



REPRESENTACIÓN Y ORGANIZACIÓN DE DOCUMENTOS DIGITALES: DETALLES Y PRÁCTICA SOBRE LA ONTOLOGÍA DIMH

REPRESENTATION AND ORGANIZATION OF DIGITAL DOCUMENTS: DETAILS AND PRACTICE ON ONTOLOGY DIMH

Ángel Castellanos González
ETSI Informática, UNED
acastellanos@lsi.uned.es

Ana García Serrano
ETSI Informática, UNED
agarcia@lsi.uned.es

Resumen

Las Humanidades Digitales pretenden facilitar el acceso y entendimiento de documentos históricos mediante aplicaciones informáticas. En este proceso es importante la etapa de representación formal y digital de los contenidos para facilitar el posterior proceso de estos en aplicaciones, por ejemplo, de acceso y visualización, búsqueda y organización automática de contenidos. En este sentido, este trabajo presenta diferentes aproximaciones de representación en el ámbito del proyecto DIMH (accesible desde <https://dimh.hypotheses.org/>). En particular se detalla el desarrollo de una ontología para la representación de los contenidos del corpus DIMH. Para reforzar la comprensión de la estructura ontológica y mostrar la potencia de los modelos ontológicos, se presentan una serie de ejemplos prácticos y de consultas.

Palabras clave: Estructuración de la información. Aplicaciones web. Ontologías. Humanidades Digitales.

Abstract

Digital Humanities intend to facilitate the access and understanding of historical documents by means of the utilization of software applications. In this process, it is important the formal and digital representation of the content in order to facilitate their later process towards applications to the access and visualization, search and automatic organization of the contents. In this regard, this paper presents different representation approaches in the context of the DIMH project (accessible from <https://dimh.hypotheses.org/>). In particular, this paper details the development of an ontology for the representation of the content of the DIMH corpus. To strength the comprehension of the ontological structure as well as to show the power of the ontological models, a series of practical examples and SPARQL queries are introduced.

Keywords: Information Structuration. Web Applications. Ontology. Digital Humanities.

1. INTRODUCCIÓN¹

Tradicionalmente, las aplicaciones software se incorporaron a las Humanidades Digitales orientadas a la preservación digital de los documentos históricos (Drucker *et al.*, 2014). Sin embargo, una vez que se han digitalizado y anotado los documentos por humanistas expertos en cada tema, se hace necesario gestionar estas fuentes de información para su acceso, visualización y explotación por parte de los usuarios (Ruthven y Chowdhury, 2014). Para el desarrollo de aplicaciones enfocadas a la utilización de estos datos (textos digitalizados y anotaciones) es necesario responder a las siguientes preguntas, las cuales dependen fundamentalmente de la tarea u objetivo a resolver:

- ¿qué tipo de datos se han de manejar?
- ¿dónde están esos datos?
- ¿qué tipo de representación y almacenamiento es necesario?
- ¿qué tipo de acceso y consulta se necesita?

¹ Este trabajo ha sido financiado parcialmente por el proyecto DIMH (HAR2012-31117) del Ministerio de Economía y Competitividad, la beca de doctorado FPI-UNED 2014 y el proyecto MUSACCES (S2015/HUM3494) de la Consejería de Educación, Juventud y deporte de la Comunidad de Madrid.

- ¿qué se quiere mostrar y a quién?

De entre los aspectos abordados por estas preguntas, este trabajo se centra en la representación y el modelado de los datos para su uso compartido en la web. Dicha representación se basa en estándares que permiten su acceso y explotación por agentes o aplicaciones software y personas. Estos estándares incluyen, entre otros, los metadatos, con los que se describen uniformemente a los contenidos de dominios concretos, nombres de personas, etc.

En el marco del proyecto *El dibujante ingeniero al servicio de la monarquía hispánica. Siglos XVI-XVIII* (DIMH), tras una etapa de familiarización de la anotación (fichas de texto en español digitalizadas) con aspectos relevantes de los objetos de la *Colección de mapas, planos y dibujos* del Archivo General de Simancas (AGS), se realizó un proceso automático de enriquecimiento lingüístico de las fichas de la colección anterior, se generó el nuevo corpus denominado corpus DIMH y se desarrolló y publicó una aplicación web para nuevas búsquedas de información (Castellanos *et al.*, 2015).

Era entonces el momento de preguntarse: ¿Qué nuevos beneficios puede recibir el investigador humanista de esta tarea de anotación y de su digitalización?, ¿qué más se puede obtener automáticamente para facilitar su investigación? Una posible respuesta a estas cuestiones es el desarrollo de aplicaciones que estructuran automáticamente la información para su posterior aprovechamiento en conceptos que van más allá de un primer nivel de organización, como es el alcanzado con la anotación sobre la base de metadatos.

Tras una breve presentación de tres aproximaciones de estructuración del conocimiento a partir de textos anotados, se ejemplificarán dos de ellas: una con una descripción breve, y, la segunda, con cierto nivel de detalle, por ser una de las contribuciones de este artículo. Esta última se corresponde con el proceso de creación de una ontología de dominio a partir del corpus DIMH, que incorpora el conocimiento previo aportado por los investigadores humanistas del proyecto DIMH.

En este artículo se incluye, en primer lugar, una breve descripción de la colección AGS original y del corpus DIMH. A continuación, se contextualizan e introducen con cierto detalle dos aproximaciones automáticas de estructuración del conocimiento aplicadas al corpus textual DIMH. Se presenta la ontología desarrollada sobre los conceptos del corpus y sus relaciones y a continuación se refuerza la comprensión de la estructura ontológica con una serie de ejemplos prácticos sobre la herramienta de desarrollo utilizada, Protégé, que ya es un estándar tanto en el ámbito académico como en el profesional. Finalmente, se realizan varios ejemplos de consultas SPARQL (un lenguaje estándar de la web semántica para la consulta de datos representados mediante ontologías en formato RDF (*Resource Description Framework*), para mostrar la potencia de los modelos ontológicos.

1.1. La colección de mapas, planos y dibujos del AGS

El contenido principal de la *Colección de mapas, planos y dibujos* del Archivo General de Simancas (AGS)² es información textual en formato de fichas (7792 fichas) relacionadas con mapas, planos y dibujos (en formato de imagen), a las que es posible acceder con búsquedas simples o avanzadas, esto es filtradas según los sus campos descriptivos (metadatos).

La interfaz con facilidades de búsqueda (con o sin filtros)³, permite una navegación clasificada (marco de la izquierda) por lenguas, materias posibles y por pertenecer a 157 clases que creemos generadas a partir de los títulos de las fichas⁴.

Una vista de la información que se ofrece al usuario es la de la Figura 1:



Sección:	Material cartográfico AGS
Número de control:	BAB20101018842
Título uniforme:	Bayona (Francia). Fortificaciones. Planos. 1639 
Título:	[Planta de la fortificación de Bayona] [Material cartográfico] / Galvarreta
Área de datos:	Sin escala
Publicación:	[1639]
Descripción física:	1 plano : ms. ; 40,7 x 55,8 cm
Notas:	Nombre del autor tomado de la documentación adjunta Referencias: Sin publicar Tinta negra. Explicación con clave alfabética Conservación [Buena] Fecha tomada de la documentación adjunta Manuscrito sobre papel AGS. Secretaría de Guerra, 01265. Con consulta del Consejo de Guerra. Madrid, 8 de marzo de 1639
Nombre jerárquico lugar:	Francia-Aquitania-Bayona 
Autores secundarios:	Galvarreta 
Tipo de publicación:	 Mapas
Ejemplares:	Archivo General de Simancas. Signatura: MPD, 67, 083. Ubicación Anterior: GYM, 01265 Préstamo:  Disponible  Copia digital

Figura 1. Ficha mostrada por el buscador AGS.

La información textual de estas fichas originales está codificada con metadatos en el formato RDF-DC⁵. Un ejemplo es la ficha siguiente, donde todos los términos entre los caracteres <y> son del metalenguaje que describe los metadatos:

² Accesible a través de <http://www.mcu.es/ccbae/es/mapas/principal.cmd>.

³ La interfaz con facilidades de búsqueda (con o sin filtros) se encuentra en <http://www.mcu.es/ccbae/es/consulta...id=30485>.

⁴ Dicha clasificación puede encontrarse en <http://www.mcu.es/ccbae/es/consulta>.

⁵ Accesible desde <http://www.mcu.es/ccbae/es/consulta/busqueda.cmd>.

```
<rdf:Description>
<dc:relation> Referencias: Mapas, planos y dibujos (Años 1503-1805). Volumen I: p.
405</dc:relation>
<dc:coverage>-S.XVIII</dc:coverage>
<dc:title> [...] Planos y Perfiles de la obra que se construye antes de empezarse a
fundir para recipientes del Mineral y que se deshace después de evacuada la fundición
que es la que se supone llamarse Crisol [Material gráfico no proyectable]</dc:title>
<dc:description> AGS. Secretaría de Marina, 00679. Acompaña a carta de don
Maximo de la Croix a don Ricardo Wall, Chaves, 16 de julio de 1762</dc:description>
<dc:description> Tinta y colores. Con explicación</dc:description>
<dc:description> Manuscrito sobre papel.</dc:description>
<dc:type> Ilustraciones y Fotos</dc:type>
<dc:language> spa </dc:language>
dc:date> 1762 </dc:date>
<dc:identifier>
http://www.mcu.es/ccbae/es/consulta/registro.cmd?id=178725</dc:identifier>
<dc:format> image/jpeg </dc:format>
<dc:subject> Hornos metalúrgicos-Dibujos </dc:subject>
</rdf:Description>
```

1.2. El corpus DIMH

Cuando se descargan las 7792 fichas de la colección original, lo que se obtiene son 8 ficheros. Por tanto, el primer paso en el proyecto DIMH fue extraer esta información y generar un corpus en el que cada ficha esté contenida en un fichero XML independiente.

A continuación, se realizó un proceso de anotación de la colección que consistió en el enriquecimiento del contenido textual de las fichas con las entidades nombradas (`<nes>`) y por categorías (`<nes_person>`, `<nes_organization>`, `<nes_location>`, `<nes_misc>`), los sintagmas nominales (`<sintagmas>`) y los lemas (`<lemas>`). El resultado es un nuevo fichero en formato XML con los campos relativos a cada tipo de información identificada. Un ejemplo es el siguiente:

```
<Ficha id="178124">
<Idioma>spa</Idioma>
<Publicacion>1787</Publicacion>
<Formato>image/jpeg</Formato>
<Referencias> Referencias: Sin publicar</Referencias>
```

<Tipo>Otros</Tipo>

<Notas> AGS. Secretaría y Superintendencia de Hacienda, 00776. Con carta de Miguel de Vallejo a Pedro de Lerena. Guadalajara, 5 de julio de 1787. </Notas>

<Tematica>Vicuña (Tejidos)-Muestras</Tematica>

<Materia>-S.XVIII</Materia>

<Titulo> [Una muestra de vicuña negra] [Objeto tridimensional]</Titulo>

<nes> Guadalajara Guadalajara Hacienda Miguel_de_Vallejo Pedro_de_Lerena

</nes>

<nes_person>Miguel_de_Vallejo Pedro_de_Lerena</nes_person>

<nes_organization>Hacienda</nes_organization>

<nes_location>Guadalajara Guadalajara</nes_location>

<nes_misc></nes_misc>

<lemas> [una muestra de vicuña negra] [objeto tridimensional] Vicuña (tejido) - muestra AGS. secretaría y superintendencia de Hacienda, 00776 . con carta de Miguel a Pedro. Guadalajara, 5 de julio de 1787. otro referenciar: sin publicar - S.XVIII

</lemas>

<sintagmas> muestra_vicuña_negro_objeto vicuña_negro vicuña_tejido_muestra secretaria_superintendencia_hacienda superintendencia_hacienda hacienda carta_miguel_pedro miguel_pedro pedro guadalajara julio_1787 </sintagmas>

</Ficha>

Todas las anotaciones originales de las fichas están, generalmente, hechas en español (notas, referencias, materia, temática.). El metadato del idioma se refiere al del título de los mapas, planos o dibujos en su soporte original (en español hay 7123).

Tras el estudio del contenido de las fichas, se identificó que los datos almacenados se pueden clasificar al menos en los siguientes tipos:

1. Fecha: número de cuatro cifras que indica el año en el que se realizó la obra.
2. Referencias: indica el libro o revista donde podemos encontrar la obra.
3. Creador – Autor: autor de la obra.
4. Tipo: Se han distinguido únicamente cinco tipos de obras: Ilustraciones y fotos; Mapas; Manuscritos; Libros; Otros
5. Idiomas – Lengua: indica la lengua en la que está redactada la obra original, codificadas por ejemplo: Spa – Español, Fre – Francés, Eng – Inglés, Lat – Latín, Ger – Alemán, Ita – Italiano, Por – Portugués, Dut – Holandés, Cat – Catalán.
6. Temática: los valores son muy diversos y cambiantes en cada obra. Algunos de los valores posibles son los siguientes: Cañones dibujos; Ballenas dibujos; Fauna dibujos;

Artillería dibujos; Máquinas dibujos; Barcos dibujos; Uniformes Militares Dibujos; Hospitales planos; Faros planos; Barcos planos; Fortificaciones planos; Dibujos de arquitectura; Municiones.

7. Técnica utilizada para la realización de la obra. Por ejemplo:
 - a. Tinta. Puede ser aguada, negra.
 - b. Tinta y colores. Los colores pueden ser: negro/a, rojo/a, dorado/a, granate, verde, amarillo/a, gris, verde, sepia.
 - c. Lápiz negro
 - d. Todas ellas pueden ser: Con explicación o Con rotulación
8. Soporte – Impreso en – Escrito en. Por ejemplo: Pergamino; Manuscrito sobre papel; Copia; Grabado (solo ilustraciones y fotos); Impreso.

Sin embargo, desde la interfaz de búsqueda on-line del AGS, los filtros no permiten realizar búsquedas sobre las clases anteriores, ya que, algunos de estos datos se encuentran agrupados en un mismo campo de información (metadato) llamado *descripción* o *notas*.

2. LA ESTRUCTURACIÓN DEL CONOCIMIENTO

En una primera etapa del proyecto DIMH se desarrolló y publicó una aplicación web para búsquedas de contenido, tanto basadas en palabras clave como configurables por el usuario, sobre algunos de los parámetros de la estructura de metadatos de los documentos del corpus o sobre el tipo de información lingüística contenida o no en el corpus (García Serrano, 2015)⁶.



Figura 2. Buscador DIMH configurable.

⁶ Aplicación accesible desde <http://albali.lsi.uned.es/DemoSimancasConfigurableWS/>.

¿Qué ventajas recibe el humanista con esta aplicación web? Al ser un buscador sobre la base de las fichas enriquecidas, no solo se busca sobre los metadatos originales, sino también sobre el contenido de los nuevos metadatos que enriquecieron la descripción.

Sin embargo, el objetivo de la estructuración de la información es identificar, ordenar y crear modelos de conocimiento sobre la base de técnicas de: clasificación, *clustering* y otras.

Las técnicas de clasificación permiten organizar colecciones de objetos, identificar sus características comunes y discriminantes y denotar los objetos de forma consistente sobre la base de las etiquetas. Un modelo de clasificación muy utilizado es la taxonomía, que es básicamente un sistema de denotación de objetos (cada una de las etiquetas de los niveles constituyen un vocabulario controlado).

Las ontologías son modelos de conocimiento que estructuran la información en una jerarquía de clases. Son difíciles de construir, pero, si se construyen y reutilizan, son muy útiles (Cimiano *et al.*, 2014: 1-178). Sin embargo, tanto en las taxonomías como en las ontologías para su construcción, es necesario disponer previamente del conocimiento relacionado con los objetos a clasificar. El consorcio internacional W3C mantiene el RDF, metalenguaje estándar para codificar ontologías sobre la base de triplets (objeto, atributo valor), denominándose base de conocimiento a una ontología con los ejemplares de sus clases (tipos o categorías) o instancias (objetos concretos, en nuestro caso, mapas, planos o dibujos).

Las ontologías sentaron las bases para la construcción de la web semántica, y algunas permiten la población de la nube denominada Linked Open Data (*LOD cloud*). Por ejemplo, el proyecto DBpedia ha generado información semántica a partir de la Wikipedia inglesa y desde el 2011 extrae información de Wikipedia en 15 de sus versiones (idiomas)⁷. Uno de ellos es el español y tiene asignado un sitio web y un SPARQL Endpoint para hacer las consultas, sobre sus 100 millones de triplets RDF⁸. Tanto en DBpedia como en cualquier otro nodo del LOD, se consulta con un lenguaje de acceso estándar como es el SPARQL, bastante lejano al lenguaje natural.

Las técnicas de *clustering* persiguen agrupar los datos de manera automática de acuerdo a similitudes, correferencias o paralelismos que se puedan encontrar entre ellos. Por ejemplo, es esperable que dos documentos compartiendo un vocabulario similar pertenezcan a la misma categoría. Relacionadas también con la organización automática de los datos, están las técnicas probabilísticas (tales como LDA), ampliamente utilizadas en el campo de las Humanidades Digitales (Yang *et al.*, 2011: 96-104; Meeks y Weingart, 2012: 1-6).

Este tipo de técnicas asume que existe una estructura latente en los datos basada en el conjunto de temáticas que estos abordan. Así mismo cada una de las temáticas es definida

⁷ Accesible desde <http://wiki.dbpedia.org/Documentation>.

⁸ Accesible desde <http://es.dbpedia.org/sparql>.

por la distribución de términos (de entre los incluidos en los datos o descripciones) que mejor la representa. Por lo tanto, cada uno de los documentos individuales puede ser expresado mediante una distribución probabilística a lo largo del conjunto de temáticas latente. La organización de los datos quedará pues reducida a encontrar automáticamente la estructura latente que mejor defina a dichos datos.

A diferencia de la técnica anterior (clasificación mediante ontologías), las aproximaciones probabilísticas o las basadas en *clustering* no requieren de información previa. Es decir, no es necesario que se establezca a priori las clases o se defina el modelo ontológico necesario para estructurar los documentos. Sin embargo, estas técnicas presentan también algunas desventajas. Es necesario parametrizar el funcionamiento de los algoritmos para crear la representación final de los datos, por ejemplo, el número de *clusters* (o grupos) en los cuales se desea estructurar los datos. Aquellos algoritmos que no lo requieren, sí que necesitan establecer de manera manual algún criterio de parada o de selección final de los grupos a generar.

Del mismo modo, las técnicas probabilísticas en general, o como *Latent Dirichlet Allocation* (LDA) en particular, necesitan que se establezca el número de temáticas que los datos abordan. Además, estas técnicas sufren el problema de dificultad en el análisis de los resultados ya que la estructura latente detectada a menudo no se corresponde con conceptos reales o entendibles por las personas. En consecuencia, la organización resultante tiene que ser analizada por expertos para interpretar su significado (Mimno, 2012: 3.1-19; Newman y Block, 2006: 753-767).

Como lo deseable es contar con técnicas que automáticamente infieran la estructura u organización desde los datos, el *Análisis de Conceptos Formales* (o FCA, por sus siglas en inglés) es una técnica, que permite organizar un conjunto de datos, mediante la inferencia de relaciones entre los mismos de acuerdo a las características que los describen (Wille, 1992: 493-515). Dado un conjunto de documentos representado por un agregado de características (texto, metadatos), FCA agrupa (en conceptos formales) a aquellos documentos que comparten el mismo grupo de particularidades. Además, es posible definir una relación de orden parcial sobre dichos conceptos formales, lo que permite ordenarlos en una representación jerárquica (retículo). La estructuración de los datos resultado, que se basa en conceptos definidos por la serie de características que han dado lugar a su creación, es fácilmente visualizable y facilita su comprensión.

3. APROXIMACIONES DE ESTRUCTURACIÓN DEL CONOCIMIENTO EN DIMH

Con objeto de mostrar el comportamiento y las funcionalidades de las diferentes aproximaciones, se presenta brevemente la aplicación de la técnica de FCA al corpus DIMH

y, a continuación, se describe con detalle el desarrollo de la ontología DIMH y su forma de uso.

3.1. La técnica FCA aplicada al corpus DIMH

En el proyecto DIMH, se ha realizado el diseño, desarrollo e implementación de un modelo basado en conceptos utilizando la técnica FCA (García Serrano y Castellanos, 2016: 379-400), que pertenece a la misma área de investigación que la técnica LDA, ampliamente utilizada en el campo de las Humanidades Digitales. Una de las ventajas sobre LDA, es que no es necesario decidir a priori el número de conceptos o nodos de la estructura en red que se genera y otra es que permite una visualización más comprensible.

La aplicación de la tecnología basada en FCA a los metadatos de la colección AGS permite la organización de los contenidos sobre la base de conceptos formales o grupos de objetos (fichas correspondientes a mapas, planos y dibujos) y sus atributos o términos elegidos automáticamente o manualmente. El proceso consiste en los siguientes pasos:

1. Extracción de información de las fichas: No toda la información contenida en las fichas era susceptible de ser utilizada en el modelado. Se decidió extraer el contenido perteneciente a los campos: publicación, referencia, notas, entidades nombradas (nes_*), temática, materia, título; y no tener en cuenta a: publicador, formato, tipo, lemas y sintagmas. Así mismo, se aplicó un proceso de eliminación de palabras vacías (sin contenido semántico).
2. Creación del contexto formal: El contexto formal se describe mediante una matriz de adyacencia (tabla) que indica cuándo un término (atributo) aparece o no en el contenido de cada ficha (objeto).
3. Reducción del contexto formal: El contexto formal generado incluye toda la terminología susceptible de representar las fichas. Sin embargo, esto da lugar a información redundante o poco significativa. Por ello, previamente a la ejecución del algoritmo de FCA se redujo el contexto formal (Cigarrán, 2008). Brevemente, la reducción consiste en encontrar aquella terminología que permita identificar más relaciones entre las fichas, sin pérdida significativa de información, mejorando tanto el tiempo de ejecución del algoritmo de FCA como sus resultados.
4. Ejecución del algoritmo de FCA: Es en esta etapa donde se lleva a cabo la generación de los conceptos formales. Sin entrar en detalles técnicos, en esta fase se generan todos los posibles (que agrupan fichas que comparten un conjunto de términos o atributos), y las relaciones de orden entre los mismos. El algoritmo

ejecutado se trata de una implementación propia del algoritmo de *Next Neighbourhoods* (Carpinetto y Romano, 2004; Cigarrán, 2008).

Realizadas las pruebas de usuarios, estos mostraron su interés en algunas agrupaciones generadas automáticamente por la técnica FCA aplicada al corpus DIMH y echaron en falta otras que ellos deseaban encontrar.

Por esta razón en el proyecto DIMH se decidió, por una parte, modificar el modelo FCA para incorporar esta información concreta y por otra parte construir una ontología, que además de la información contenida en las fichas, incorporara los detalles de interés aportados en el denominado *conocimiento previo* explicitado por los expertos humanistas del proyecto en forma de terminología.

Para tener en cuenta el conocimiento previo en el modelo FCA (ver lista completa en la sección 3.2), durante la etapa de reducción del contexto formal, cada vez que una de las entidades que aparece en dicha información es detectada en el texto de una de las fichas, se obliga a que esta entidad sea incluida como atributo relevante para la ficha, independientemente de su frecuencia de aparición y, por lo tanto, estará en el retículo generado (resultado) tras la aplicación de FCA.

Mediante este refinamiento todo el conocimiento previo que aparezca en las fichas será tenido en cuenta, y el modelo resultante organizará tanto la información deseada por los historiadores como el resto de la información aportada por las fichas.

Puesto que para explorar los conceptos formales es necesario desarrollar una interfaz, se decidió por una interfaz de búsqueda. Para permitir esta búsqueda, todos los conceptos formales generados se han indexado (almacenamiento) junto con su información. Por ejemplo, si se quiere buscar información relacionada con mapas, se obtendrían los resultados de la Figura 3, donde pueden observarse los conceptos formales asociados a los mapas del corpus DIMH (con indicación del número de fichas que incorporan). Como se puede ver, se obtiene información agrupada de diferentes ámbitos geográficos de los mapas (*Galicia, Andalucía, Madrid, Ceuta*), diferentes características (*azul, legajo*) o diferentes protagonistas (*duque, marqués de la Ensenada, magestad, Julián de Arriaga*), etc.

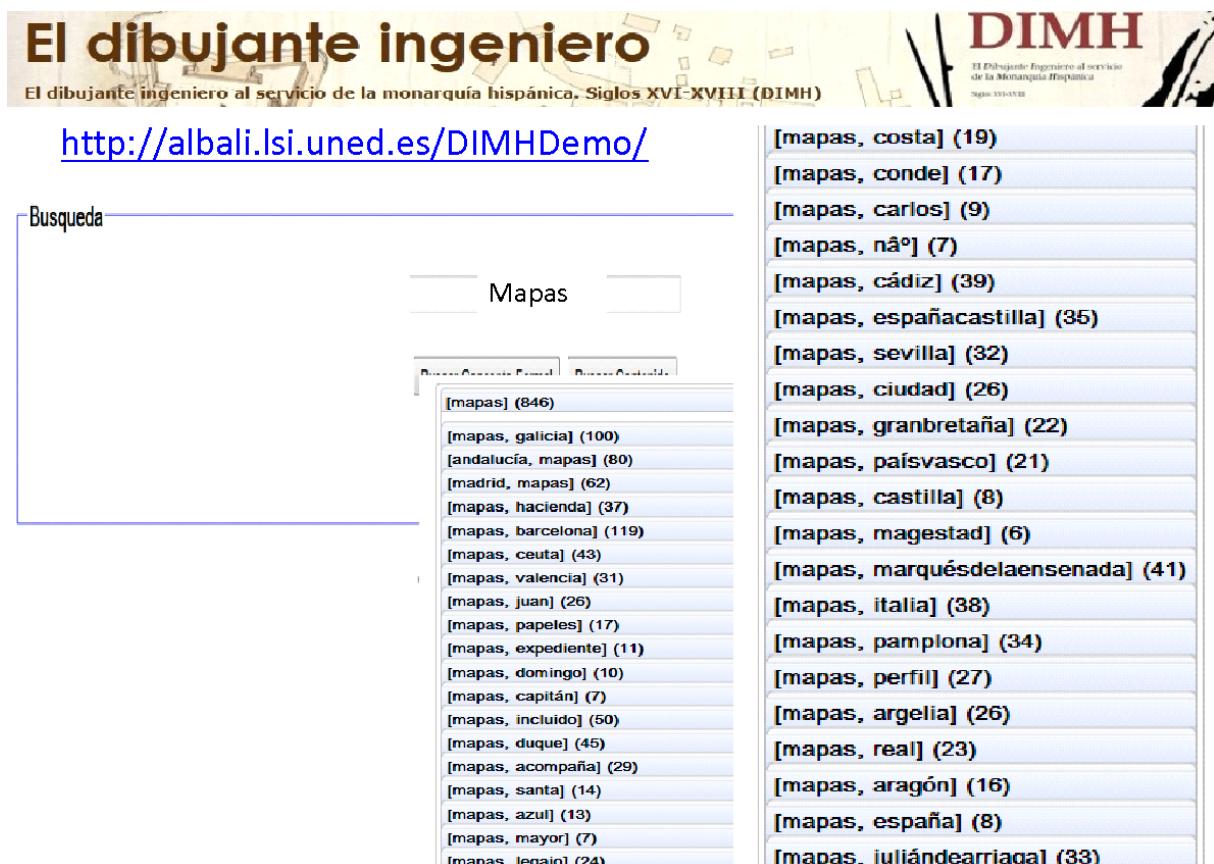


Figura 3. Navegador por los conceptos formales generados automáticamente (FCA).

La interfaz también permite buscar fichas a partir de consultas basadas en palabras clave. Para permitir esta búsqueda se han indexado todas las fichas junto con sus contenidos. Esta interfaz está dirigida a satisfacer una necesidad de información muy concreta. Por ejemplo, si se quiere buscar información de pinturas al óleo, con la búsqueda *pintura oleo* la interfaz devolverá una lista de fichas, relacionadas con esa consulta. También permite seleccionar alguna ficha de la lista para ver su contenido; y una vez que una ficha sea seleccionada, la interfaz permite visualizar los conceptos formales en los que se ha agrupado dicha ficha.

3.2. Conocimiento previo definido por los expertos humanistas

Tras el análisis de la interfaz presentada en el punto anterior, los expertos humanistas acordaron que sería interesante incluir la información que ellos ya tenían acerca de los datos presentes en la colección y que era de su interés. En concreto, la terminología seleccionada para representar el conocimiento previo de interés, en una primera versión, es la siguiente:

1. Sistema de representación (71 términos)

1.1 TÉCNICA (14 términos)

1.1.1 Base: papel; tela; entelado; manuscrito sobre papel;

1.1.2 Tipo técnica: Tinta; tinta negra; sombreado; lápiz; aguada; grabado; plumilla; colores; dibujo;

1.2 COLORES (13 términos)

sepia; marrón; azul y gris; encarnado(s); violeta; ocres; amarillo; azul ; gris; rosa de ocho vientos; rojo; lis; verde;

1.3 TIPOS DE REPRESENTACIÓN (25 términos)

1.3.1 Planta arquitectónica; plano; planta; perspectiva; alzado

1.3.2 Clasificación tipo contenido: Bosquejo; perfil(es); cuaderno/quaderno; traza/traça/trasa; croquis;

1.3.3 Clasificación funcional: proyecto/proyectado; mapa; descripciones de partes/términos; porción; detalle; rasguño; levantamiento;

1.3.4 dibujo; orientada al NO; planta ichnográfica; descripción/discreption/description

1.3.5 representación/representación; retrato/ritratto; relieve;

1.3.6 (h)orizontal; vertical; (h)exagonal;

1.3.7 relación; desinios/desegno/designio/desño/disinio/diseño;

1.4 VARIA (19 términos)

1.4.1 escala; cálculo; albañil; instrucción; autores; precios; modelo; rúbrica; conforme; emplazamiento; rosa de ocho vientos; rosa de los vientos de ocho puntas; propuesta de arreglos; con lis en rosa de ocho vientos

1.4.2 clave: alfabética; numérica; alfanumérica

1.4.3 manuscrito; distancias

2. Categorías urbanas / geográficas (7 términos)

Puerto; canal; bahía; plaza; Ciudad / citta; costa; Tierras; ciudadela; fuerte;

3. Tipologías (59 términos)

3.1 MILITAR (23 términos)

cuartele; trincheras; pabellones; murallas; castillo / castello / castel; baluarte / baluarte / semibaluarte; almacén; ciudadela; murallas; arsenal; cortina; caballero ; torreón; torre(s) / torrione; arboladuras; revellín; baterías; barracones; alcazaba vieja ; bastión/baestión ; foso; pabellones; fuerte/fortificacion(es)/fortaleza/fortezza/forte;

3.2 MARÍTIMA (7 términos)

bocana; puerto/puertto; astillero; casas de grúas; arsenal; dársena; navío;

3.3 INDUSTRIAL (5 términos)

fábrica; fábrica de tabacos; almahazen/almagacen /almacén; aserradero; nagacenes;

3.4 OBRA PÚBLICA (6 términos)

obras; casas de grúas; polea; oficinas; caminos; gradas;

3.5 ARQUITECTURA (11 términos)

edificio; muro(s); tenaza; doble tenaza; bloques; repartimientos; almacén; estrada; entrada; terraza; doble terraza;

3.6 CIUDAD (7 términos)

extramuros; villa; alcantarillado; plaza/plaça; dependencias; puerta; plaza

4. Medidas (25 términos)

toesas/tuesas/tuesas; toesas reales; codos; escala; baras; pies reales; braseaje/braceaje; varas castellanas; varas de Castilla; varas reales; varas cúbicas; cana/caña/canya; pies geométricos; pies comunes; pies castellanos; pies de Castilla; pie romano; pie; pulgada; milla; palmos; legua; legua real; ana; braza

5. Autores (nombres citados) (21 términos)

Tanlete, Jose(ph); Arredondo, Antonio de; Montalvo, Lorenzo de; Cramer, Agustín; Patiño, José; Buceta, Ventura; Martínez de la Vega, Dionisio; Próspero, Jorge (Marqués de Verboom) [1665-1744]; Cafaro, Donato Antonio; Caballero (Cavallo), Bruno; Montero de Espinosa, Jose(ph) (capitán de navío); Picchiatti, Francesco Antonio; Guemes y Horcasitas, Juan Francisco (Gobernador de La Habana); Castellón, José (Joseph); Liguera, Juan de; Marqués de la Ensenada; Domingo y Cueba, Francisco; Reggio, Andrés; Tavira, Ramón; Tiburcio Spannocchi (Espanoqui, Espanochio, Espanochi, FraTiburcio, Tribulcio Hispanochi e Hispanoqui)

3.3. Ontología DIMH

Una ontología es una conceptualización de un dominio, en este caso el dominio es el descrito por las fichas de la colección DIMH. En particular, la ontología aquí presentada se basa en la información incluida en las fichas de DIMH y en las entidades que aparecen en ellas, así como en la terminología identificada en el denominado conocimiento previo. Los conceptos principales de la ontología están representados por la jerarquía de las siguientes clases (Figura 4), con raíz la clase obligatoria inicial o raíz *thing*:

- *Ficha*: con la que el resto de las clases principales estarán relacionadas.
- *Entity*: que representa a las entidades que aparecen en las fichas. Estas entidades son de tipo: localización (*Location*), organización (*Organization*) y miscelánea (*Miscellaneous*).
- *Person*: que representa a las personas que aparecen en las fichas. Pueden ser de tipo: autores de las obras (*Authors*), autoridades que dirigían las obras (*Authorities*) y otros (*Others*).

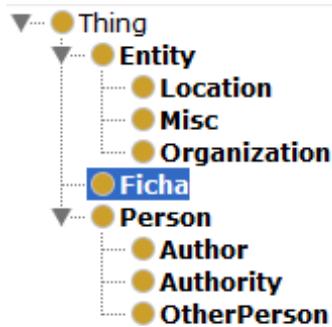


Figura 4. Jerarquía de clases de la ontología.

Para representar en la ontología aquellos campos o metadatos en las fichas que describen los diferentes tipos de información que contienen, se ha aprovechado la utilidad *data property* o propiedad del concepto o clase al que se asocia la propiedad. Así la clase *Ficha* contiene a las siguientes *data properties*:

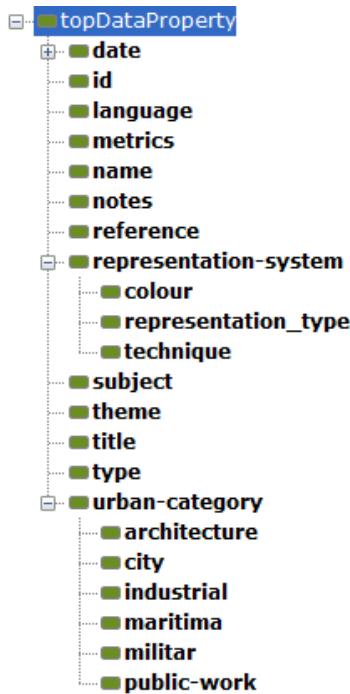


Figura 5. Jerarquía de propiedades de la clase Ficha.

Se hace notar que los campos o metadatos de las fichas en la ontología implican la forma en que se estructura el conocimiento. Sin embargo, se puede observar que en la Figura 5, aparecen dos propiedades nuevas, *representation-system* y *urban-category*, que a su vez tienen un conjunto de sub-propiedades. Estas propiedades aparecen en conocimiento previo. Las sub-propiedades pueden tomar un conjunto de valores determinados (rango). Por ejemplo, en la Figura 6, se ve la especificación de los posibles valores de la propiedad *colour*.

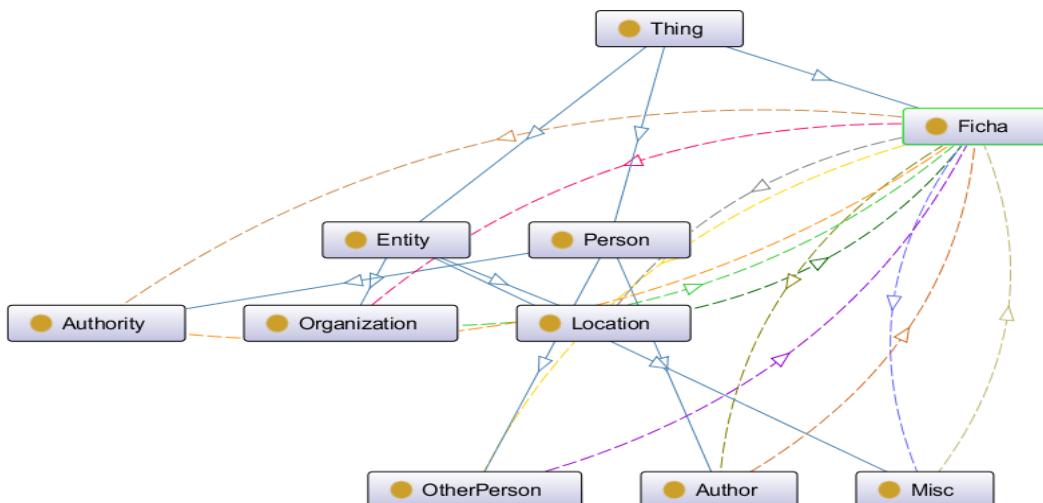
Ranges +
 ● {"AMARILLO", "AZUL", "AZUL_Y_GRIS", "COLORES", "ENCARNADO", "GRIS", "LIS", "MARRÓN", "OCRES", "ROJO", "ROSA_DE_OCHOVIENTOS", "SEPIA", "VERDE", "VERDES", "VIOLETA"} ? @ × ○

Figura 6. Rango especificado de valores.

Las clases están relacionadas mediante un conjunto de relaciones de dos tipos: *is-A* y su inversa *has-A*. Por ejemplo, si se tiene la relación entre ficha y autor *hasAuthor*, se ha de tener también la inversa *Author isAuthorOf Ficha*. Todas las relaciones definidas en la ontología pueden verse en la Figura 7. La forma en que esas relaciones conectan a las clases con su estructura se muestra en la Figura 8.



Figura 7. Relaciones definidas en la ontología.



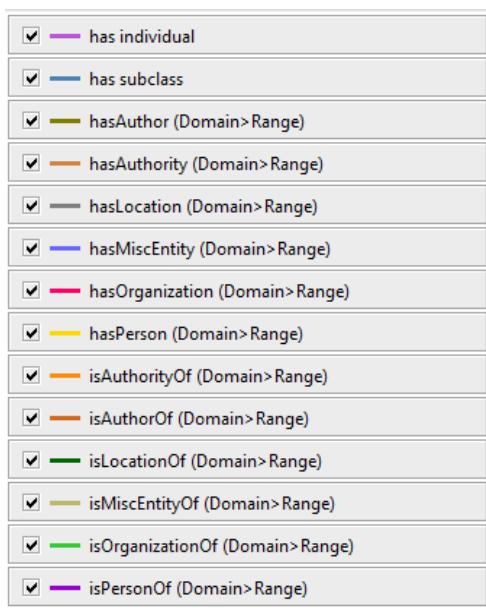


Figura 8. Relaciones entre clases de la ontología.

Estos son los componentes básicos de la ontología, clases, propiedades y relaciones y representan la estructura conceptual de un dominio de conocimiento. Para que represente a un dominio concreto, los rangos y los conceptos tienen que estar bien especificados e instanciados en objetos concretos.

3.4. Poblado de la ontología

El proceso de poblado de una ontología (*population* en inglés) consiste en la inclusión de los datos concretos (las fichas de la colección DIMH en nuestro caso y su contenido) en la estructura de la ontología (clases y relaciones). Para poblar la ontología, se ha desarrollado un analizador (*parser*) en Java (automático), que toma como entrada el contenido de las fichas (en formato XML) y puebla la ontología utilizando la librería Apache Jena Library⁹. Este *parser*, para cada ficha:

1. Lee el contenido y crea un nuevo objeto de tipo Ficha en la ontología.
2. Identifica las fechas que aparecen.
3. Identifica las menciones de los términos del vocabulario del conocimiento previo, para crear la *data property* relacionada (*representation-system* y *urban-category*).
4. Identifica los autores que aparecen, para crear un Nuevo autor (si no está aún) o relacionar la ficha con el autor, mediante las relaciones *hasAuthor* y *isAuthorOf*.

⁹ Accesible desde <http://jena.apache.org/index.html>.

5. Identifica otras entidades de persona, para crear una nueva instancia de persona o relacionar la ficha con la persona existente mediante las relaciones *hasPerson* y *isPersonOf*.
6. Identifica al resto de tipos de entidades (localización, organización o miscelánea) para crear una nueva instancia o relacionar a la ficha con la entidad existente a través de las relaciones *hasXXX* y *isXXXOf*.

En la Figura 9 puede verse el efecto producido en la ontología tras el análisis de una ficha.

Object property assertions	
■	hasPerson san_felipe
■	hasLocation orán
■	hasLocation san_felipe
■	hasLocation argelia
■	hasPerson torelli
■	hasLocation hercules
■	hasLocation plaza_de_orán

Data property assertions	
■	title "entrada encubierta que se necesita de acer alrededor de san philipe por darle defensa que no la tiene material cartográfico orán argelia fortificaciones planos 1693"
■	id "176576"
■	notes "ags guerra y marina legajos 02939 con carta del marqués de robledo al marqués de villanueva urgiendo la reparación de la construcción de la fortaleza de orán para defenderse de los turcos y mequineses orán 25 de julio de 1693 incluye una relación de los reparos y nuevas fortificaciones que necesita la plaza de orán y sus castillos orán 25 de julio de 1693 incluye una relación de los reparos y nuevas fortificaciones que necesita la plaza de orán y sus castillos nombre del autor tomado de la documentación adjunta manuscrito sobre papel sumario planta poligonal irregular del castillo de san felipe en tinta explicación cortados los extremos de las esquinas al verso pone dos plantas del castillo de san felipe castillo de sn phelipe"
■	representation_type "RELACIÓN"
■	technique "TINTA"
■	other-dates "Sat Jul 25 00:00:00 CET 1693"^^date
■	architecture "ESTRADA"
■	language "spa"
■	representation_type "PLANTA_ARQUITECTÓNICA"
■	subject "argeliaorán"
■	publication-date "Thu Jan 01 00:00:00 CET 1693"^^date
■	militar "CASTILLO"
■	militar "FUERTE"
■	technique "MANUSCRITO_SOBRE_PAPEL"
■	reference "referencias sin publicar"
■	technique "PAPEL"
■	type "MAP"
■	city "PLAZA"

Figura 9. Resultado de la incorporación de una ficha en la ontología (poblado).

La figura 10 muestra una lista de autores incluidos en la ontología.

- ◆ [Andres_Reggio](#)
- ◆ [Antonio_de_Arredondo](#)
- ◆ [Bruno_Caballero](#)
- ◆ [Francisco_Domingo_y_Cueba](#)
- ◆ [Jorge_Prospero](#)
- ◆ [Jose_Castellon](#)
- ◆ [Jose_Montero_de_Espinosa](#)
- ◆ [Jose_Patiño](#)
- ◆ [Juan_Francisco_Guemes_y_Horcasitas](#)
- ◆ [Lorenzo_Montalvo](#)
- ◆ [Ramon_Tavira](#)
- ◆ [Tiburcio_Spannocchi](#)

Figura 10. Vista parcial de los autores de la ontología.

Cada entidad, persona y autor está relacionada con cada Ficha en la que aparece (ejemplo de autor en la Figura 11).

- [isAuthorOf dimh#184119](#)
- [isAuthorOf dimh#179067](#)
- [isAuthorOf dimh#179772](#)
- [isAuthorOf dimh#183520](#)
- [isAuthorOf dimh#178718](#)
- [isAuthorOf dimh#179334](#)
- [isAuthorOf dimh#176994](#)
- [isAuthorOf dimh#180235](#)
- [isAuthorOf dimh#177582](#)
- [isAuthorOf dimh#177622](#)
- [isAuthorOf dimh#176976](#)
- [isAuthorOf dimh#184123](#)
- [isAuthorOf dimh#179066](#)
- [isAuthorOf dimh#179501](#)
- [isAuthorOf dimh#176834](#)
- [isAuthorOf dimh#179523](#)
- [isAuthorOf dimh#184117](#)
- [isAuthorOf dimh#180229](#)
- [isAuthorOf dimh#184116](#)
- [isAuthorOf dimh#179871](#)
- [isAuthorOf dimh#184124](#)
- [isAuthorOf dimh#184118](#)
- [isAuthorOf dimh#177415](#)

Figura 11. Vista parcial de las fichas cuyo autor es Andres Reggio.

4. PRACTICANDO CON LA ONTOLOGÍA

A continuación, se presentan unos ejemplos para familiarizarse con la ontología. Para ello hay que descargar la mencionada herramienta Protégé. Protégé es un software de código libre para la creación y edición de ontologías que se puede descargar¹⁰ e instalar, ya sea en su versión Frames o estándar OWL (*Web Ontology Language*) orientado a la Web Semántica (la opción preferida en este trabajo)¹¹. Tras completar la instalación basta con ejecutar el programa para acceder al editor y comenzar a desarrollar o cargar una ontología previamente desarrollada.

Existen varios tutoriales que explican de manera detallada los primeros pasos a seguir para familiarizarse con la herramienta y comenzar con la creación de ontologías, tanto para Protégé-Frames como para Protégé-OWL. Por ejemplo, el relacionado con la configuración de Protégé y la creación de una ontología¹².

4.1. Primeros pasos

Los primeros pasos están enfocados a conocer con más detalle la ontología DIMH facilitada, sus clases (*Class*), instancias (*Instance*), relaciones (*Object property*) y propiedades (*Data Properties*).

Puesto que la población de la ontología se ha hecho mediante el análisis automático del texto del corpus DIMH, es posible que haya algunos errores. Por ejemplo, en la vista de la Figura 12 pueden verse algunas de las instancias pertenecientes a la clase *Location*. Y *cadiz_de_14*, no se corresponde a una localización (la correcta sería *cádiz*) sino a un error derivado del poblado automático.

¹⁰ Accesible desde <http://protege.stanford.edu/download/download.html>.

¹¹ Accesible desde <http://www.w3.org/2004/OWL/>.

¹² Accesible desde <http://protegewiki.stanford.edu/wiki/Protege4Pizzas10Minutes>. Una explicación detallada acerca de las ontologías y cómo crear una en <http://protege.stanford.edu/...ontology101.pdf>. Otros tutoriales de uso de Protégé en <http://owl.cs.manchester.ac.uk/publications/talks-and-tutorials/protg-owl-tutorial/>.

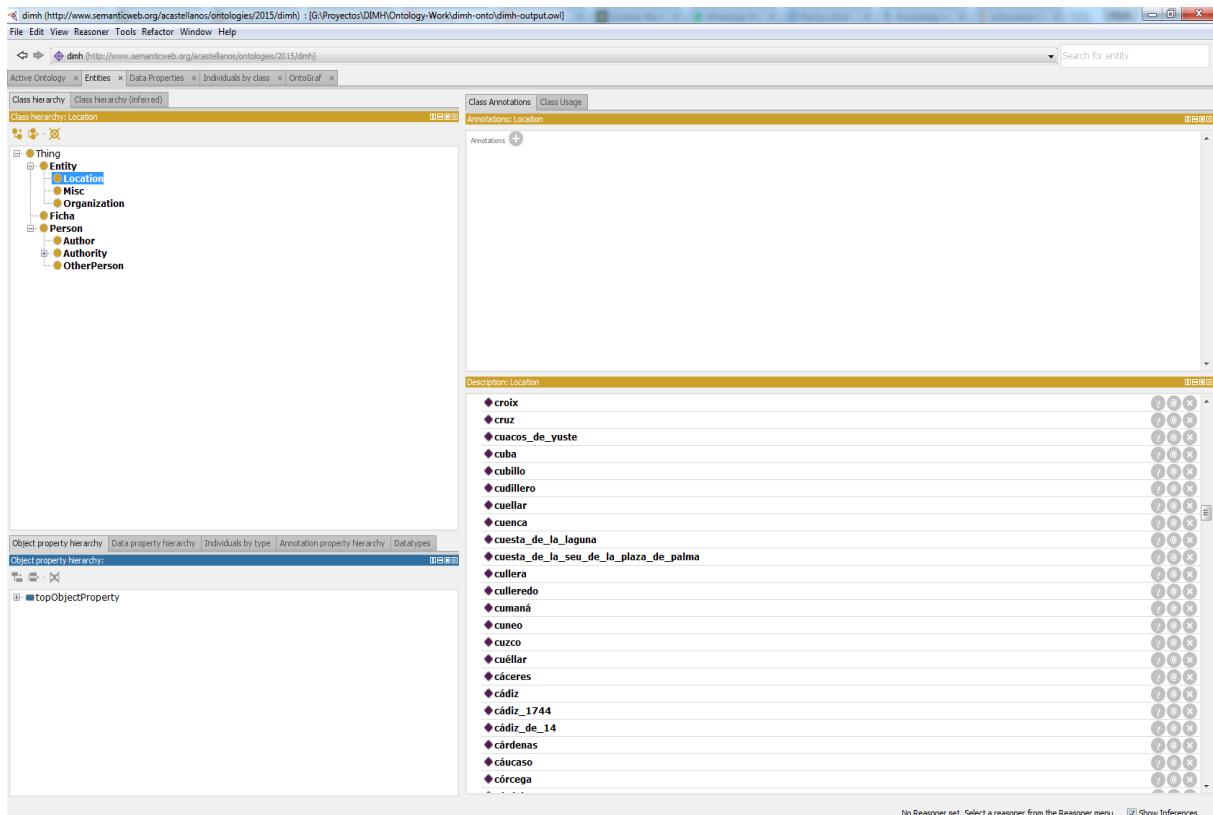


Figura 12. Vista de las entidades de tipo *Location*.

También se pueden encontrar fichas que citen en su contenido a personas, las cuales no hayan sido identificadas. No habrán sido identificadas aquellas personas para las que, a pesar de que figuren en el contenido de la ficha, no exista una relación entre ellas y la ficha. Es decir, que no exista una relación (*Object Property*) de tipo *hasAuthor*, *hasAuthority*, *hasMilitarRangeAuthority*, *hasOhterAuthority*, *hasNobilityAuthority* o *hasPerson* entre la ficha y la persona. Por ejemplo, en la ficha de la Figura 13 ocurre con *Carlos Desnaux*.

Property assertions: dimh#176607	
Object property assertions	+
hasLocation galicia	?
hasLocation la_coruña	?
hasLocation santiago_de_compostela	?
hasLocation santiago	?
hasLocation españa	?
Data property assertions	+
colour "GRIS"	?
subject "españagalicia la coruña provinciasantiago de compostela"	?
representation_type "PLANO"	?
representation_type "PORCIÓN"	?
publication-date "1751-01-01"^^dateTime	?
id "176607"	?
representation_type "PROYECTO"	?
technique "PAPEL"	?
language "spa"	?
militar "CUARTELES"	?
representation_type "PLANTA_ARQUITECTÓNICA"	?
technique "MANUSCRITO_SOBRE_PAPEL"	?
reference "referencias sin publicar"	?
title "santiago de compostela la coruña construcciones militares planos 1748 plano que expresa la porcion fabricada y la que resta sin fabricar del cuartel proyectado por el ingeniero en jefe carlos desnaux material cartográfico"	?
type "MAP"	?
notes "ags secretaría de guerra legajos 03684 nota sobre construcción de otros pesebres contiene elevación y planta de dicho cuartel coloreado en gris verde y carmín explicación con clave numérica ags secretaría de guerra legajos 03684 entre otros documentos con carta del conde de itre al marqués de la ensenada santiago 3 de marzo de 1751 título tomado de la documentación adjunta manuscrito sobre papel"	?

Figura 13. Vista de la ficha 176607.

Aquellas fichas identificadas en el punto anterior deberían ser completadas manualmente para que tengan en cuenta a las personas no identificadas. Para ello, se ha de crear la relación (*Object Property*) entre la ficha y la persona identificada según los siguientes dos casos.

Si la persona identificada ya existe entre las instancias ya definidas en la ontología (en las clases *Author*, *MilitarRanges*, *Nobility*, *Other* o *OtherPerson*), se tendrá que crear una nueva relación entre la instancia de la ficha y la instancia de la persona. La relación (*Object Property*) tendrá que ser del tipo correspondiente a la clase de la persona, por ejemplo, si la instancia de persona es de tipo *Author* la relación deberá ser de tipo *hasAuthor* y tendrá que crearse tanto esta relación como su inversa (*isAuthorOf*).

Para más información acerca de las relaciones (*Object Property*) se puede consultar la sección correspondiente en el tutorial de Protégé anteriormente citado (sección 4.6). En particular, la manera más sencilla para crear una nueva relación entre una ficha y un autor ya existentes es a través de la visualización de la propia ficha, pinchando en  en la parte de *Object Properties* (destacada en azul en la Figura 14):

The screenshot shows the Protégé interface with the following details:

- Left Panel (Description):** Shows 'Description: dimh#176576' and a 'Types' section with 'Ficha' selected. Buttons for 'Same Individual As' and 'Different Individuals' are also present.
- Right Panel (Object properties):** Titled 'Object property assertions: dimh#176576'. It lists the following assertions:
 - hasPerson san_felipe
 - hasLocation orán
 - hasLocation san_felipe
 - hasLocation argelia
 - hasPerson torelli
 - hasLocation hercules
 - hasLocation plaza_de_orán
- Bottom Panel (Data properties):** Titled 'Data property assertions'. It lists the following assertions:
 - title "entrada encubierta que se necesita de acer alrededor de san philipe por darle defensa que no la tiene material cartográfico orán fortificaciones planos 1693"
 - notes "ags guerra y marina legajos 02939 con carta del marqués de robledo al marqués de villanueva urgiendo la reparación de la construcción de la fortaleza de orán para defenderse de los turcos y mequineses orán 25 de julio de 1693 incluye una relación de los reparos y nuevas fortificaciones que necesita la plaza de orán y sus castillos orán 25 de julio de 1693 incluye una relación de los reparos y nuevas fortificaciones que necesita la plaza de orán y sus castillos nombre del autor tomado de la documentación adjunta manuscrito sobre papel sumario planta poligonal irregular del castillo de san felipe en tinta explicación cortados los extremos de las esquinas al verso pone dos plantas del castillo de san felipe castillo de sn phelipe"
 - representation_type "RELACIÓN"
 - subject "argeliaorán"
 - technique "TINTA"
 - other-dates "1693-07-25"^^dateTime
 - architecture "ESTRADA"
 - language "spa"
 - representation_type "PLANTA_ARQUITECTÓNICA"
 - id "176576"
 - militar "FUFRTF"

Figura 14. Vista de la parte de *Object Properties* de una ficha.

Si la persona identificada no estuviese incluida entre las entidades ya definidas en la ontología (en una de las clases *Author*, *MilitarRanges*, *Nobility*, *Other* o *OtherPerson*), se tendría que crear una nueva instancia para esta persona. En caso de tener que crear una nueva entidad, se considerará que es de la clase *OtherPerson*. Para más información acerca de cómo crear una nueva instancia se puede consultar la sección a este respecto en el tutorial de Protégé antes citado. En la interfaz de Protégé las instancias se crean a través de la interfaz de la clase a la que vayan a pertenecer. Esto es, hay que pinchar en la clase a la cual va a pertenecer la instancia (*OtherPerson* en este caso), y pinchar en para crear una nueva instancia (marcado en azul en la Figura 15):



Figura 15. Vista de las instancias de una clase.

4.2. Nuevos componentes en la ontología

Esta parte está enfocada a mostrar la interacción con la ontología definiendo manualmente nuevos componentes de la ontología como son: definir una nueva clase, definir nuevas relaciones, o crear instancias.

Por ejemplo, se desea definir una nueva clase en la ontología que se llame *ObjetoPrincipal* que dependa de la clase *Thing* y además que represente al objeto principal sobre el que versa una ficha, como la de la Figura 16. Allí se puede ver que el objeto principal de la ficha sería la *Fortaleza de Orán* y el objetivo de la ficha la reparación de la fortaleza.

Property assertions: dimh#176576

Object property assertions

- hasPerson san_felipe
- hasLocation orán
- hasLocation san_felipe
- hasLocation argelia
- hasPerson torelli
- hasLocation hercules
- hasLocation plaza_de_orán

Data property assertions

- title "entrada encubierta que se necesita de acer alrededor de san philipe por darle defensa que no la tiene material cartográfico orán argelia fortificaciones planos 1693"
- notes "ags guerra y marina legajos 02939 con carta del marqués de robledo al marqués de villanueva urgiendo la reparación de la construcción de la fortaleza de orán para defenderse de los turcos y mequineses orán 25 de julio de 1693 incluye una relación de los reparos y nuevas fortificaciones que necesita la plaza de orán y sus castillos orán 25 de julio de 1693 incluye una relación de los reparos y nuevas fortificaciones que necesita la plaza de orán y sus castillos nombre del autor tomado de la documentación adjunta manuscrito sobre papel sumario planta poligonal irregular del castillo de san felipe en tinta explicación cortados los extremos de las esquinas al verso pone dos plantas del castillo de san felipe castillo de sn phelipe"
- representation_type "RELACIÓN"
- subject "argeliaorán"
- technique "TINTA"
- other-dates "1693-07-25"^^dateTime
- architecture "ESTRADA"
- language "spa"
- representation_type "PLANTA_ARQUITECTÓNICA"
- id "176576"
- militar "FIIFRTF"

Figura 16. Vista de la ficha 176576.

Para más información acerca de cómo crear una nueva clase se puede consultar la sección correspondiente en el tutorial de Protégé antes citado (sección 4 del tutorial). En la interfaz de Protégé las clases se crean en la interfaz de la jerarquía de clases (Figura 17).



Figura 17. Interfaz de la jerarquía de clases.

Hay que definir las nuevas relaciones (*Object Property*) en la ontología para relacionar las entidades de esta nueva clase con las entidades de la clase *Ficha*, relaciones que se han de llamar *isObjetoPrincipal* y su inversa *hasObjetoPrincipal*, que relacionará entidades de la clase *Ficha* con entidades de la clase *ObjetoPrincipal*. Para asegurar que una relación es la inversa de la otra, se debe definir como tal en la descripción de la relación. Del mismo modo, para asegurar que solo pueda relacionar entidades de estas clases, a la hora de definir las relaciones se debe fijar como dominio y rango de la clase que corresponda, *Ficha* u *ObjetoPrincipal*. Para más información acerca de la creación de relaciones (*Object Property*) puede consultar la sección correspondiente del tutorial de Protégé antes citado (sección 4.4 del tutorial).

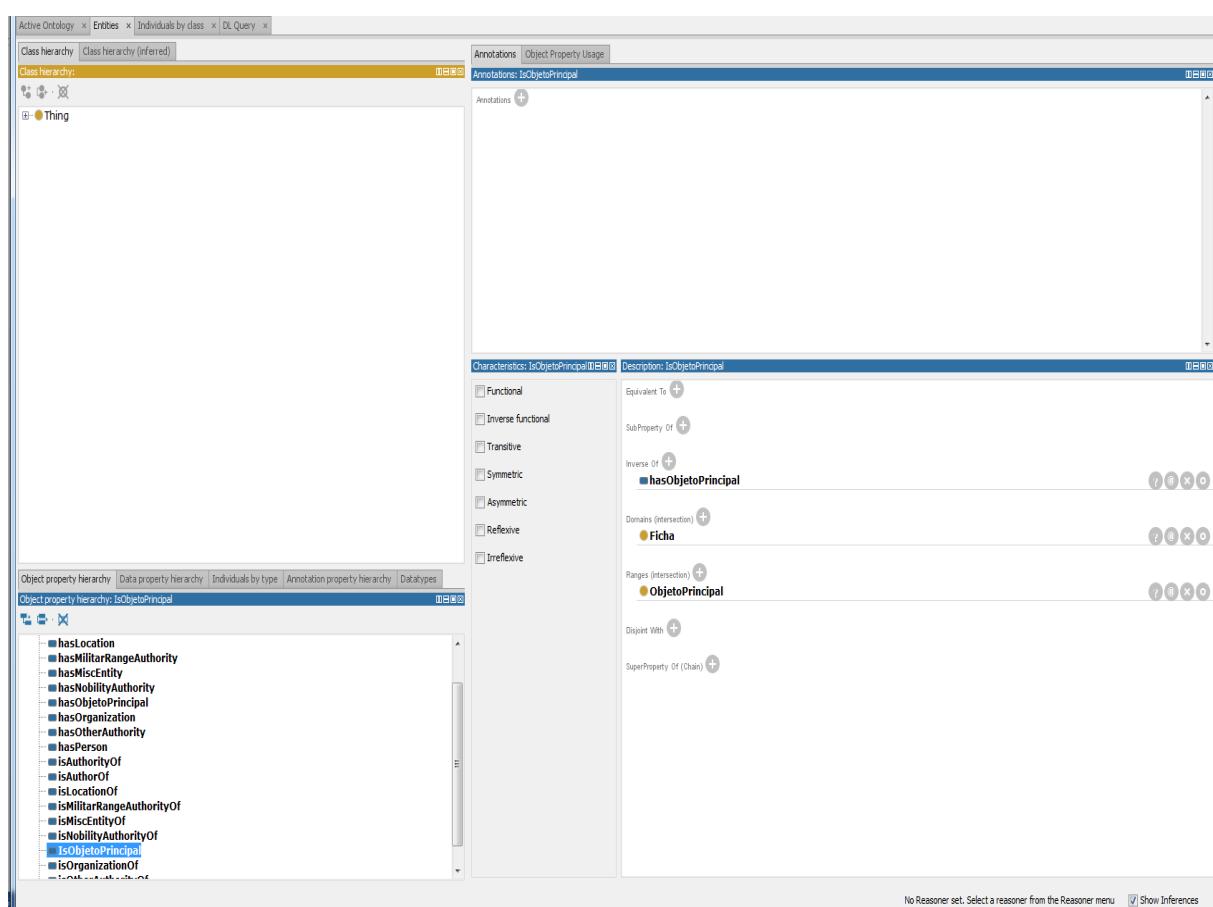


Figura 18. Vista de creación de relaciones (*isObjetoPrincipal*, *hasObjetoPrincipal*)

4.3. Consultando la ontología

Este ejemplo está enfocado a la explotación de la ontología, esto es a la ejecución de consultas estructuradas sobre el contenido de la ontología. Antes de realizar esta tarea es recomendable consultar alguna bibliografía sobre el lenguaje SPARQL¹³, también hay otros

¹³ Por ejemplo, <http://skos.um.es/TR/rdf-sparql-query/-sparqlSyntax>.

recursos que incluyen diferentes ejemplos de consultas SPARQL, ejecutadas sobre la versión española de DBpedia¹⁴.

La base de datos en formato RDF (estándar de la web semántica) que contiene la información de la ontología DIMH, se puede consultar mediante su *end-point*¹⁵. Algunas preguntas posibles (y sus respuestas) son las siguientes:

1. Seleccionar fichas donde aparece como autor *Tiburcio Spannocchi*.

PREFIX dimh:

```
<http://www.semanticweb.org/acastellanos/ontologies/2015/dimh#>
select ?ficha
where {
?ficha dimh:hasAuthor dimh:Tiburcio_Spannocchi
}
```

Ficha
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 176870
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 179801
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184070
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184073
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184078
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184085
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184100
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 191574
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 191581

2. Seleccionar fichas de *edificios militares en Aragón* y mostrar los autores.

PREFIX dimh:

```
<http://www.semanticweb.org/acastellanos/ontologies/2015/dimh#>
select ?ficha ?author ?militar
where {
?author rdf:type dimh:Author.
?ficha dimh:hasAuthor ?author.
?ficha dimh:hasLocation dimh:aragón.
?ficha dimh:militar ?militar
}
```

¹⁴ Se puede consultar en <http://es.dbpedia.org/...Ejemplos de consultas SPARQL>.

¹⁵ En <http://albali.lsi.uned.es:8890/sparql>.

Ficha	author	militar
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 179999	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Jorge_Prospere	"CUARTELES"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 180454	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Jose_Pati%C3%B1o	"FUERTE"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 180454	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Jose_Pati%C3%B1o	"TORREÓN"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 181537	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Jorge_Prospere	"ALMACÉN"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184073	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Tiburcio_Spannocchi	"CASTILLO"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184073	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Tiburcio_Spannocchi	"FUERTE"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184073	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Tiburcio_Spannocchi	"TORREÓN"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184078	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Tiburcio_Spannocchi	"FUERTE"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184100	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Tiburcio_Spannocchi	"FUERTE"
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184100	http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - Tiburcio_Spannocchi	"MURALLAS"

3. Seleccionar fichas de *Tiburcio Spanocchi* de *edificios militares*, ordenadas por fecha.

PREFIX dimh:

<<http://www.semanticweb.org/acastellanos/ontologies/2015/dimh>>

PREFIX xsd: <<http://www.w3.org/2001/XMLSchema>>

select ?ficha ?militar ?date

where {

?ficha dimh:hasAuthor dimh:Tiburcio_Spannocchi.

?ficha dimh:militar ?militar.

?ficha dimh:publication-date ?date.

}

Ficha	militar	date
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184070	"FOSO"	1602-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184070	"FUERTE"	1602-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184070	"MURALLAS"	1602-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 176870	"BALUARTE"	1595-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 176870	"CASTILLO"	1595-01-01T00:00:00+01:00

http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 176870	"FUERTE"	1595-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 179801	"FUERTE"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184073	"CASTILLO"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184073	"FUERTE"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184073	"TORREÓN"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184085	"CASTILLO"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184085	"FUERTE"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184085	"TORREÓN"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 191574	"CASTILLO"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 191574	"FUERTE"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 191574	"TORREÓN"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 191581	"CASTILLO"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 191581	"FUERTE"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 191581	"TORREÓN"	1594-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184078	"FUERTE"	1593-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184100	"FUERTE"	1593-01-01T00:00:00+01:00
http://www.semanticweb.org/acastellanos/ontologies/2015/dimh - 184100	"MURALLAS"	1593-01-01T00:00:00+01:00

5. CONCLUSIONES

En este trabajo se ha abordado la representación y gestión de datos relacionados con el ámbito de las Humanidades Digitales (en particular de un conjunto de documentos históricos) para su posterior acceso, visualización y explotación en el marco del proyecto *El dibujante ingeniero al servicio de la monarquía hispánica. Siglos XVI-XVIII* (DIMH).

Para ello, se han presentado diferentes aproximaciones de estructuración automática de los datos. Sobre la base de estas aproximaciones, se ha mostrado cómo aplicar la técnica FCA y sus interfaces de y visualización y búsqueda y, en segundo lugar, se ha descrito con cierto detalle el desarrollo de una ontología para la representación automática de los datos,

integrando el conocimiento previo sobre los datos aportados por los expertos humanistas del proyecto DIMH.

Mediante la descripción detallada de dicha ontología, la presentación de una serie de ejemplos prácticos para la familiarización de esta tecnología y su posible refinamiento manual y la definición y ejecución de varias consultas SPARQL sobre la misma, se ha pretendido mostrar la utilidad de una ontología en la representación de contenidos en el ámbito de las Humanidades Digitales. Con ello se espera facilitar la exploración y entendimiento de los datos en la colección DIMH, tanto por usuarios no expertos, como por expertos humanistas en la temática de la colección de planos, mapas y dibujos.

AGRADECIMIENTOS

Los autores agradecen la colaboración de la profesora catedrática de la UNED, Dña Alicia Cámara, sin cuyo excelente trabajo en Humanidades, no podríamos haber realizado el presentado. Este trabajo ha sido parcialmente financiado por los proyectos españoles: Musacces (S2015/HUM3494) y VEMODALEN (TIN2015-71785-R).

REFERENCIAS BIBLIOGRÁFICAS

- CARPINETO, C. y ROMANO, G. (2004). *Concept Data Analysis: Theory and Applications*. New Jersey: John Wiley & Sons.
- CASTELLANOS, A., GARCÍA SERRANO, A. y CIGARRÁN, J. (2015). "Concept-based Organization for Semi-Automatic Knowledge Inference in Digital Humanities: Modelling and Visualization". En *Actas de la 10th Conference of the International Society for Knowledge Organization (ISKO)*, France. Recuperado de https://www.researchgate.net/profile/Ana_Garcia-Serrano/publication/291102972_Concept-based_Organization_for_semi-automatic_Knowledge_Inference_in_Digital_Humanities_Modelling_and_Visualization/links/569e109c08ae950bd7a863fe.pdf el 29/04/2017.
- CIGARRÁN, J. (2008). *Agrupación de Resultados de Búsqueda Mediante Análisis Formal de Conceptos*. Tesis doctoral, UNED.
- CIMIANO, P., UNGER, C. y McCRAE, J. (2014). "Ontology-Based Interpretation of Natural Language". En *Synthesis Lectures on Human Language Tech*, 7.2, 1-178.

- DRUCKER, J., KIM, D., SALEHIAN, I. y BUSHONG, A. (2014). *Introduction to digital Humanities. Concepts, Methods, and Tutorial for Students and Instructors*. Recuperado de <https://archive.org/Int...ToDigitalHumanities> el 27/04/2017.
- GARCÍA SERRANO, A. (2015). *Some Hints on the Modelling and Visualization at the DIMH Spanish Project*. Comunicación oral Nº 48 en el EADHDay (European Association for Digital Humanities Day), celebrado en la UNED el 8 de octubre de 2015, Recuperado de <https://canal.uned.es/mmobj/index/id/45319> el 28/04/2017.
- GARCÍA SERRANO, A. y CASTELLANOS, A. (2016). "Conceptualización, acceso y visibilidad de la información en el proyecto DIMH". En *El dibujante ingeniero al servicio de la monarquía hispánica. Siglos XVI-XVIII*, Cámara Muñoz, A. (ed.), 379-400. Madrid: Fundación Juanelo Turriano. Recuperado de https://issuu.com/juaneloturriano/docs/el_dibujante_ingeniero-issuu?e=1641776/35252175 el 28/04/2017.
- MEEKS, E. y WEINGART, S. (2012). "The Digital Humanities Contribution to Topic Modeling". *Journal of Digital Humanities*, 2.1, 1-6. Recuperado de <http://journalofdigitalhumanities.org/2-1/dh-contribution-to-topic-modeling/> el 29/04/2017.
- MIMNO, D. (2012). "Computational Historiography: Data Mining in a Century of Classics Journals". *Journal on Computing and Cultural Heritage*, 5.1, 3.1-19.
- NEWMAN, D.J., BLOCK, S. (2006). "Probabilistic Topic Decomposition of an Eighteenth-Century American Newspaper". *Journal of the American Society for Information Science and Technology*, 57. 6, 753-767.
- RUTHVEN, I. y CHOWDHURY, G. (2014). *Cultural Heritage Information Access and Management*. Londres: Facet Publishing. Recuperado de http://www.facetpublishing.co.uk/title.php?id=9306&category_code=402 el 28/04/2017.
- WILLE, R. (1992). "Concept Lattices and Conceptual Knowledge Systems". *Computers & Mathematics with Applications*, 23.6, 493-515. Recuperado de <http://ac.els-cdn.com/pdf> el 28/04/2017.
- YANG, T.I., TORGET, A.J. y MIHALCEA, R. (2011). "Topic Modeling on Historical Newspapers". En *5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 96-104. Oregon: ACL. Recuperado de <http://anthology.aclweb.org/W/W11/W11-15.pdf - page=108> el 28/04/2017.