de los registros de un archivo es necesaria la descripción de las circunstancias que rodearon su creación y uso. Con esta información, los usuarios pueden entender los archivos completamente, una vez que sabrán el contexto en el cual la organización o persona operó y creó los archivos. La denominación EAC fue utilizada con el objetivo de enfatizar las relaciones con EAD y técnicamente, constituye una DTD de XML y SGML que, en este caso regula la codificación de la información sobre el contexto archivístico, proporcionando un formato electrónico para los registros de autoridad ajustados a la ISAAR(CPF) [12].

1.3 Técnicas de recuperación de la información en la archivística.

El proceso de recuperación se lleva a cabo generalmente mediante consultas a la base de datos donde se almacena la información estructurada, mediante un lenguaje de interrogación adecuado. Es necesario tener en cuenta los elementos clave que permiten hacer la búsqueda, determinando un mayor grado de pertinencia y precisión, como son: los índices, palabras clave y tesauros [13].

1.3.1 Índices.

Listado de términos normalizados que representan el contenido de un recurso. Algunos tipos son:

Índice de materias: términos ordenados según las materias que trata la base de datos y el buscador.

- **Índice alfabético:** listado de términos alfabéticamente.
- **Índice KWIC:** tipo de índice permutado en el que el contenido temático de una obra se representa mediante palabras clave de su título o de otra fuente de información del documento.
- **Índice KWOC:** tipo de índice permutado que varía en su presentación respecto al índice KWIC, en que las palabras clave aparecen como un encabezamiento en línea separada. Bajo cada encabezamiento aparece la totalidad de los títulos, completos o truncados, que contienen la palabra clave de que se trata [13].

1.3.2 Palabras clave (*Keywords*).

Término significativo en lenguaje natural que representa el contenido del documento.

En la búsqueda de información esta opción es esencial ya que permite acotar y precisar información. El problema recae en definir la palabra exacta que representa el contenido, por ello es conveniente utilizar especificadores. Por ejemplo si se utiliza la palabra flor en cualquier buscador se puede estar buscando, la floristería más cercana, una imagen de flores o un estudio sobre las flores en las distintas estaciones del año [13].

1.3.3 Tesauros.

Es un listado terminológico controlado sobre un área o ámbito de conocimiento que mantiene entre sí relaciones semánticas y genéricas. Su principal característica es que los términos están ordenados jerárquicamente, permitiendo la precisión terminológica en la búsqueda de información [13].

Componentes:

- **Descriptores admitidos o preferentes:** son aquellos términos normalizado que el tesauro los considera aptos para asignarlos a un documento y que posteriormente facilite la recuperación.
- **Descriptores no admitidos:** son aquellos que aún estando normalizados no se consideran adecuado para utilizarlos, suelen ser sinónimos, términos no utilizados en el campo de actuación [13].

Relaciones:

- **Jerárquicas:** indican cuando un término es más específico que otro.
- **Asociativas:** indican que los términos guardan alguna relación.
- **Sinónimos:** indican que dos términos son sinónimos y cuál de ellos se utiliza como admitido [13].

Uno de los problemas que surgen en la búsqueda de información es si lo que se recupera es "mucho o poco" es decir, dependiendo del tipo de búsqueda se pueden recuperar multitud de documentos o simplemente un número muy reducido. A este fenómeno se denomina "Silencio" o "Ruido documental" [13].

Silencio documental: Son aquellos documentos almacenados en la base de datos pero que no han sido recuperados, debido a que la estrategia de búsqueda ha sido demasiado específica o que las palabras clave utilizadas no son las adecuadas para definir la búsqueda.

Ruido documental: Son aquellos documentos recuperados por el sistema pero que no son relevantes. Esto suele ocurrir cuando la estrategia de búsqueda se ha definido demasiado genérica [13].

1.4 Mecanismos de búsqueda y recuperación de la información en los sistemas de gestión archivística.

En el mundo moderno ha ido evolucionando a gran escala la archivística, a través de sistemas que se están automatizando y optimizando cada vez más, uno de los países impulsores de este proceso es España, el cual cuenta con dos de los sistemas más referenciados, portal PARES y Archivo3000. Otras aplicaciones que se destacan por las funcionalidades que brindan para el proceso de gestión archivística son ICA-AtoM, Albalá y ArchiVenHIS. A continuación se presenta un estudio de los mecanismos de búsquedas de dichos sistemas, con el fin de identificar la utilización de técnicas de Web semántica que sirvan de referencia para la elaboración de la propuesta de solución.

1.4.1 Archivo3000.

Es un software de última generación para la gestión integral de archivos. Su eficiencia se basa en el tratamiento integrado de toda la documentación, que es almacenada en un solo contenedor de datos. Archivo3000 es una aplicación que está en constante proceso de mejora y adaptación a los nuevos estándares y productos comerciales que aparecen en el mercado, gran parte de su trabajo se centra en conseguir una aplicación lo más abierta y multiplataforma posible [14]. Los documentos pueden ser buscados por palabra clave o usando un lenguaje de consultas. Mientras busca, el programa calcula

automáticamente todas las formas de palabras para proveer búsquedas completas de texto en ocho idiomas [15].

1.4.2 PARES.

Ofrece un acceso libre y gratuito, no solo al investigador, sino también a cualquier ciudadano interesado en acceder a los documentos con imágenes digitalizadas de los Archivos Españoles. Ofrece una búsqueda sencilla que consiste en un localizador de unidades documentales mediante un texto libre y un filtro de fechas. Consta de una búsqueda avanzada, esta consiste en localizar las unidades mediante filtros por archivos y fondos, por índices y signaturas. Contiene un inventario dinámico de archivos que informa sobre los documentos de cada archivo que son accesibles, indicando los que están digitalizados [16].

1.4.3 ICA-AtoM.

Es un proyecto del Consejo Internacional de Archivos que consiste en el desarrollo de un programa informático basado en los estándares internacionales de descripción archivística (ISAD (G), ISAAR (CPF), ISDIAH). Esta iniciativa, patrocinada por organismos de diversos países, tiene por objetivo proporcionar gratuitamente un software de licencia libre a aquellas instituciones que quieren tener sus fondos accesibles en línea. El sistema está totalmente basado en web, su manejo es sencillo para el usuario, es flexible y personalizable y está traducido a varios idiomas. El usuario puede recuperar información mediante un cuadro de búsqueda simple que está presente en todas las páginas. El sistema busca la palabra clave consultada en el texto completo de los campos de todas las descripciones archivísticas. ICA-AtoM almacena cualquier descripción que contenga el texto que se busca y lo muestra en la pantalla en una lista de resultados, indicando el título de la descripción y las primeras líneas con el alcance y contenido del campo. Posee una búsqueda avanzada que cuenta con operadores (*and, or, not*) para refinar la búsqueda con más de un criterio, así como palabras clave o frases que se utilizan para buscar coincidencias con los campos de las descripciones almacenadas [17].

1.4.4 Albalá.

Albalá es el sistema de Gestión Integrado de Centros Archivísticos, desarrollado por la empresa Baratz, Servicios de Teledocumentación. Herramienta potente, completa, fácil de usar y sumamente flexible. Está diseñado para satisfacer las necesidades de todo tipo de archivos ya sean públicos o privados, administrativos o históricos. Se adecúa a las normas internacionales para la descripción archivística y de autoridades. El sistema dispone de mecanismos de navegación jerárquica, así como búsquedas por texto libre, junto con la presentación de información identificativa y documental [18].

1.4.5 ArchiVenHIS.

Actualmente en la Universidad de Ciencias Informáticas (UCI) se llevan a cabo proyectos para el desarrollo de software con el propósito de informatizar la sociedad cubana y exportar a otros países. En el año 2007 se desarrolló para el Archivo General de la Nación (AGN) de Venezuela el Sistema de Gestión de Archivos Históricos (SGDAH), ArchiVenHIS, el cual se encuentra muy ligado a los flujos de trabajo y a otras particularidades de la institución.

Es una herramienta desarrollada sobre tecnologías libres y estándares abiertos (PHP, CodeIgniter, MySQL, ISAD (G)) que automatizan el proceso de descripción archivística de la institución y la localización de los documentos con base en los metadatos descritos con el propósito de brindar servicios (préstamos, consulta digital, digitalización, fotocopias, transcripciones), ofrece varias vías para la localización de los documentos tomando como base los metadatos incorporados durante la descripción (búsqueda sencilla, búsqueda avanzada y explorar).

Debido a que no se pudo obtener información referente a cómo se implementaron en los sistemas de gestión archivística PARES, Archivo3000 y Albalá los mecanismos de búsqueda y recuperación de la información, se decidió realizar pruebas a partir de criterios de búsquedas introducidos para comprobar si los resultados obtenidos evidencian la implementación de mecanismos semánticos. Se pudo observar que a pesar de tener funcionalidades como búsquedas sencillas y complejas, agilización de los procesos de localización de documentos y acceso a la información, recuperación rápida de documentos, son sistemas de solo consulta, pues ninguno de estos arroja resultados en los que se puedan apreciar la interpretación de significados sino que son coincidencias únicamente sintácticas.

ICA-AtoM es un sistema de acceso libre desarrollado en symfony, se pudo comprobar que las funciones de búsquedas están basadas en la localización de coincidencias sintácticas en palabras o frases de las descripciones almacenadas.

En el caso de ArchiVenHIS como es un producto de desarrollo por el grupo de proyecto deXcriba, se pudo comprobar que este sistema realiza principalmente tres tipos de búsquedas, búsqueda sencilla o general, búsqueda avanzada y búsqueda exploratoria.

En la búsqueda sencilla o general son introducidos criterios que son utilizados para encontrar coincidencias con los campos descritos en el área Identificación de la norma ISAD (G), específicamente con el Título, con el campo Alcance y contenido del área Contenido y Estructura devolviendo los documentos que coinciden sintácticamente con ese criterio de búsqueda introducido.

Mediante la búsqueda exploratoria no hay que especificar ningún criterio, solamente se busca de forma manual en el árbol de los niveles del archivo los niveles que poseen documentos que se visualizan.

La búsqueda avanzada cuenta con un mayor número de criterios que te permiten refinar la búsqueda de los documentos, al igual que en la búsqueda sencilla se busca coincidencias contra el campo Título, Alcance y Contenido, la fecha que se introduce se compara con el campo *Fechas* del área de Identificación.

Por todo lo anteriormente expuesto en cuanto a los diferentes mecanismos de búsqueda y recuperación de la información se puede llegar a la conclusión que el sistema gestor de documentos históricos realiza búsquedas sintácticas, ya que los resultados de las consultas no son más que las coincidencias sintácticas que se encuentran entre los criterios y los campos de la norma ISAD (G) que han sido descritos en los documentos.

Con el estudio de los principales conceptos y definiciones asociados a la gestión documental de archivos, especialmente las técnicas de recuperación de la información, se pudo comprender la necesidad de que los sistemas afines a la gestión archivística, cuenten con mecanismos de búsqueda cada vez más eficientes, que le permitan al investigador el acceso rápido y preciso a la información deseada.

El análisis de las tendencias actuales en cuanto a los mecanismos de búsquedas implementados en los sistemas de gestión archivística, evidenció que no se hace uso de técnicas semánticas para la recuperación de la información. En todos los casos se emplean búsquedas por sintaxis donde los resultados arrojados responden a coincidencias meramente sintácticas con los criterios de búsqueda introducidos.

Aplicación de las tecnologías de la Web Semántica a la gestión archivística.

En el presente capítulo se exponen los conceptos y definiciones fundamentales asociados a las tecnologías de la Web Semántica, arquitectura, lenguajes y estándares para la representación del conocimiento. Se describen los principales aportes de aplicación de la Web Semántica a la gestión archivística y a partir de ello se propone la integración de dichas tecnologías en el desarrollo del sistema de gestión de documentos históricos, deXcriba.

El uso generalizado de Internet ha impulsado el intercambio de información personal, académica y comercial. El crecimiento exponencial de los datos que se maneja en la web ha generado la necesidad de crear mecanismos inteligentes de clasificación y recuperación de la información que permitan aprovechar todo el conocimiento que en ella reside. Bajo esta filosofía ha surgido el proyecto Web Semántica creado por Berners-Lee [19], cuyo objetivo consiste en extender el actual modelo de web, utilizando una serie de vocabularios que permiten enriquecer la descripción de los recursos disponibles en la red y de esta forma hacerlos semánticamente accesibles. De este modo, la web se convierte en una plataforma universal para el intercambio de información.

El concepto de Web Semántica fue creado por Tim a Berners-Lee creador de la *World Wide Web (WWW)*, quien la define como: “El primer paso es colocar los datos en la web de un modo en que las máquinas puedan entenderlos naturalmente o convertirlos a esa forma. Esto crea lo que se le llama una Web Semántica: una red a de datos que pueden ser procesados directamente o indirectamente por máquinas [20].”

La Web Semántica es una extensión de la actual web, en la cual la información se da mediante un significado bien definido, lo que facilita que los ordenadores y las personas trabajen en cooperación. Aspira a mejorar sustancialmente la interacción entre los sistemas informáticos y los seres humanos, dotando a los primeros de una mayor inteligencia y autonomía y a los segundos de un nuevo universo conceptual que satisfaga la necesidad de convertir la información en conocimiento, es añadir metadatos semánticos a los contenidos que existen en la web, estos metadatos deben ser entendibles tanto por los humanos como por las máquinas [20].

2.1 Tecnologías de la Web Semántica.

La infraestructura de tecnologías y lenguajes necesarios para la implementación de la Web Semántica se esquematiza en una torre de siete capas según Tim Berners-Lee, como se muestra en la Figura 1 a continuación [21]:

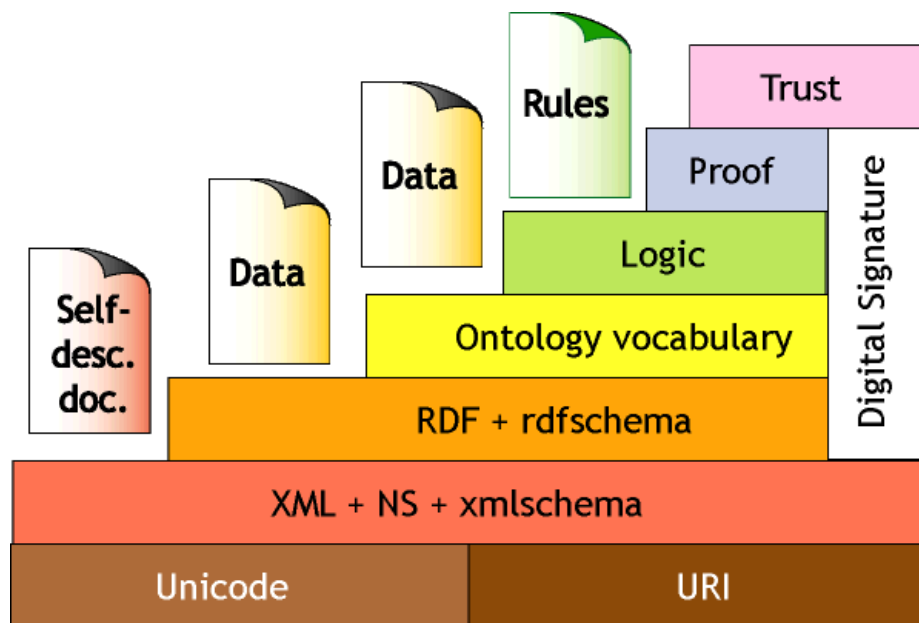


Figura 1 TORRE SEMÁNTICA [22].

La torre de siete capas está estructurada por niveles que establecen una jerarquía entre los mismos:

- **Nivel de recursos:** Identificación de recursos Uniformes (URIs), su importancia está en que se puede señalar a cualquier documento o recurso de cualquier tipo en el universo de la información.
- **Nivel sintáctico:** Definición de lenguajes de etiquetado para añadir semántica a las páginas web, mediante las tecnologías XML y XMLSchema.
- **Nivel de Descripción de Recursos:** Descripción de recursos de la web, utilizando para ello el estándar RDF³.
- **Nivel de Ontologías:** El conocimiento es consensuado y reutilizable a través del uso de ontologías.

³ **RDF.** Es un lenguaje para representar recursos web.

- **Resto de niveles:** La lógica ayuda a inferir conocimiento basado en ontologías y la seguridad da un cierto grado de confianza con el uso de firmas digitales [21].

A continuación se describe el rol de cada una de las capas de la arquitectura de la Web Semántica:

- **Capa Unicode:** Es un estándar que permite la codificación de texto, que garantiza que la información no contenga símbolos extraños.
 - **Capa URI:** Son cadenas de caracteres que identifican inequívocamente un recurso físico abstracto.
 - **Capa XML + NS + Xmlschema:** Es la capa más técnica de la Web Semántica ya que posibilita la comunicación entre agentes.
 - **Capa RDF + RDFSCHEMA:** Define un lenguaje universal con el que se puede expresar diferentes ideas en la WS.
 - **Capa Ontología:** Sirve para clasificar la información agregando clases y propiedades a los recursos.
 - **Capa Lógica:** Se especifican reglas de inferencias para las ontologías.
 - **Capa de Pruebas:** Se intercambiarán “pruebas” escritas en el lenguaje unificador de la Web Semántica.
 - **Capa de Confianza:** Especifica que antes de que no se compruebe la veracidad de las fuentes de información los agentes no ejecuten alguna acción.
- **Capa de Firma digital:** Las firmas digitales se utilizan para que los computadores y los agentes estén seguros que la información proviene de una fuente confiable [21].

La Web Semántica se compone, principalmente, por metalenguajes y estándares para la representación del conocimiento, así como: XML, XML Schema, RDF, RDF Schema y OWL.

La especificación OWL describe la función y relación de cada uno de los componentes de la Web Semántica, tales como:

- **XML:** aporta la sintaxis superficial para los documentos estructurados, pero sin dotarles de ninguna restricción sobre el significado. Representa un lenguaje de marcas diseñado para describir los datos.
- **XML Schema:** es un lenguaje para definir la estructura de los documentos codificados en XML.
- **RDF:** es un modelo de datos para los recursos y las relaciones que se puedan establecer entre ellos.

Aporta una semántica básica para este modelo de datos que puede representarse mediante XML.

- **RDF Schema:** es un vocabulario para describir las propiedades y las clases de los recursos RDF, con una semántica para establecer jerarquías de generalización entre dichas propiedades y clases.
- **OWL:** añade más vocabulario para describir propiedades y clases, tales como relaciones entre clases, cardinalidad, igualdad, tipologías de propiedades más complejas, caracterización de propiedades o clases enumeradas.[22]

Estos lenguajes universales se combinan para aportar descripciones explícitas de los recursos de la web, ya sean catálogos, formularios, mapas u otro tipo de objeto documental. De esta forma el contenido queda desvelado, como los datos de una base de datos accesibles por web o las etiquetas inmersas en el documento, normalmente en XHTML o directamente en XML y las instrucciones de visualización definidas en una hoja de estilos aparte. Estas etiquetas hacen posible, a los gestores de contenidos, interpretar los documentos y realizar procesos inteligentes de captura y tratamiento de información. Además de los metalenguajes y estándares para la representación del conocimiento existen otros conceptos que juegan un papel fundamental dentro de la Web Semántica, tales como:

Ontologías:

Según diferentes autores Neches [23], Gruber [24] y Guarino [25] las ontologías pueden ser entendidas como la suma de una serie de conceptos relevantes que representan el conocimiento compartido por los miembros de un dominio concreto, las relaciones que establecen entre sí estos conceptos y los axiomas definidos sobre estos conceptos y relaciones. Según Clyde [26], las ontologías suponen en sí un canal de comunicación entre personas y máquinas ya que permiten establecer un puente entre el lenguaje natural y la manera en la que se comunican entre sí las máquinas.

Posteriormente Standard Upper Ontology Working Group de IEEE define que una ontología está formada por «un conjunto de conceptos, axiomas y relaciones entre términos que describen un dominio de interés» [27].

Para representar conocimiento de dominio, las ontologías poseen los siguientes componentes:

- ✓ **Conceptos:** ideas básicas que se intentan formalizar. Pueden ser clases de objetos, métodos, planes, estrategias y procesos de razonamiento.
- ✓ **Relaciones:** representan la interacción y enlace entre los conceptos de dominio. Suelen formar la taxonomía de dominio: “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 que pertenecen a un concepto determinado.
- ✓ **Axiomas:** proposiciones que se declaran sobre las relaciones que deben cumplir los elementos de la ontología. Por ejemplo: “Si A y B son de clase C, entonces A no es subclase de B”, “Para todo A que cumpla condición C1, A es B” [28].

La Web Semántica ha atraído a investigadores, laboratorios, empresas e instituciones de los cinco continentes y sigue ganando en popularidad. Existe un gran interés desde el entorno corporativo, el sector público y el mundo académico por hacer de la Web Semántica una realidad, ya que se piensa que puede ser una pieza importante para el progreso de la sociedad de la información. Para ello se está invirtiendo un gran esfuerzo en desarrollar:

- La infraestructura necesaria para su despliegue.
- Aplicaciones que demuestren la viabilidad y el beneficio de la Web Semántica y a la vez motiven el desarrollo y consumo de infraestructura.
- Nuevas soluciones para resolver problemas específicos e ideas que mejoren, amplíen y/o exploten las posibilidades de la Web Semántica [28].

2.2 Aplicación de las tecnologías de la Web Semántica a la gestión archivística.

A continuación se describen las principales propuestas de utilización de tecnologías de la Web Semántica en la gestión archivística. De manera general son pocas las propuestas si se mira desde la óptica más ambiciosa, las principales investigaciones van encaminadas al diseño de modelos ontológicos asociados a este dominio de aplicación. Diferentes autores han publicado numerosos artículos, uno de los que ha marcado pauta dentro de este dominio por sus diferentes publicaciones es el doctor Juan Pablo Palacios Escalona, contribuyendo en el desarrollo de varias estructuras conceptuales para el dominio de los Archivos Digitales. De sus publicaciones más exitosas se pueden citar:

- *Integración de Archivos Digitales en la Web a partir del Sistema de Gestión Parlamentario 'SIAP'.*
- *Semantic Web Digital Archives Integrated Architecture.*
- *E-government: A Legislative Ontology for the 'SIAP' Parliamentary Management System.*

- *E-government: Iniciativas Legislativas en la Web Semántica.*
- *Web Digital Archives Integrated Architecture.*

Las publicaciones anteriores, son extraídas del historial investigador del autor, guardan una estrecha relación con el proceso evolutivo de su tesis doctoral titulada: Modelo de unificación Semántica de ontologías, aplicado al dominio de los Archivos Digitales, esta contribuye con el desarrollo de nuevos modelos ontológicos específicos a este dominio, partiendo de estándares importantes y fundamentales para la descripción del contenido de cualquier archivo. Esta propuesta desarrolla un modelo ontológico que plasma, en un lenguaje formal, los conceptos definidos en las normas ISAD (G) e ISAAR (CPF) y define un modelo conceptual para la unificación semántica de estas ontologías específicas. A continuación se muestra como ha sido descrito el procedimiento para la creación de la ontología para la norma ISAD (G) y la ISAAR (CPF):

2.2.1 Ontología para la Norma ISAD (G).

La ontología propuesta para la norma ISAD (G) recibe el nombre de **ISAD-O**. El lenguaje utilizado para su codificación ha sido OWL.

El concepto '*UNIT_OF_DESCRIPTION*' es el término central de **ISAD-O** y conforme a la especificación de ISAD (G), se puede describir por medio de 26 propiedades que relacionan al término central con otros términos del dominio en cuestión (*Ver Anexo 1*). Seis de ellos son requeridos u obligatorios para la descripción de cualquier Unidad de Descripción (entendiéndose por Unidad de Descripción todo documento o conjunto de documentos, que tratados como un todo, constituyen la base única de una descripción) se trata de los siguientes:

- ✓ *NAME_OF_CREATOR'*
- ✓ *'TITLE'*
- ✓ *'REFERENCE_CODE'*
- ✓ *'DATE'*
- ✓ *'EXTENT_MEDIUM'*
- ✓ *'LEVEL_OF_DESCRIPTION'*

A continuación se muestran en la siguiente tabla los campos de la norma ISAD (G), las propiedades que han sido modeladas en dicho campo y los nuevos términos que son necesarios incluir en la relación.

Campo de la norma ISAD (G)	Propiedades	Términos que relaciona
Código de referencia	isReferencedByCountry isReferencedByIdentifier isReferencedByRepository	Country Identifier Repository
Title	isTitled	Elementos de tipo texto(String)
Date	isDated	creationDate accumulationDate
Level of description	ISAD_G-Fond ISAD_G-Sub-Fond ISAD_G-Series ISAD_G-Sub-Series ISAD_G-File ISAD_G-Item	
Extent and medium of the unit of description	hasPhysicalExtent	PhysicalExtent
Administrative/ Biographical history	createdBy accumulatedBy mantainedBy	ISAAR_Authority
Immediate source of acquisition or transfer	hasTransferRecord hasImmediateSourceOfAcquisitionOrTransfer	TransferRecord
Scope and content	hasContentInformation hasScopeInformation	Content Scope

Accruals	hasInformationOnAction	Action
System of arrangement	internallyStructured isOrdered classifiedAccordingTo	
Conditions governing access	accessAffectedBy	Restriction
Conditions governing reproduction	reproductionAffectedBy	Restriction
Language/scripts of material	symbolSystemOfMaterial scriptOfMaterial languageOfMaterial	SymbolSystem Script Language
Physical characteristics and technical requirements	hasPhysicalCharacteristic	PhysicalCharacteristic
Finding aids	findingAid	
Existence and location of originals	originalISAD_G-UnitOfDescription	originalArchivalMaterial
Existence and location of copies	copyISAD_G-UnitOfDescription	copyArchivalMaterial
Related units of description	relatedISAD_G-UnitOfDescription	
Publication note	isReferencedIn	Publication
Note	hasnote	

Además del término central descrito en la ontología, se modelan un conjunto de conceptos definidos mediante:

- ✓ valores literales: *string*, *character*, *integer*. Se encuentran aquellos conceptos que también se definen como valores literales pero están descritos en algún esquema particular XML o RDF-S. Este tipo de conceptos ofrecen mayor riqueza semántica, ya que se pueden utilizar sus esquemas en parseadores o razonadores para extraer nuevos datos o para comprobar su consistencia. Tal es el caso de 'TITLE' y 'REFERENCE_CODE'.
- ✓ conceptos cuyo rango de valores no se define exclusivamente mediante literales sino que también puede expresarse mediante instancias a otros conceptos, como es el concepto 'DATE'.

2.2.2 Ontología para ISAAR (CPF).

La ontología propuesta para la norma ISAAR (CPF) recibe el nombre de **ISAAR-O**, se ha codificado en OWL.

El concepto 'AUTHORITY_RECORD' representa el término central de esta ontología. Cada Registro de Autoría se puede describir mediante nueve propiedades generales (Ver Anexo 2):

- ✓ 'IDENTITY_CODE'
- ✓ 'NAMES'
- ✓ 'RELATIONSHIPS'
- ✓ 'HAS_NOTE'
- ✓ 'ENTRY'
- ✓ 'RELATED_AUTHORITY_ENTRY'
- ✓ 'HAS_DATE_AND_PLACE'
- ✓ 'NON_PREFERRED_TERM'
- ✓ 'OTHER_INFORMATION'

En **ISAAR-O** se han definido los conceptos 'CORPORATE_BODIES', 'PERSONS' y 'FAMILIES' como conceptos especializados del concepto genérico 'AUTHORITY_RECORD', heredando así, todas las propiedades y atributos de esta súper-clase.

Cada concepto especializado posee a su vez, propiedades que no son compartidas por los demás conceptos ontológicos. Por ejemplo, los conceptos 'PERSONS' y 'FAMILIES' comparten una propiedad en común que es la nacionalidad, pero esta propiedad no tiene sentido dentro del concepto 'CORPORATE_BODIES' ya que una corporación (según ISAAR(CPF)) no tiene nacionalidad. En cambio, el concepto 'CORPORATE_BODIES' posee una propiedad llamada 'legal_number', que no es requerida en las otras entidades. Una de las propiedades necesarias para describir un 'AUTHORITY_RECORD' es 'date'; pero, en este caso, esta propiedad no sólo representa una 'fecha simple', tal como 'día/mes/año' sino que es definida como un concepto particular que permite enlazar una 'fecha simple' con un lugar de ubicación. Es importante mencionar estas pequeñas diferencias conceptuales, ya que posteriormente afectarán al proceso de unificación.

Entre los resultados de la tesis doctoral de Juan Pablo Palacios Escalona se encuentra una propuesta para la unificación de ontología mediante la definición de una metodología.

La metodología de unificación semántica de ontologías propone dos fases para llevar a cabo el modelo de unificación:

- ✓ una primera fase encargada del diseño y configuración del modelo de unificación.
- ✓ una segunda fase encargada del proceso de unificación como tal, es decir, de la creación de una Ontología Global a partir de ciertas ontologías específicas.

La fase de diseño es quizás, la etapa más importante del proceso de unificación, ya que en esta fase se define la información que será la base conceptual del proceso de unificación. Se llama fase de diseño porque en realidad se están diseñando o modelando los roles de cada uno de los componente participantes dentro del modelo de unificación.

La fase de unificación está integrada por una serie de procesos que permiten crear una nueva Ontología Global a partir de los conceptos descritos en las diversas ontologías del nivel de ontologías específicas del dominio (SDO).

Otra de las propuestas encontradas fue la investigación titulada: Sobre la interoperabilidad Semántica en las descripciones Archivísticas digitales creado por Salvador Sánchez Alonso, Miguel Ángel Sicilia Urbán y Gonzalo de Rato Leguina, este artículo presenta un trabajo orientado a facilitar la interoperabilidad Semántica (concepto que cobra su pleno sentido en el contexto de la Web Semántica) de las descripciones Archivísticas. Este propósito se pretende conseguir mediante la representación de ISAD (G) en un lenguaje de ontologías específicamente orientado a la web y estableciendo correspondencias con ontologías de conocimiento general.

La parte de la investigación que realmente aporta novedad con respecto a otros trabajos similares es la correspondencia de ISAD (G) con ontologías de conocimiento general, pues describe cómo enlazar la ontología de dominio particular (en este caso del dominio de la archivística) con ontologías de conocimiento general con el objetivo de obtener interoperabilidad semántica.

Para ello se utiliza la base de conocimiento del proyecto *OpenCyc*, la cual contiene más de 300.000 conceptos que conforman, según sus autores, «una ontología fundamental de toda la realidad consensual humana», junto con millones de aserciones sobre esos conceptos. [29]

La técnica de correspondencia con *OpenCyc* (válida para cualquier otra ontología fundamental) parte de una descripción estructurada como ISAD (G) o de un modelo ontológico como el descrito anteriormente y lleva a cabo un proceso de revisión sistemática de cada término y sus descripciones, produciendo una correspondencia con la ontología fundamental. Los pasos pueden resumirse en lo

siguiente:

- a) Encontrar en la ontología fundamental uno o más términos que subsuman a la categoría que se está considerando.
- b) Comprobar que la correspondencia es consistente, incluyendo al resto de los supra-términos y predicados de éstos en la ontología fundamental.
- c) Proporcionar los predicados adicionales necesarios, en su caso, para caracterizar al nuevo concepto.
- d) Editar en Protégé la definición formal.
- e) Iterar las fases anteriores, con la comprobación de diferentes expertos o grupos de ellos [29].

Es importante reseñar que la ontología descrita para la norma ISAD (G) está escrita en OWL mientras que la ontología *Opencyc* lo está en un lenguaje propio denominado CycL, no compatible con OWL, por lo que sería necesario definir esta última en CycL o bien exportar el conocimiento de *Opencyc* a OWL.

Después del análisis de las propuestas encontradas se propone utilizar en el desarrollo del sistema de gestión de documentos históricos la descripción del modelo ontológico específico para la norma ISAD (G) que se propone en la tesis doctoral de Juan Pablo Palacios Escalona, puesto que modela todas las áreas de información que conforman el registro o descripción archivística de los documentos de archivo, proporcionando un vocabulario y organización de conceptos que representan un marco de trabajo conceptual. Por el contrario el modelo específico para la norma ISAAR (CPF) descrito en esta misma investigación deja fuera conceptos importantes en la descripción de productores que son necesarios en el dominio de la archivística, dado a que esta ontología fue creada por una versión no actualizada de la norma. Por este motivo no se incluye en la propuesta del modelo ontológico, quedando como línea futura de investigación el diseño de una ontología por la versión enriquecida de ISAAR (CPF) que establece los elementos de descripción organizados en cuatro áreas. La ontología para la norma ISAD (G) puede venir a mitigar la situación problemática de la presente investigación, si se parte del hecho que mientras mejor procesada y organizada esté la información en la web, a un nivel que se establezca su significado será más fácil la selección y adquisición de la misma y contribuirá al inicio de una Web Semántica que interprete, entienda y distribuya en virtud de las diferencias contextuales. Para la unificación de los distintos modelos ontológicos que finalmente se decidan utilizar se propone emplear la metodología descrita en la tesis doctoral de Palacios con el fin de lograr interoperabilidad semántica.

Herramientas para el modelado de ontologías.

En el presente capítulo se describen los paradigmas de representación del conocimiento, así como las herramientas para el diseño de modelos ontológicos y los razonadores más referenciados que son utilizados para medir la consistencia de la ontología propuesta.

Actualmente existen diferentes formas de representación del conocimiento, desde simples listas de términos hasta lógicas muy expresivas en las que cualquier conocimiento arbitrario puede ser expresado. Históricamente, los tres principales paradigmas de representación del conocimiento han sido los sistemas basados en redes (redes semánticas y otros formalismos basados en grafos, los sistemas basados en marcos o *frames* y los sistemas basados en lógica, en particular, las llamadas lógicas descriptivas)[30].

3.1 Redes semánticas.

Las redes semánticas son un conjunto de formalismos que representan gráficamente conceptos y sus relaciones. En todos ellos, los conceptos son representados como nodos y las relaciones como arcos entre ellos. Sowa [31] presenta una revisión sobre los muchos subtipos de redes semánticas existentes.

Las redes semánticas tienen su origen en la década de los 50 y aún mantienen aplicaciones prácticas de importancia. Sin embargo, dejaron de emplearse como sistemas generales de representación del conocimiento a partir de la crítica de Woods [32], que mostró que las redes semánticas no podían utilizarse como un formalismo lógico riguroso, al carecer de una caracterización semántica precisa de los conceptos y de sus relaciones, que dificulta el razonamiento sobre el conocimiento.

3.2 Sistemas basados en marcos.

Cada “marco” (*frame*) representa un concepto y se le añaden categorías (*slots*) que pueden tener especificaciones (*fillers*). Por ejemplo, un marco que representa el concepto “vino” puede tener un *slot* “color”, que puede ser rellenado con los valores “blanco”, “rosado” o “tinto”.

La ventaja más obvia de los marcos respecto a las redes semánticas es que todos los valores no tienen que estar representados explícitamente: se permite dar a los *slots* valores por defecto y hay una noción

explícita de herencia que permite definir nuevos marcos a partir de otros ya existentes. Por ello se consideraron muy rápidamente como una alternativa preferible a las redes semánticas [33].

3.3 Sistemas basados en lógicas descriptivas.

El tercer paradigma de representación del conocimiento son las lógicas descriptivas, que proporcionan fundamentos lógicos rigurosos a los sistemas basados en marcos. La lógica descriptiva, es una familia de formalismos lógicos para la representación y el razonamiento sobre conjuntos de conceptos, sus relaciones y las instancias o clases concretas. Provee teorías y sistemas para expresar información estructurada, permitir su acceso y poder razonar de forma semánticamente precisa. La lógica de descripciones es un subconjunto de la Lógica de Primer Orden [33].

Básicamente un Sistema Basado en Lógica Descriptiva (SBDL) se caracteriza por tener una base de conocimiento (KB) compuesta por dos componentes. El primero, el Tbox, el cual indica la terminología del dominio. El segundo, el Abox, que contiene las aserciones acerca de los individuos particulares del dominio con base en el vocabulario expresado en la terminología. En la Figura 2, se puede observar la arquitectura de un SBDL. [33]

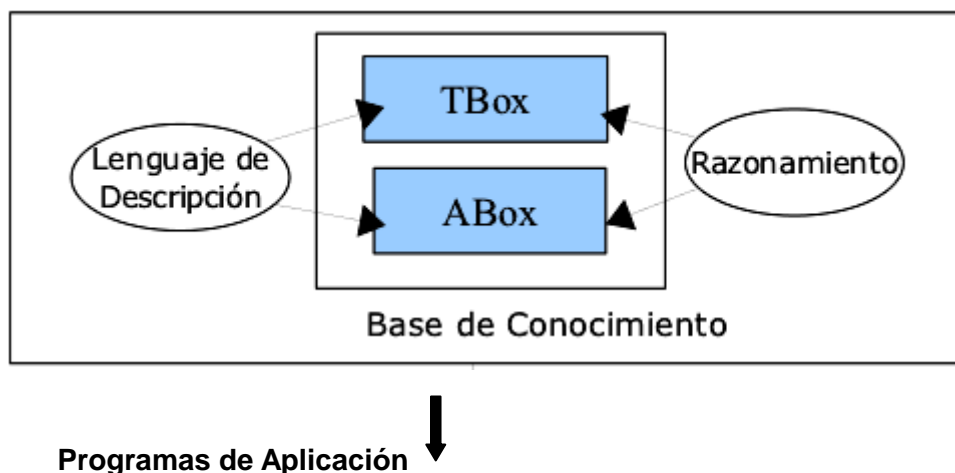


Figura 2 Arquitectura de un SBDL. [33]

3.4 Tareas de razonamiento.

Los razonadores son herramientas para comprobar que una ontología se ha descrito de forma adecuada. Validación, comprobación de conceptos no alcanzables o clasificación son características habituales en ellos. Un detalle importante es que requieren que la ontología responda al perfil OWL-DL para asegurar su correcto funcionamiento. De los existentes los más usados son, Protégé, Pellet,

Racer y FaCT++, que pueden ser integrados a herramientas para el desarrollo de ontologías como Protégé o SWOOP [34].

3.4.1 Pellet: es un razonador *open source* para OWL-DL construido en JAVA, soporta la expresividad de la lógica descriptiva, emplea razonamiento sobre SHOIN (D) y SHIN (D), implementa estrategias de TBox, soporte nominal (clases definidas por extensión), absorción, ramificación semántica. Su razonamiento individual, de tipos de datos (*DataType*) y la optimización de las respuestas a las consultas de Abox, hacen que sea adecuado para aplicaciones Web Semánticas de sonido [35]. Este razonador implementa las mejores técnicas de optimización, lo que hace que su desempeño sea bueno, en especial cuando debe evaluar ontologías con mayor complejidad y expresividad.

3.4.2 FaCT++: soporta la lógica SHOIQ (D) y ha sido desarrollado bajo el proyecto europeo WonderWeb. Está implementado en C++ y corresponde a una nueva versión del razonador DL FaCT, que fue originalmente implementado en Lisp⁴. Para realizar inferencia utiliza los algoritmos *tableaux* e implementa nuevas características y optimizaciones, que permite personalizar para adicionar nuevas tácticas de razonamiento y la capacidad de razonar con lógicas descriptivas más potentes y cercanas a la expresividad de OWL-DL.

Es un buen razonador para la TBox de una ontología. Sin embargo, hay dos desventajas importantes de FaCT++. Por un lado, no tiene soporte para otros tipos de dato que no sean *string* o *integer*, como sí ocurre, por ejemplo, con Pellet. Por otro lado, tampoco tiene soporte para el razonamiento con la A-Box de la ontología. En cualquier caso, ambas carencias no son del todo definitivas, ya que es posible utilizar el lenguaje de consultas para RDF: SPARQL, que permite consultar el modelo inferido de una ontología OWL-DL. En el caso de los tipos de datos en un entorno web, es preferible que exista soporte para los tipos de datos de XML Schema.

Entre las ventajas que ofrece se tienen: posee licencia GPL, que todavía tiene mantenimiento y sigue en desarrollo, trabaja de forma eficiente con la T-Box de ontologías de tamaño grande y mediano [35].

3.4.3 Racer: es un razonador DL para la lógica descriptiva, desarrollado en Lips. Inicialmente fue creado por la Universidad de Hamburgo, pero luego se convirtió en software propietario y su nombre comercial es RacerPro. Soporta OWL DL excepto para los nominales (clases definidas por una enumeración de sus miembros, que implementa como definiciones parciales) y para tipos de datos no estándar; maneja largas ABoxes en combinación con largas y expresivas Tboxes y provee servicios de

⁴ **LISP.** Es un lenguaje de programación creado por el profesor John McCarthy a finales de los años 50, principalmente para el procesamiento de listas.

inferencia sofisticados [35].

3.5 Herramientas para el diseño de ontologías.

Actualmente existen varias herramientas cuya meta es automatizar las tareas de diseño, implementación, integración y anotación de ontologías, entre las más relevantes se pueden mencionar:

3.5.1 Protégé: desarrollado en el Stanford Medical Informatics de la Universidad de Stanford. Es un entorno abierto y extensible para el desarrollo y administración de ontologías. Desde 1987 Protégé ha ido evolucionando hasta convertirse hoy en día en una herramienta multiplataforma, extensible, libre y con soporte. Maneja varios lenguajes ontológicos, entre los que se encuentran: RDF(S), FLogic, Prolog, DAML+OIL y OWL. Además de ser un editor es considerado un entorno de desarrollo que provee no solo herramientas para el diseño sino que incluye razonadores que permiten chequear la consistencia de la ontología [22].

3.5.2 OntoEdit: desarrollado por el AIFB de la Universidad de Karlsruhe. Es un entorno para modelar y administrar ontologías, basado en una arquitectura extensible. Ofrece una herramienta gráfica que permite desarrollar y mantener ontologías codificadas en diversos lenguajes, entre los que destacan: F-Logic, RDF-Schema y DAML-OIL [22].

3.5.3 SWOPP: es un editor de ontologías similar a Protégé, Una de las principales características de SWOPP es la posibilidad de particionar automáticamente las ontologías y trabajar de forma independiente con cada módulo desde el editor, se destaca por su capacidad para resolver consultas SPARQL y por justificar las inferencias realizadas por su razonador (Pellet). Es muy útil en la depuración de ontologías [30].

Luego del estudio realizado de las herramientas utilizadas para la edición de ontologías, se propone para el diseño del modelo ontológico el editor de ontología Protégé. Este más que un editor es considerado todo un entorno de desarrollo que provee no solo herramientas de diseño, sino que además incluye razonadores que permiten chequear la consistencia de las ontologías modeladas, entre los que destaca Pellet y FACT. Soporta simultáneamente los tres paradigmas de representación del conocimiento. Es un sistema abierto, gratuito y posee una interfaz fácil de usar e intuitiva. Esta herramienta permite interactuar con diferentes bases de datos relacionales e implementar algoritmos de integración y visualización de ontologías, maneja varios lenguajes ontológicos, entre los que se encuentran: RDF(S), FLogic, Prolog, DAML+OIL y OWL.

3.6 Validación de la propuesta.

La propuesta del modelo ontológico para la norma ISAD (G) ha sido avalada por la tesis doctoral

titulada Modelo de unificación Semántica de ontologías, aplicado al dominio de los Archivos Digitales, del doctor Juan Pablo Palacios Escalona, quien se ha especializado en el manejo del conocimiento mediante modelos ontológicos, contribuyendo en el desarrollo de varias estructuras conceptuales para el dominio de los Archivos Digitales. Este investigador ha publicado un grupo de trabajos que guardan estrecha relación con el proceso evolutivo de su tesis doctoral, dentro de las que se pueden citar:

- *Semantic Web Digital Archives Integrated Architecture.*
- *E-government: A Legislative Ontology for the 'SIAP' Parliamentary Management System.*
- *E-government: Iniciativas Legislativas en la Web Semántica.*
- *Integración de Archivos Digitales en la Web a partir del Sistema de Gestión Parlamentario 'SIAP'.*
- *Web Digital Archives Integrated Architecture.*

Cada una de las aportaciones expuestas en la tesis doctoral ha sido publicada, paulatinamente, en diferentes congresos y conferencias internacionales lo que avala, en cierto modo, la viabilidad conceptual de las ontologías y el modelo de unificación propuesto. La investigación realizada en esta tesis doctoral ha sido parcialmente financiada por el proyecto DAWIS-UPM, del Ministerio de Ciencia y Tecnología en España.

Por otro lado la propuesta de la utilización de la herramienta Protégé puede ser avalada por su uso en la mayoría de las investigaciones científicas donde se modelan ontologías de dominio general y específico. A continuación se referencian algunas de estas publicaciones:

- Roldán García, María del Mar, Molina-Castro, Joaquín, Aldana Montes, José Francisco. A Protégé Plugin for Querying and Reasoning on Persistent OWL Ontologies. SWAP, 2008.
- Torres Garrigós, David. Evolución de DATEX II a un modelo semántico del autor, Universidad de Valencia, 2009.
- Jiménez Ruiz, Ernesto. Ontologías en Informática del autor Ismael Sanz, 2007.
- Sánchez Alonso, Salvador, Sicilia Urbán, Miguel Ángel, de Rato Leguina, Gonzalo. Sobre la interoperabilidad semántica en las descripciones archivísticas digitales, España, 2008.
- Codina, Lluís, Rovira, Cristófol. Recursos sobre la Web Semántica, España, 2006.

- Morales del Castillo, José Manuel, Hurtado Martín, Germán. Integrando las Tecnologías de Web Semántica en la Archivística, Universidad Carlos III de Madrid y la Universidad de Gante (Bélgica) ,2011.
- Aguirre Helguero, Norka Natalia. Un agente basado en un razonador de ontologías, Buenos Aires, 2011.

En la presente investigación se realizó un estudio de los mecanismos de búsquedas implementados en los sistemas de gestión archivística, en el cual se pudo evidenciar que ninguno de estos emplea técnicas semánticas en los procesos de recuperación de la información, debido a que las búsquedas que realizan arrojan resultados meramente sintácticos.

A partir del estudio que se realizó de las principales propuestas de aplicación de las tecnologías de la Web Semántica a la gestión archivística, se propone un modelo ontológico para la norma ISAD (G) en el que se modelan todas las áreas de información que conforman el registro o descripción archivística de los documentos de archivo. Además, se propone también la utilización de la metodología descrita en la tesis doctoral de Juan Pablo Palacios Escalona dirigido a la unificación de las ontologías con el objetivo de lograr mayor interoperabilidad semántica.

Se realizó un estudio de las herramientas utilizadas para la edición de ontologías quedando seleccionada para el diseño del modelo ontológico propuesto el editor Protégé ya que es un sistema abierto, gratuito y posee una interfaz fácil de usar e intuitiva.

1. Definir un modelo ontológico para la versión enriquecida de ISAAR (CPF) que establece los elementos de descripción organizados en cuatro áreas.
2. Implementar el modelo ontológico en el sistema de gestión de documentos históricos.
3. Realizar un estudio de las herramientas que son empleadas en la integración de ontologías.

- [1] Los Archivos Históricos como fuente de conocimiento y cultura para los estudiantes universitarios: Archivo Histórico. In: [En línea]. [Citado el 4 Octubre de 2011]. [Disponible en: <http://archivohistorico.villaclara.cu/estudios/los-archivos-historicos-y-las-universidades>].
- [2] Recuperación de Información: Historia. In: [En línea]. [Citado el 4 Octubre de 2011]. [Disponible en: <http://recuperacion7.blogspot.com/2010/10/historia.html>].
- [3] Tipos de Buscadores. In: [En línea]. [Citado el 4 Octubre de 2011]. [Disponible en: http://www.hipertexto.info/documentos/tipos_buscador.htm#motores].
- [4] Tipos de Buscadores. In: En línea]. [Citado el 5 Octubre de 2011]. [Disponible en: http://www.hipertexto.info/documentos/tipos_buscador.htm#motores].
- [5] HEREDIA, Antonia. Archivística general. Teoría y práctica. 5ta Edición. Sevilla, 1991.
- [6] MENA MUGICA, Mayra. Gestión documental y organización de archivos. La Habana: Félix Varela. ISBN 959-258-950-X, 2005.
- [7] Normas y Estándares En Comunicación - Ensayos - Zigma37. In: [En línea]. [Citado el 15 de Octubre de 2011]. [Disponible en: <http://www.buenastareas.com/ensayos/Normas-y-Estandares-En-Comunicacion/820701.html>].
- [8] Comité de Normas de Descripción del CIA, Norma Internacional General de Descripción Archivística-ISAD (G), Madrid: Ministerio de Educación Cultura y Deporte-Subdirección de los Archivos Estatales, 2000.
- [9] Comité de Normas de Descripción del CIA, Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias, 2004.
- [10] Comité de Buenas Prácticas y Normas (ICA/CBPS), Norma internacional para describir instituciones que custodian fondos de archivo, Londres, Reino Unido, 2008.
- [11] DELGADO GÓMEZ, Alejandro, Introducción a Encoded Archival Description (EAD): Mitos y oportunidades, Ayuntamiento de Cartagena-Servicio de Archivos y Bibliotecas Centro Cultural Ramón Alonso Luzzy, 2005.
- [12] ENCONDING ARCHIVAL CONTEXTS – EAC. Encoding Archival Context: EAC. .”
- [13] BÚSQUEDA Y RECUPERACIÓN DE INFORMACIÓN. [En línea]. [Citado el 16 de Octubre de 2011]. [Disponible en: http://www.mariapinto.es/e-coms/recu_infor.htm].

- [14] 3000 Informática. SL. Archivo 3000 W. [Citado el 16 Octubre de 2011] [Disponible en: <http://www.archivo3000.com/general/a3wp.htm>].
- [15] Archivarius 3000. [Citado el 16 Octubre de 2011]. [Disponible de: <http://www.likasoft.com/es/document-search/features.shtml>].
- [16] Portal de Archivos Españoles. [En línea]. [Citado el 17 Octubre de 2011] [Disponible en: <http://pares.mcu.es/>].
- [17] ARAÑA CRUZ, Jonathan, HERRERA TEJADA, Clara, 2011. ICA-AtoM, una buena herramienta para la difusión de los archivos en la web. March 2011. S.l.: s.n.
- [18] MANZANARES, Daniel, MOCHÓN, Gonzalo, MARTÍN OYA, María, FERNÁNDEZ, Paula, 2003. Directorio Español de Software para la Gestión Bibliotecaria, Documental y de Contenidos. 2003. S.l.: s.n.
- [19] Berners-Lee: Weaving the Web. [En línea]. [Citado el 18 Octubre de 2011] [Disponible en: <http://www.w3.org/People/Berners-Lee/Weaving/Overview.html>].
- [20] Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. *The Scientific American*, 284 (5), 34-43.
- [21] SARANGO ROMERO, Darwin Leonardo, 2011. Publicación de datos universitarios observando los principios de Linked Data. Ecuador: Universidad Técnica Particular de Loja.
- [22] PALACIOS ESCALONA, Juan Pablo. Modelo de unificación Semántica de ontologías, aplicado al dominio de los Archivos Digitales. S.l.: Departamento de Ingeniería de Sistemas Telemáticos, 2005.
- [23] RIVER, Neches. Enabling technology for knowledge sharing. *AI Magazine*, 12 (3), 33-56, 1991.
- [24] GRUBER, Tom. Towards principles for the design of ontologies used for knowledge sharing, *Int. Workshop on Formal Ontology*, 1993.
- [25] GUARINO, Nicola, Formal Ontology, Conceptual Analysis and Knowledge Representation, *Int. Journal of Human and Computer Studies*, 43(5-6), 1995.
- [26] CLYDE, William. & Hissaple, J.K.D. (2002). A collaborative approach to ontology design. *Communications of the ACM*, 45 (2), 42-47.
- [27] Standard Upper Ontology Working Group de IEEE (SUO WG, <http://suo.ieee.org/>).
- [28] AGUIRRE HELGUERO, Norka Natalia. Un agente basado en un razonador de ontologías. S.l.:

Universidad de Buenos Aires, 2011.

[29] SÁNCHEZ ALONSO, Salvador, SICILIA URBÁN, Miguel Ángel, DE RATO LEGUINA, Gonzalo, 2008. Sobre la interoperabilidad semántica en las descripciones archivísticas digitales, 2008.

[30] SANZ, Ismael, JIMÉNEZ RUIZ, Ernesto. Ontologías en Informática, 2007.

[31] SOWA, John. "Principles of Semantic Networks: Explorations in the Representation of Knowledge", Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[32] WOODS, William. What's in a Link: Foundations for Semantic Networks, en BOBROW, D.G. y COLLINS, Representation and Understanding, Academic Press, 1975.

[33] FLORES VITELLI, Ivan. Introducción al Razonamiento Sobre Ontologías, Universidad Central de Venezuela Facultad de Ciencias Escuela de Computación, 2011.

[34] TORRES GARRIGÓS, David. Evolución de D_AT_EX II a un modelo semántico. Universidad de Valencia, 2009.

[35] Razonadores semánticos: un estado del arte. In: Revista de la Facultad de Ingeniería, 2010.

- HEREDIA, Antonia. Archivística general. Teoría y práctica. 5ta Edición. Sevilla, 1991.
- MENA MUGICA, Mayra. Gestión documental y organización de archivos. La Habana: Félix Varela. ISBN 959-258-950-X, 2005.
- DELGADO GÓMEZ, Alejandro, Introducción a Encoded Archival Description (EAD): Mitos y oportunidades, Ayuntamiento de Cartagena-Servicio de Archivos y Bibliotecas Centro Cultural Ramón Alonso Luzzy, 2005.
- ARAÑA CRUZ, Jonhatan, HERRERA TEJADA, Clara, 2011. ICA-AtoM, una buena herramienta para la difusión de los archivos en la web, 2011.
- MANZANARES, Daniel, MOCHÓN, Gonzalo, MARTÍN OYA, María, FERNÁNDEZ, Paula, 2003. Directorio Español de Software para la Gestión Bibliotecaria, Documental y de Contenidos, 2003.
- SARANGO ROMERO, Darwin Leonardo. Publicación de datos universitarios observando los principios de Linked Data. Ecuador: Universidad Técnica Particular de Loja, 2011.
- PALACIOS ESCALONA, Juan Pablo. Modelo de unificación Semántica de ontologías, aplicado al dominio de los Archivos Digitales. S.I.: Departamento de Ingeniería de Sistemas Telemáticos, 2005.
- RIVER, Neches. Enabling technology for knowledge sharing. AI Magazine, 12 (3), 33-56, 1991.
- GRUBER, Tom. Towards principles for the design of ontologies used for knowledge sharing, Int. Workshop on Formal Ontology, 1993.
- GUARINO, Nicola, Formal Ontology, Conceptual Analysis and Knowledge Representation, Int. Journal of Human and Computer Studies, 43(5-6), 1995.
- CLYDE, William. & Hissaple, J.K.D. A collaborative approach to ontology design. Communications of the ACM, 45 (2), 42-47, 1991.
- AGUIRRE HELGUERO, Norka Natalia. Un agente basado en un razonador de ontologías. S.I.: Universidad de Buenos Aires, 2011.
- SÁNCHEZ ALONSO, Salvador, SICILIA URBÁN, Miguel Ángel, DE RATO LEGUINA, Gonzalo, 2008. Sobre la interoperabilidad semántica en las descripciones archivísticas digitales. In: 2008.
- SANZ, Ismael, JIMÉNEZ RUIZ, Ernesto. Ontologías en Informática, 2007.
- FLORES VITELLI, Iván. Introducción al Razonamiento Sobre Ontologías. S.I. Universidad Central de Venezuela Facultad de Ciencias Escuela de Computación, 2011.

DAVID TORRES GARRIGÓS. Evolución de DATEX II a un modelo semántico. Universidad de Valencia, 2009.

Razonadores semánticos: un estado del arte. In: Revista de la Facultad de Ingeniería, 2010.

MORALES DEL CASTILLO, José Manuel, HURTADO MARTÍN, Germán. Integrando las Tecnologías de Web Semántica en la Archivística. Universidad Carlos III de Madrid, 2011.

Archivo: Es uno o más conjuntos de documentos, sea cual sea su fecha, su forma y soporte material acumulados en un proceso natural por una persona o institución pública o privada en el transcurso de su gestión, conservados, para servir como testimonio en formación para la persona o institución que los produce.

ISAD (G): (Norma Internacional General de Descripción Archivística) es una normativa internacional para la descripción de documentos. Contiene reglas generales para la descripción archivística que pueden aplicarse con independencia del tipo documental o del soporte físico de los documentos.

ISAAR (CPF): (Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias) sirve de guía para elaborar registros de autoridad de Archivos que proporcionan descripciones de entidades (instituciones, personas y familias) asociadas a la producción y a la gestión de archivos.

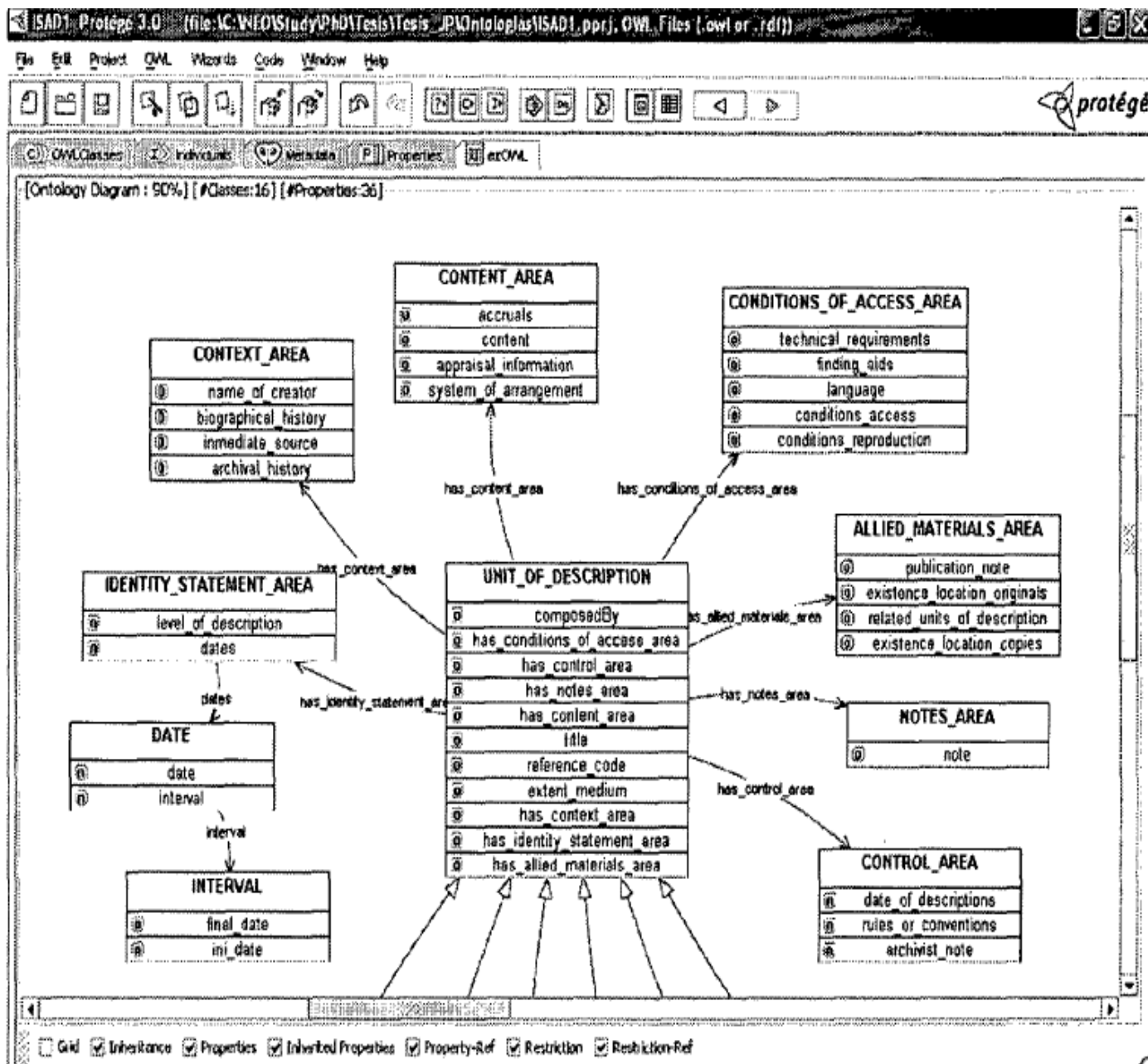
Índice KWIC: Tipo de índice permutado en el que el contenido temático de una obra se representa mediante palabras clave de su título o de otra fuente de información del documento.

Ontología: Una ontología está formada por «un conjunto de conceptos, axiomas y relaciones entre términos que describen un dominio de interés».

World Wide Web: es un sistema de distribución de información basado en hipertexto o hipermedios enlazados y accesibles a través de Internet. Con un navegador web, un usuario visualiza sitios web compuestos de páginas web que pueden contener texto, imágenes, vídeos u otros contenidos multimedia y navega a través de ellas usando hiperenlaces.

Web Semántica: Vendría a ser una extensión de la web actual dotada de significado, de manera que pudiera ser interpretada tanto por agentes humanos como por agentes computarizados.

Anexo 1 Propiedades de la UNIT_OF_DESCRIPTION en ISAD-O



Anexo 2 Gráfico parcial de la ontología de ISAAR (CPF)

