

# Trabajo Práctico N°4: Ontologías

Julieta Francisconi - Franco Santellán

## CONSIGNA 1

Seleccione seis ontologías pertenecientes a diferentes categorías.

Elabore un informe conteniendo los siguientes datos:

- Nombre – URL.
- Grupo u organismo de desarrollo
- Categoría. En caso de ser de dominio, especifique de qué dominio se trata.
- Descripción.
- Cantidad de términos (aprox.).
- Taxonomía (jerarquía de componentes) o modelo

### • Ontología 1

|                                |   |
|--------------------------------|---|
| <b>Nombre URL</b>              | <b>HydrOntology</b><br>URL: <a href="http://mayor2.dia.fi.upm.es/oeg-upm/files/hydrontology/hydrOntology_GeoLinkedData.owl">http://mayor2.dia.fi.upm.es/oeg-upm/files/hydrontology/hydrOntology_GeoLinkedData.owl</a>   |
| <b>Organismo de Desarrollo</b> | Desarrollado por: Vilches – Blázquez en 2009  |
| <b>Categoría</b>               | Se trata de una ontología global del dominio hidrográfico.  |
| <b>Descripción</b>             | Inicialmente, esta ontología fue creada como una ontología local que establecía correspondencias entre diferentes fuentes de datos (catálogos de características, diccionarios geográficos, etc.) del Instituto Geográfico Nacional Español (IGN-E). Su propósito era servir como un marco de armonización entre los productores cartográficos españoles. Más tarde, la ontología se ha convertido en una ontología de dominio global y que trata de cubrir la mayor parte de los conceptos del dominio hidrográfico. |
| <b>Cantidad de Términos</b>    | <i>No se especifica</i>   |
| <b>Modelo</b>                  | <i>No se especifica</i>   |

### • Ontología 2

|                                |  |
|--------------------------------|--|
| <b>Nombre URL</b>              | <b>PhenomenOntology</b><br>URL: <a href="http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/ontologies/152-phenomenontology/index.html">http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/ontologies/152-phenomenontology/index.html</a>  |
| <b>Organismo de Desarrollo</b> | Desarrollado por: OntologyEngineeringGroup (OEG) en colaboración con el Instituto Geográfico Nacional (IGN)  |
| <b>Categoría</b>               | Se trata de una ontología de dominio geoespacial.  |
| <b>Descripción</b>             | El desarrollo de PhenomenOntology está basado en distintos recursos heterogéneos de información geoespacial pertenecientes al IGN, tales como: Nomenclátor Conciso, Nomenclátor Geográfico Nacional, Base Cartográfica Numérica 1:200.000, Base Cartográfica Numérica 1:25.000 y Base Topográfica Numérica 1:25.000. |
| <b>Cantidad de Términos</b>    | <i>No se especifica</i>  |
| <b>Modelo</b>                  | <i>No se especifica</i>  |

### • Ontología 3

|                                |   |
|--------------------------------|---|
| <b>Nombre URL</b>              | <b>Ontología Gestión de Recursos Humanos</b><br><b>URL:</b> <a href="http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/ontologies/99-hrontology/index.html">http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/ontologies/99-hrontology/index.html</a> |
| <b>Organismo de Desarrollo</b> | Desarrollado por: Boris Villazón-Terrazas, Jaime Ramírez y Asunción Gómez-Pérez.  |
| <b>Categoría</b>               | Se trata de una ontología de dominio: Recursos Humanos.   |
| <b>Descripción</b>             | La Ontología Gestión de Recursos Humanos, actuará como un "lenguaje común" mediante un conjunto de vocabularios controlados para describir los detalles de una oferta de trabajo y el currículum vitae del demandante de empleo.                    |
| <b>Cantidad de Términos</b>    | La ontología está compuesta por trece ontologías modulares. La ontología tiene 3206 conceptos, 277 atributos, 1428 instancias y 1592 axiomas (la mayoría de ellos para establecer valores de atributos de clase)                                    |
| <b>Modelo</b>                  |   |

• Ontología 4

|                                |  |
|--------------------------------|--|
| <b>Nombre URL</b>              | <b>Web semántica:</b> un nuevo enfoque para la organización y recuperación de información en el web.<br><b>URL:</b> <a href="http://bvs.sld.cu/revistas/aci/vol13_6_05/aci030605.htm">http://bvs.sld.cu/revistas/aci/vol13_6_05/aci030605.htm</a>  |
| <b>Organismo de Desarrollo</b> | Lic. Keilyn Rodríguez Perojo y Lic. Rodrigo Ronda León   |
| <b>Categoría</b>               | Categoría de Dominio: Web Semántica  |
| <b>Descripción</b>             | Esta ontología especifica la estructura general de la web semántica y sus principales componentes.   |
| <b>Cantidad de Términos</b>    | 21 términos  |
| <b>Modelo</b>                  | <pre> graph TD     SGML[SGML] -- "estándar definido por" --&gt; XML[XML]     XML -- "es la base sintáctica de" --&gt; WS[Web Semántica]     XML -- "define la sintaxis de" --&gt; RDF[RDF]     RDF -- "es la base sintáctica de" --&gt; WS     RDF -- "es una extensión de" --&gt; WA[Web Actual]     WA -- "definida en" --&gt; HTML[HTML]     RDF -- "actúa como empaquetador de" --&gt; MET[METADATOS]     MET -- "definidos en" --&gt; EAD[EAD]     MET -- "definidos en" --&gt; DC[Dublin Core]     MET -- "definidos en" --&gt; TEI[TEI]     MET -- "definidos en" --&gt; MARC[MARC]     RDF -- "esquemas de" --&gt; URI[URI]     URI -- "localizar mediante" --&gt; WS_S[WEB SERVICES]     WS_S -- "interoperan entre" --&gt; WS_S     WS_S -- "interoperan entre" --&gt; AI[AGENTES INTELIGENTES]     AI -- "establecer" --&gt; RC[REDES DE CONFIANZA]     AI -- "trabajando con" --&gt; FD[FIRMAS DIGITALES]     AI -- "entre" --&gt; APL[APLICACIONES]     XML -- "posibilita la estructuración de" --&gt; TAX[TAXONOMIA]     XML -- "posibilita crear" --&gt; TM[TOPIC MAPS]     TAX -- "legibles por humanos como" --&gt; TES[TESAUROS]     TAX -- "legibles por máquina como" --&gt; TM2[TOPIC MAPS]     TES -- "se combinan par aformar" --&gt; ONT[ONTOLOGIAS]     TM2 -- "se combinan par aformar" --&gt; ONT     ONT -- "permitirán a" --&gt; WS_S     ONT -- "permitirán a" --&gt; AI   </pre> |



## • Ontología 6

|                                |  |
|--------------------------------|--|
| <b>Nombre URL</b>              | <b>Mantis</b><br>URL: <a href="https://www.yumpu.com/es/document/view/19123696/departamento-de-informatica-universidad-de-castilla-la-mancha-/5">https://www.yumpu.com/es/document/view/19123696/departamento-de-informatica-universidad-de-castilla-la-mancha-/5</a>  |
| <b>Organismo de Desarrollo</b> | MANTIS Software Company  |
| <b>Categoría</b>               | Categoría de Dominio: Mantenimiento de Software  |
| <b>Descripción</b>             | Ontología para la gestión de proyectos de mantenimiento de software  |
| <b>Cantidad de Términos</b>    | 8 aproximadamente  |
| <b>Modelo</b>                  | <pre> graph TD     Proyecto[Proyecto] -- "+proyecto" --&gt; Proyecto     Proyecto -- "dividir" --&gt; Proyecto     Proyecto -- "+sub-proyecto" --&gt; Proyecto     Proyecto -- "engloba" --&gt; Elemento[Elemento]     Proyecto -- "pretende" --&gt; Objetivo[Objetivo]     Proyecto -- "limita-a" --&gt; Precondición[Precondición]     Precondición -- "se refiere-a" --&gt; Restricción[Restricción]     Restricción -- "referencia-a" --&gt; Elemento     Artefacto[Artefacto] --&gt; Elemento     Actividad[Actividad] --&gt; Elemento     Recurso[Recurso] --&gt; Elemento     Agente[Agente] --&gt; Elemento   </pre> |

## CONSIGNA 2

¿Qué es una ontología lingüística? Explique y ejemplifique.

La ontología lingüística es una ontología, cuyos conceptos están considerablemente relacionados con los significados de las unidades lingüísticas, los términos del campo temático. La misma puede ser considerada como un tipo especial de base de datos léxica y un tipo especial de ontología. El papel de las ontologías lingüísticas aumenta en aplicaciones relacionadas con el procesamiento del lenguaje natural. Algunos ejemplos son: WordNet, MicroKosmoc, entre otros.

## CONSIGNA 3

Según lo expresado por Oscar Corcho en su artículo “Ontologybaseddocumentannotation: trends and open researchproblems”, responda:

a. En los que respecta a las ontologías Dublin Core y FOAF (Friend of a Friend):

i. ¿Cuál es el dominio de aplicación?

Según Oscar Corcho, DublinCore (<http://www.dublincore.org/>) está siendo utilizada para especificar las características de documentos electrónicos. Donde se especifican un conjunto predefinido de características del documento como por ejemplo creador, fecha, contribuyente, descripción, formato, etc.

Por otro lado, dado que las ontologías FOAF son una red de documentos que describen una red de personas (y otras cosas), su mayor aplicación se ve mayormente en redes sociales.

ii. ¿En qué lenguaje están implementadas?

Las anotaciones de DublinCore se pueden implementar en lenguajes como RDF y XML. FOAF se implementa en RDF Schema.

iii. ¿En qué lugar se encuentran dentro de la progresión semántica propuesta por Lassila y Mc Guinness (2001)? Justifique.

DublinCore se encuentra al inicio de la progresión, esto seguramente se debe a que fue creado con el objetivo de generar metadatos en un documento.

Por otro lado, FOAF se encuentra en la mitad de la progresión ya que se trata de una ontología más estructurada que contiene clases básicas como por ejemplo Agente, Persona, Organización, Grupo, Proyecto, Documento, Imagen, etc.

b. Explique las relaciones y diferencias semánticas entre diferentes anotaciones: Dublin Core – Tesauro sobre información geográfica (GettyThesaurus of GeographicNames) – Ontología en el dominio de los viajes. Ejemplifique.

*Algunas relaciones semánticas entre Dublin Core, Tesauro sobre información geográfica y Ontología en el dominio de los viajes son:*

Las anotaciones de DublinCore describen principalmente propiedades del propio documento sin proporcionar demasiados detalles sobre su contenido (sólo algunas palabras clave y descripciones de lenguaje natural en propiedades como asunto o descripción). En cambio, las anotaciones basadas en ontologías se dedican a describir el contenido del documento, y no sus propiedades generales. Por último, los tesauros y los vocabularios controlados pueden utilizarse en ambos enfoques para proporcionar

términos acordados en ámbitos específicos. En consecuencia, todos los enfoques se complementan entre sí.

En cuanto a las diferencias, las anotaciones de DublinCore son más ambiguas que las anotaciones basadas en un tesoro o vocabulario controlado, y éstas son también más ambiguas, en general, que las anotaciones basadas en ontologías. Por ejemplo, DublinCore recomienda las mejores prácticas (no normativas) para la mayoría de los valores que se utilizarán al describir documentos; Las anotaciones basadas en tesauros proporcionan directrices claras sobre los términos a utilizar; Y finalmente las anotaciones basadas en ontologías normalmente incluyen instancias de relación que dan 'se refieren a clear', mientras que en enfoques basados en ontología algunos de estos valores serán referencias a otras instancias en la ontología.

Para ejemplificar se toma la página HTML mostrada en la Figura que pertenece al sitio web de una agencia de viajes y resume la información sobre un vuelo desde Madrid a Seattle el 8 de febrero de 2003.

### Flight details

|                 |  |
|-----------------|--|
| <b>Outbound</b> | <p>Leaving from <b>Madrid</b> - Barajas - Spain<br/>on Saturday 08 February 2003 at <b>11:50</b><br/>Arriving in <b>Chicago</b> - O'Hare International - United States of America<br/>same day at <b>14:10</b><br/>Airline: American Airlines<br/>Flight No. AA 7615<br/>Type of aircraft: Airbus Industrie A340 All Series PAX/H</p> <p>Leaving from <b>Chicago</b> - O'Hare International - United States of America<br/>on Saturday 08 February 2003 at <b>16:48</b><br/>Arriving in <b>Seattle</b> - Seattle/Tacoma International - United States of America<br/>same day at <b>19:23</b><br/>Airline: American Airlines<br/>Flight No. AA 1605<br/>Type of aircraft: non referenced/B</p> |
|-----------------|--|

#### Ejemplo de uso de DublinCore para la anotación de documentos

Si anotáramos esta página HTML con DublinCore, incluiríamos información como: el colaborador y creador (servicio de reserva de vuelos [www.flightbookings.com](http://www.flightbookings.com)), la fecha (1 de enero de 2003), una descripción ("detalles de vuelo para un viaje entre Madrid y Seattle vía Chicago el 8 de febrero de 2003"), el formato del documento (HTML), el idioma del documento ('en', Inglés) entre otros.

#### Ejemplo de uso de tesauros para la anotación de documentos

Supongamos que queremos anotar el documento con un tesoro como el de Nombres Geográficos (TGN) de Getty. En este caso incluiríamos información como la siguiente:

- Madrid es una referencia al término con la ID 7010413 en el tesoro, que se refiere a la ciudad de Madrid en España
- España es una referencia al término con la ID 1000095, que se refiere al Reino de España en Europa



- Chicago es una referencia al término con la identificación 7013596, que refiere a la ciudad de Chicago en Illinois, los EEUU
- Estados Unidos de América es una referencia al término "Estados Unidos" con la identificación 7012149, que se refiere a la nación de Estados Unidos
- Seattle es una referencia al término con la identificación 7014494, que refiere a la ciudad de Seattle en Washington, los EEUU.

Este no es el único tesoro que podemos usar para anotar el documento. También podríamos usar los códigos IATA (<http://www.iata.org/index.htm>) para referirnos a los diferentes aeropuertos que aparecen en el documento:

- Barajas es una referencia al código IATA MAD O'Hare International es una referencia al código IATA CHI
- Seattle / Tacoma International es una referencia al código IATA SEA.

Del mismo modo, podríamos usar otros tesoros para referirse a nombres de aerolíneas, modelos de avión, etc.

### Ejemplo de uso de ontologías para la anotación de documentos

Supongamos que existe una ontología en el dominio itinerante que queremos utilizar para anotar el documento en la Figura 3. Esta ontología contendría conceptos como Vuelo, Ubicación, Aeropuerto, etc., y propiedades como lugar de salida y llegada, precio de boleto, etc. Las anotaciones de documentos basadas en ontologías suelen contener tres tipos de información: 1) *Las instancias de concepto*, que relacionan una parte del documento con uno o varios conceptos en una ontología. Por ejemplo, 'Detalles de vuelo' puede representar una instancia del concepto Vuelo, y puede llamarse AA7615\_Feb08\_2003, aunque las instancias de concepto no tienen necesariamente un nombre. 2) *Los valores de atributo*, que relacionan una instancia de concepto con parte del documento, que es el valor de uno de sus atributos. Por ejemplo, 'American Airlines' puede ser el valor del atributo companyName. 3) *Instancias de relación*, que relacionan dos instancias de concepto por alguna relación específica de dominio. Por ejemplo, el vuelo AA7615\_Feb08\_2003 y la ubicación Madrid pueden conectarse mediante la relación departurePlace.

c. ¿Para qué se utilizan las recomendaciones de W3C RDF, RDFS y OWL?  
¿Cómo se utilizan? Indique brevemente.

El lenguaje RDF se utiliza para realizar anotaciones basadas en ontologías. Estas anotaciones pueden basarse en ontologías implementadas en RDF Schema u OWL.

RDF, RDF Schema y OWL son recomendaciones del W3C.

Un código RDF puede referirse bien a un esquema RDF o a una ontología OWL, y debe insertarse también dentro del código HTML de la página Web o en un recurso Web diferente que se refiera al que se está anotando.



#### d. ¿Qué son las herramientas para anotación de metadatos basadas en ontologías?

Según Corcho, las herramientas de anotación basadas en ontologías están diseñadas principalmente para permitir la inserción y mantenimiento de marcas basadas en ontologías en páginas Web.

La mayoría de estas herramientas surgieron con la aparición de la Web Semántica. Fueron concebidas por primera vez como herramientas que podrían ser utilizados para aliviar la carga manual de anotaciones basadas en ontologías en las páginas web. Desde entonces han evolucionado con entornos más completos que utilizan técnicas de Extracción de Información (IE) y Aprendizaje Automático (ML) para proponer anotaciones semiautomáticas para documentos Web.

#### i. Seleccione una herramienta. Describa su funcionamiento. Ejemplifique.

##### ***OntoGloss***

Es una herramienta de anotación basada en ontología que utiliza conceptos predefinidos en ontología para marcar un documento. Tal anotación proporciona: adecuación expresiva, adecuación/adaptación semántica, incrementabilidad, uniformidad, franqueza, extensibilidad, legibilidad humana y Consistencia.

OntoGloss utiliza los conocimientos lingüísticos recogidos a través de anotaciones por la comunidad para anotar automáticamente otros documentos. Para cualquier documento anotado, un conjunto de triples de RDF se crea y se guarda en la base de datos. En la próxima visita al mismo documento, OntoGloss recupera todos los triples para el documento de la base de datos y marca todas las secciones anotadas. Mientras la estructura del documento no cambie drásticamente (lo que suele ser el caso en lingüística) esto crearía las mismas secciones anotadas.

OntoGloss utiliza UniformResourceIdentifies (URI) para identificar los recursos y representa las relaciones entre ellos. Mantiene las anotaciones separadas de los documentos reales y admite dos modos de operaciones: local y remoto. En el modo local, los datos anotados se guardan localmente y se utilizan en la anotación de documentos que se visitan por primera vez. En el modo de servidor de anotaciones remoto o compartido, el lingüista puede agregar sus datos anotados a un servidor para que la comunidad lo use.

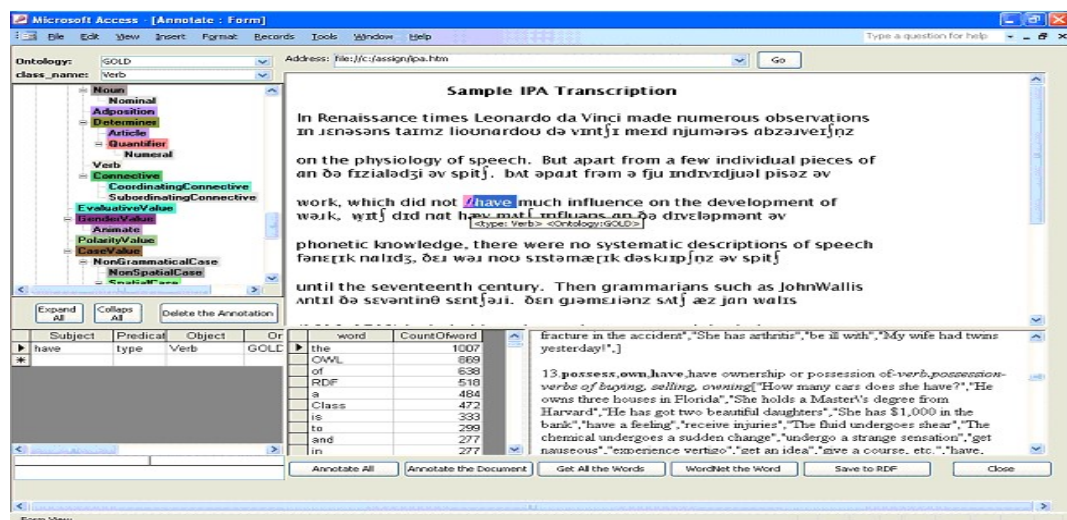
##### **OntoGloss tiene las siguientes características:**

- Usa diferentes ontologías para marcar documentos, párrafos, frases, palabras y morfemas. Es independiente de la ontología seleccionada y puede acomodar varias ontologías al mismo tiempo.
- Con la operación de arrastrar y soltar se realizan anotaciones en el documento. Al pasar el ratón sobre una selección anotada, el lingüista puede ver el tipo de anotación.

- Realiza anotaciones automáticamente nuevos documentos basados en los documentos previamente anotados.
- Tiene la capacidad de utilizar un sistema de referencia léxica. Este sistema de referencia léxica podría ya existir, p. WordNet 2.6, o puede ser construido y agregado gradualmente dentro de OntoGloss. Al igual que WordNet para el idioma inglés, este sistema puede usarse como un recurso durante el proceso de anotación proporcionando sinonimia, hiponimia y diferentes sentidos para palabras individuales.
- Exporta datos de anotación en formato RDF. Los datos RDF se pueden cargar en un repositorio RDF como Sesame 4 con capacidades de consulta.
- Mantiene la anotación separada del documento real. Los datos anotados se guardan en una base de datos y se cargan durante cada visita al documento.
- Anotar todo el documento con información general como el nombre del anotador, la fecha y otra información especificada en el DublinCore.
- Soporta servidores de anotación locales y remotos.

Cuando se visita un documento por primera vez, OntoGloss compara cada palabra con todo el texto anotado en la base de datos y asigna el mismo tipo de anotación a las palabras. Esto servirá como una sugerencia inicial y puede ser cambiado por el lingüista si es necesario. Las clases en la ontología están codificadas por colores. Un texto anotado tiene el mismo color que la clase que se utiliza en la anotación. Esto da una pista visual al lingüista sobre el tipo de marcado.

Hay muchos anotadores de texto disponibles tanto código abierto como productos comerciales. Lo diferente de un anotador lingüístico es que las palabras de la lingüística se dividen en morfemas. OntoGloss es capaz de anotar morfemas en una palabra. Por ejemplo, xxxabcestá compuesto de xxx más un sufijo -abc, un lingüista que utiliza OntoGloss es capaz de anotar cada morfema por separado. En la anotación automática de nuevos documentos, cuando OntoGloss encuentra yyyabc, puede determinar si tiene el mismo sufijo y anotarlo con la misma clase en la ontología.



Captura de pantalla de OntoGloss