

SISTEMAS BASADOS EN CONOCIMIENTOS

■ Título

Desarrollo de un buscador semántico de los proyectos existentes de la Universidad Técnica Particular de Loja.

■ Autores: nombre y apellidos completos

Hermel Santiago Aguirre Montaña

Xavier Antonio Quevedo Jumbo

■ Fecha:

19 de marzo del 2019

■ Ciclo académico:

Periodo Extraordinario 1 - 2019

- **Introducción: justificación y estado del arte del tema (es importante que se describa además del marco teórico, los trabajos relacionados)**

Justificación

La Web se ha venido desarrollado más como un medio de documentos para las personas que para los datos e información que pueden ser procesados de forma automática. Con la ayuda de la web semántica los buscadores pueden encontrar la información más relevante de un determinado tema que un usuario esté investigando, ya que dichas búsquedas se las realiza por el significado no por el contenido textual, lo que ayudará a tener información de calidad.

Es por ello que se plantea realizar un buscador semántico de los proyectos de la UTPL tomando en cuenta dos fuentes de datos en donde representaremos los mismos mediante una ontología esto con la finalidad de visualizar ante el usuario una aplicación que le permita identificar, buscar y ordenar los datos, de acuerdo a filtros, todo esto será posible con ayuda de otras tecnologías semánticas.

Estado del Arte

Concepto de web semántica [1].

Es un conjunto de actividades que están enfocadas al Internet, lo que se pretende es crear información con unos ciertos parámetros para que así sea entendible por las máquinas.

¿Para qué sirve la web semántica? [1].

Toda la información que está subida en Internet, está en una forma no estructurada, se pretende que el usuario pueda acceder a la información que en verdad le interesa.

Aspectos de la web semántica [2].

- URI: es un acrónimo de “Uniform Resource Identifier”, que permite localizar un recurso que esté en Internet.
- XML: “eXtensive Markup Language” permite la codificación de documentos complejos, formato de texto diseñado para la transmisión de datos estructurados.
- RDF: “Resource Description Framework”, es el encargado de establecer relaciones entre los datos.

- Ontologías: Recopilación de enunciados que relacionan conceptos y reglas de lógica para entender los datos.

Ventajas y Desventajas de la Web Semántica [3], [4]

Ventaja	Desventaja
Contenido semántico a la información de Internet.	Costoso y muy laborioso proceso de adaptar la información que está en Internet.
Se organiza por conceptos la información.	No toda la información se encuentra en un mismo idioma varian segun el pais.
Los buscadores encuentran la información más fácilmente.	Otro problema es la complejidad de la codificación semántica, unificar estándares semánticos y es laborioso

Proyectos UTPL

Al primer semestre del 2018, la producción científica de la UTPL superó las 1000 publicaciones indexadas en Scopus (la base de datos científica más grande del mundo), gracias a que los artículos de investigación generados responden a altos estándares de calidad y por este motivo son publicados en revistas de alto impacto (cuartil 1), como: Science, Nature Publishing Group, PNAS, Plus One, entre otras.

Un total de 17 grupos y 10 observatorios de investigación, hacen posible que la UTPL genere publicaciones de carácter científico y de divulgación social, permitiendo el desarrollo de proyectos de investigación aplicada, que luego evolucionan a procesos de innovación y emprendimiento, a través de distintas iniciativas que potencian esta actividad desde la academia.

Buscador Semántico [5].

Un buscador semántico extrae información de los datos etiquetados (cabeceras, RDF, microformatos). Si la web nuestra no contiene estas etiquetas se hace más difícil que sea entendida por los algoritmos de búsqueda.

Los algoritmos de estos buscadores generan las ontologías que definen las relaciones entre las entidades mediante el Lenguaje de Ontología Web (OWL), turtle, N3, N-Triples... Una ontología define cosas como clases, superclases, subclases, propiedades y las relaciones entre ellos. Por ejemplo, podría especificar que “si una persona trabaja para una compañía, entonces la compañía emplea a la persona” (una relación inversa).

Trabajos Relacionados [6].

1. Natural Finder

Creado por dos empresas españolas, Sitesa y Bitex. El enlace da a una demo y se puede contrastar una búsqueda con Natural Finder y otra sin ella.

2. Swotti

Creado por la empresa española, BuzzTrend.

3. Ideas Afines

Buscador de palabras y conceptos relacionados. Contiene una lista de palabras relacionadas muy amplia.

4. Hakia

Dirigido por Riza Berkan, experto en Inteligencia Artificial y Procesamiento del Lenguaje Natural. Posibilidad de realizar búsquedas en inglés.

5. Askwiki

Buscador para la Wikipedia. En inglés.

6. Powerset

Buscador para la Wikipedia. En inglés.

7. Lexxe

Promete dar respuesta con un 50% más de relevancia que cualquier buscador. (En inglés).

8. Kartoo

Devuelve los resultados de forma gráfica y en diferente tamaño según su importancia.

9. Retrievr

Original buscador de imágenes en Flickr. La búsqueda se puede hacer de dos formas: dibujando en una pantallita la forma y colores que deseas, o subiendo una fotografía que sirva de modelo.

10. Mnemomap

Recibe información de diferentes fuentes como Digg, Youtube, del.icio.es, y de imágenes, además de lo que hay en la Web.

■ **Objetivos**

Se pretende realizar un buscador semántico que permita realizar la búsqueda de los proyectos de la Universidad Técnica Particular de Loja.

El objetivo del presente software es que cuando un usuario intenta realizar una búsqueda de un proyecto de la UTPL pueda introducir en el buscador palabras claves relacionadas con el proyecto a buscar y este al momento le proporcionará los mejores resultados obtenidos.

● **Beneficiarios**

Los principales beneficiarios serán los docentes y estudiantes, ya que podrán obtener información de los diferentes proyectos, que se realizan o se realizarán en la universidad.

■ **Metodología:** describir cómo se realizará la investigación, los materiales, documentación y métodos empleados. En este punto se deberá hacer mención a un ciclo de publicación de datos enlazados. Como primer paso de la metodología identificar y seleccionar las fuentes de información con las cuales se trabajará.

■ **Potenciales resultados.**

Especificación de Fuentes de Datos

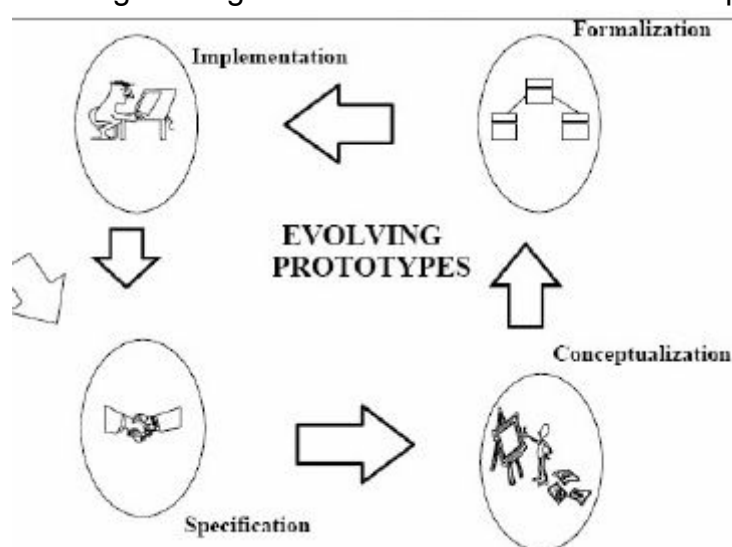
- **Completar la plantilla de especificación de datos según la plantilla que consta al final de este anuncio.**

Se incluire un archivo en formato .xlsx

Especificación, Conceptualización y Formalización del modelo ontológico

- Desarrollar las actividades de cada una de las etapas de creación de ontología según METHONTOLOGY. Documentar cada paso.

En el siguiente gráfico se detallan las actividades que contiene *Methontology*



Ontología “*Buscador Semántico Proyectos UTPL*”

ESPECIFICACIÓN

En la siguiente tabla se muestran el documento de requerimientos de la ontología que se pretende elaborar en la cual tenemos propósito, un alcance, un lenguaje de implementación, se identifican los posibles usuarios finales, los posibles usos de la ontología, los requerimientos identificados.

DOCUMENTO DE ESPECIFICACIÓN DE REQUERIMIENTOS DE LA ONTOLOGÍA	
1	PROPÓSITO
	La ontología sobre la búsqueda semántica de proyecto de la UTPL, esto permitirá que el usuario realice búsquedas de proyectos por medio de palabras las claves el mismo que nos arrojará los mejores resultados.

2	ALCANCE
	<p>Se pretende realizar un buscador semántico que permita realizar la búsqueda de proyectos dependiendo a la categorización se pueda introducir en el buscador palabras claves relacionadas con los proyectos, estas pueden ser nombres del proyecto, nombres de los investigadores entre otras y este al momento le proporcionará los mejores resultados obtenidos.</p> <p>El nivel de granularidad está directamente relacionado con las cuestiones de requisitos que se especifican en los criterios de elegibilidad.</p>
3	LENGUAJE DE IMPLEMENTACIÓN
	La ontología será implementada en el lenguaje OWL (Ontology Web Language), utilizando la herramienta de edición de ontologías Protégé.
4	POSIBLES USUARIOS FINALES
	<ol style="list-style-type: none"> 1. <i>Investigador que requiere información de los diferentes proyectos que se realizan en la UTPL.</i> 2. <i>Estudiantes de la UTPL que requieran información sobre los diferentes proyectos que se realizan en la UTPL.</i> 3. <i>Organismos de control que requieren conocer información sobre los proyectos generados en la UTPL.</i> 4. <i>Docentes que necesitan información sobre algún tema en específico de los diferentes proyectos que se generan en la UTPL.</i> 5. <i>Entidades gubernamentales que deseen conocer los proyectos que genera la UTPL.</i> 6. <i>Agentes máquina que podrían procesar la información de forma automática y con diferentes fines.</i>
5	POSIBLES USOS
	<ol style="list-style-type: none"> 1. <i>Buscar información detallada de los proyectos.</i> 2. <i>Estudiantes que puedan necesitar informacion de algun proyecto para el desarrollo de sus trabajos .</i> 3. <i>Docentes que puedan necesitar informacion de algun proyecto para el desarrollo de sus trabajos e investigaciones.</i>
6	REQUERIMIENTOS
	<p>a. Requerimientos no funcionales</p> <ul style="list-style-type: none"> - <i>La ontología se desarrollará inicialmente en el idioma español, con proyección para implementarla en inglés.</i>

	b. Requerimientos funcionales: Cuestiones de competencia 1. ¿Cuáles son los Status de los proyectos de la UTPL? 2. ¿Cuál es la fecha de inicio de los proyectos de la UTPL ? 3. ¿Cual son los tipos de proyectos existen en la UTPL? 4. ¿Tipos de miembros vinculados a los proyectos?. 5. ¿En que año han iniciado los proyectos? 6. ¿Cuál es el porcentaje de los proyectos de la UTPL? 7. ¿Cuál es el rol que cumple un miembro en un proyecto? 8. ¿Cuál es la fecha de inicio de trabajo de un miembro en un proyecto? 9. ¿Cuál es el nombre de los miembros del proyecto? 10. ¿Cuál es la identificación de los miembros del proyecto?		
7	Pre-glosario de términos		
	Términos de las cuestiones de competencia		
	Proyecto Tipo Titulo Descripción Año de inicio Fecha de inicio Fecha de fin Status Codigo del proyecto Investigador	Docente Nombre del Docente Fecha de inicio del docente Fecha fin del docente Id del docente Id del proyecto Tipo de docente	
	Objetos		
	Estudiante, docente, investigador Utpl.		

CONCEPTUALIZACIÓN

Construir glosario de términos

Definición del glosario de términos:

En la siguiente tabla se tiene el glosario de términos elaborada a partir de los datos en el cual se indica el nombre, sinónimos(si tuviese), acrónimos(si tuviese), descripción a lo que se refiere el término dentro de los datos y el tipo de tiene (concepto, atributo).

NOMBRE	SINÓNIMOS	ACRÓNIMOS	DESCRIPCIÓN	TIPO
Proyecto			Un proyecto es una planificación que consiste en un conjunto de actividades que se encuentran interrelacionadas y coordinadas.	Concepto

Status			Referencia del estado en el que se encuentra el proyecto	Atributo de Instancia
Titulo			Nombre que llevara el proyecto	Atributo de Instancia
Codigo			Código con el cual el proyecto sera identificado internamente	Atributo de Instancia
Fecha de inicio			Fecha donde inician las actividades del proyecto	Atributo de Instancia
Fecha final			Fecha donde se terminan las actividades del proyecto.	Atributo de Instancia
año			Año donde inició el proyecto	Atributo de Instancia
Descripción			Breve descripción de lo que tratara el proyecto.	Atributo de Instancia
id			Este id será el cual nos permitirá unir a un proyecto con un docente	Atributo de Instancia
Persona			Miembro que esta unido a un proyecto	Concepto
Nombre			Definira el nombre de la persona	Atributo de Instancia
Rol			Que rol cumplirá esta persona en el proyecto.	Atributo de Instancia
Image			Imagen del docente	Atributo de instancia
Email				
Tipo			Se utiliza para Proyecto, Articulos y personas el cual desempeña cosas distintas en cada una de ellas.	
ISBN			Para el registro de articulos.	
Journal			La revista en donde se publica	

Taxonomía de conceptos

Elaboración de modelo de conceptos, en el siguiente modelo se tiene los conceptos identificados en la especificación.

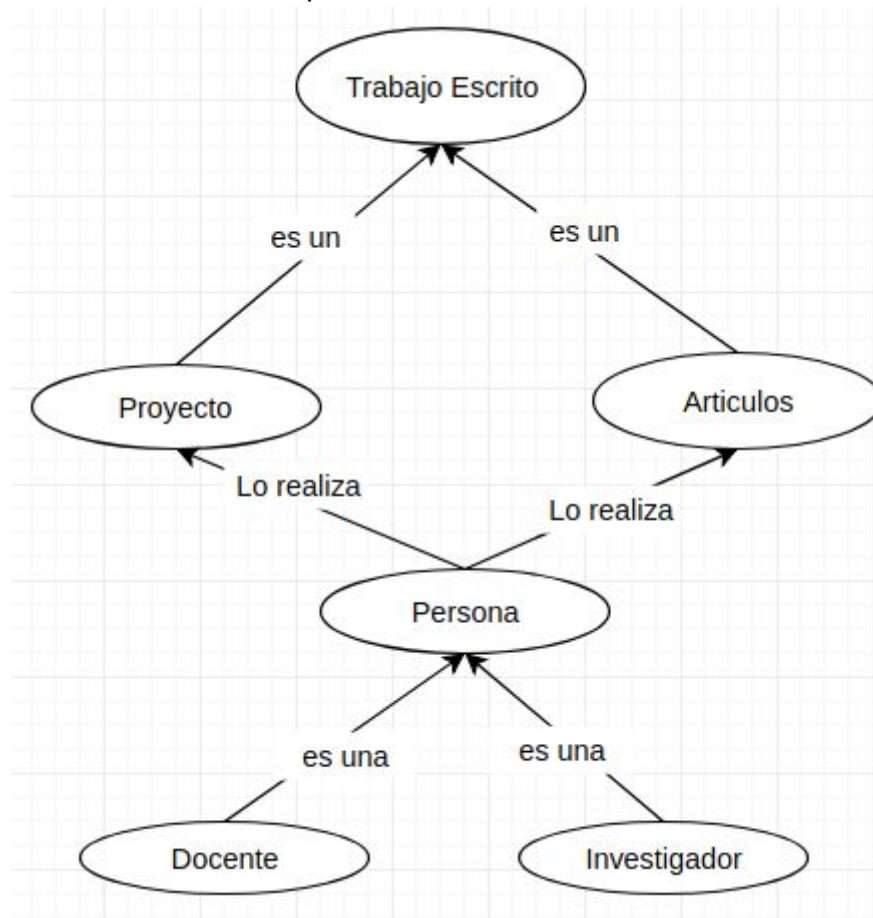
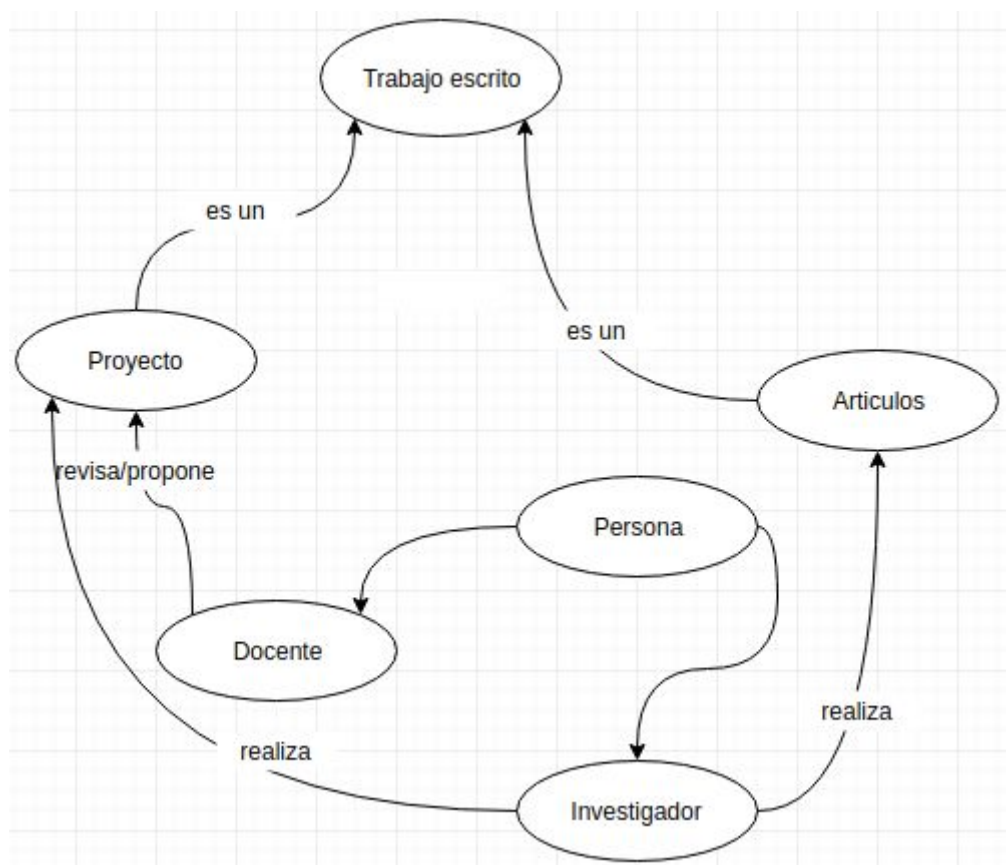


Diagrama de relaciones binarias

Elaboración de diagrama de relaciones, en el siguiente gráfico se muestra las relaciones que van a tener cada concepto.



Diccionario de conceptos

En la siguiente tabla se muestra un diccionario de conceptos se muestra la relación de cada concepto.

Nombre del concepto	Instancias	Atributos de clase	Atributos de instancia	Relaciones
Trabajo Escrito	Tesis			es un
Articulos	Cientifico Investigacion			es un
Persona			Identificacion	realiza
Docente				revisa/propone
Investigador	Docente Investigador			realiza

Tabla de atributos

En la siguiente tabla se muestra los atributos identificados

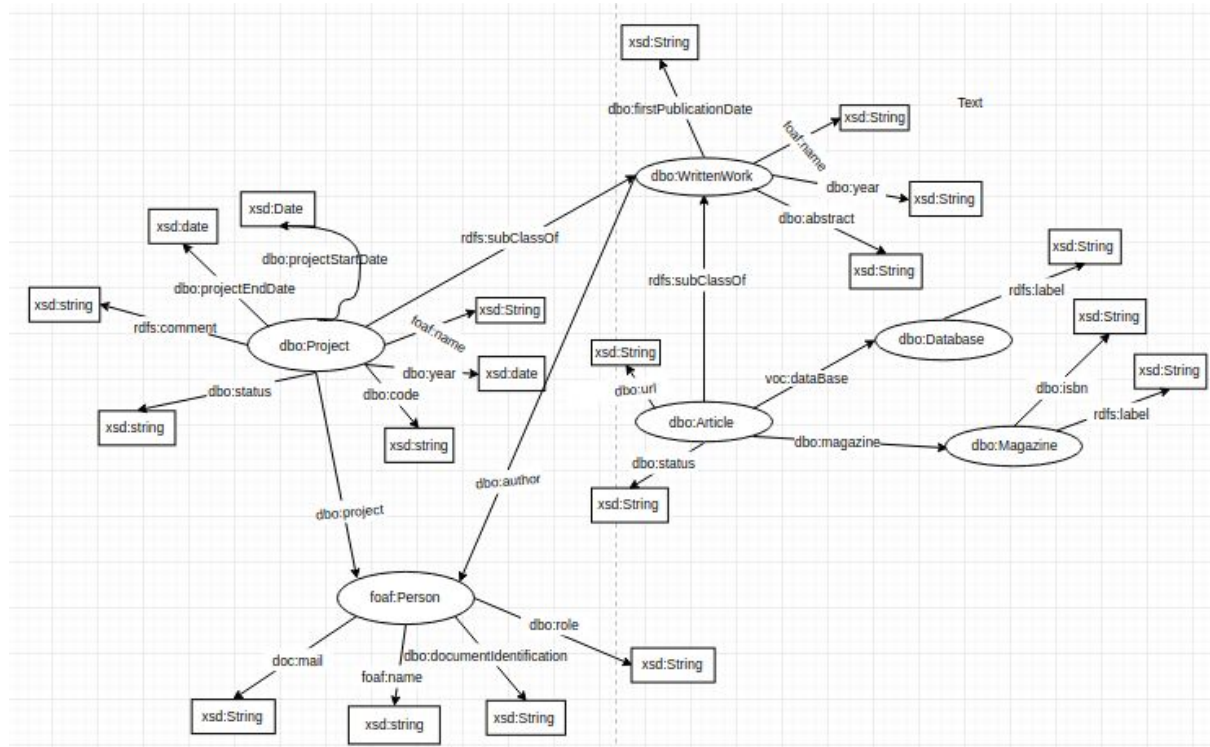
Nombre del atributo	Concepto	Tipo de valor	Rango de valores	Cardinalidad
Proyecto	TrabajoEscrito Persona	Cadena de caracteres	Nombre	(1,1)
Status	Proyecto Articulo	Cadena de caracteres	Activo Finalizado	(1,1)
Titulo	Proyecto Articulo	Cadena de caracteres	Nombre	(1,1)
Codigo	Proyecto Articulo	Cadena de caracteres	Nombre	(1,N)
Fecha de inicio	Proyecto Articulo	Fecha		(1,1)
Fecha final		Fecha		(1,1)
año		Fecha		(1,1)
Descripción		Cadena de caracteres		(1,1)
id				(1,1)
Persona				(1,N)
Nombre		Cadena de caracteres		(1,1)
rol		Cadena de caracteres		(1,1)
Identificación	Persona	Cadena de enteros		(1,1)

Tabla de relaciones binarias

Nombre de la relacion	Concepto de origen dominio	Cardinalidad maxima	Concepto de destino rango	Relacion inversa
es un	Investigador	2	Proyecto	
revisa/propone	Docente	N	Proyecto	
realiza	Investigador	N	Articulo Proyecto	

Modelo Semi-Formal de la ontología

En el siguiente modelo se muestra la ontología semi formal propuesta utilizando vocabularios como (dbpedia, foaf y propios)

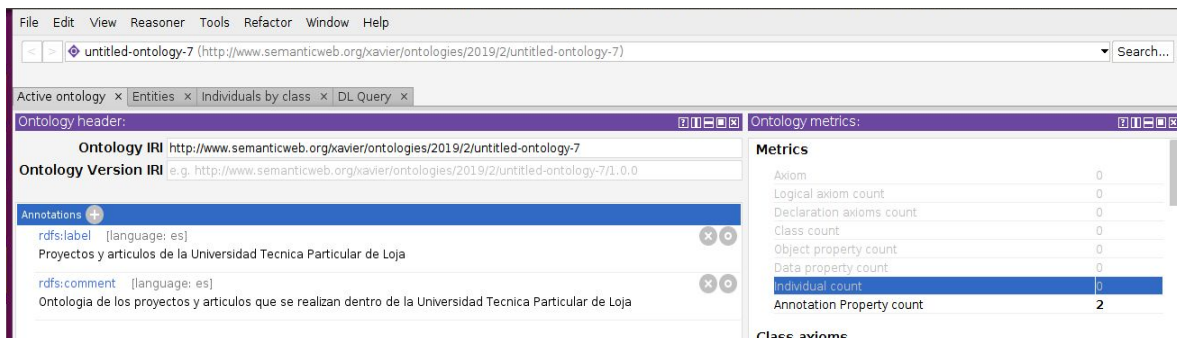


Implementación de ontología

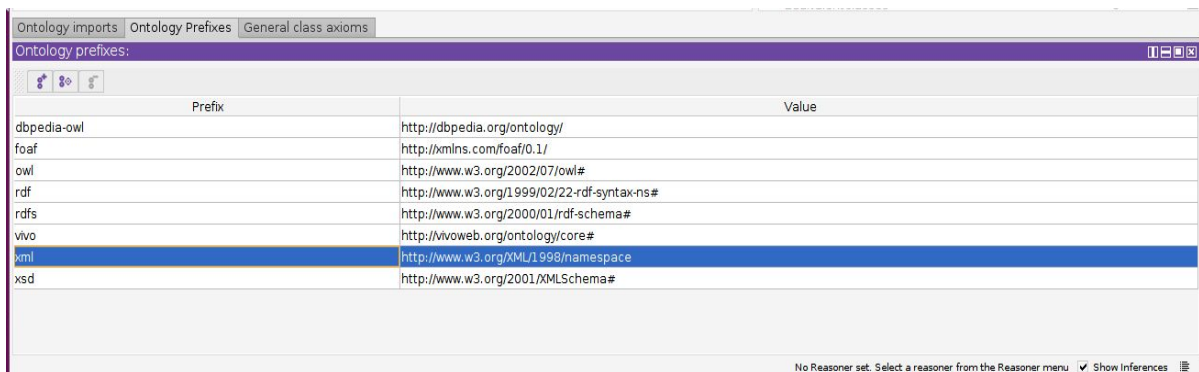
En base a las fases previas de creación de ontología, crear el modelo computable de la ontología en Protégé. Documentar el proceso y colocar las pantallas principales en el informe del proyecto.

Proceso en Protégé

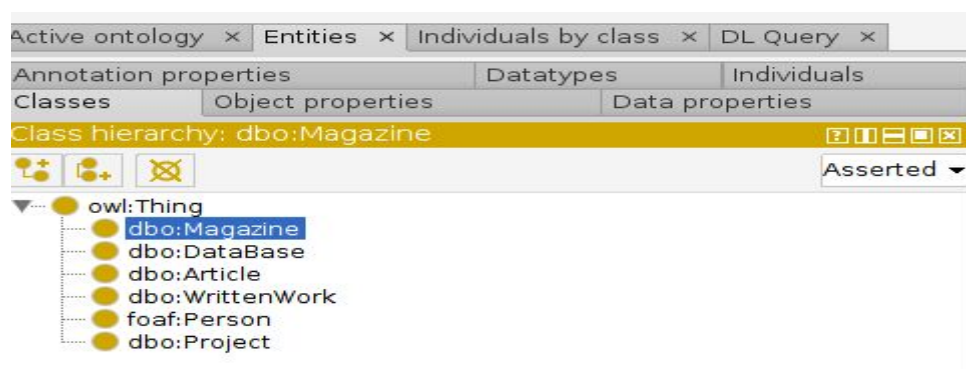
Como primer paso le asignamos una **Ontology IRI** a la ontología, en nuestro caso dejamos la que nos crea por defecto al abrir el Protégé, creamos dentro de **Annotations**, un `rdf:label` y un `rdf:comment` como una pequeña introducción de lo que se va encontrar al abrir el archivo generado, como se muestra en la captura de pantalla.



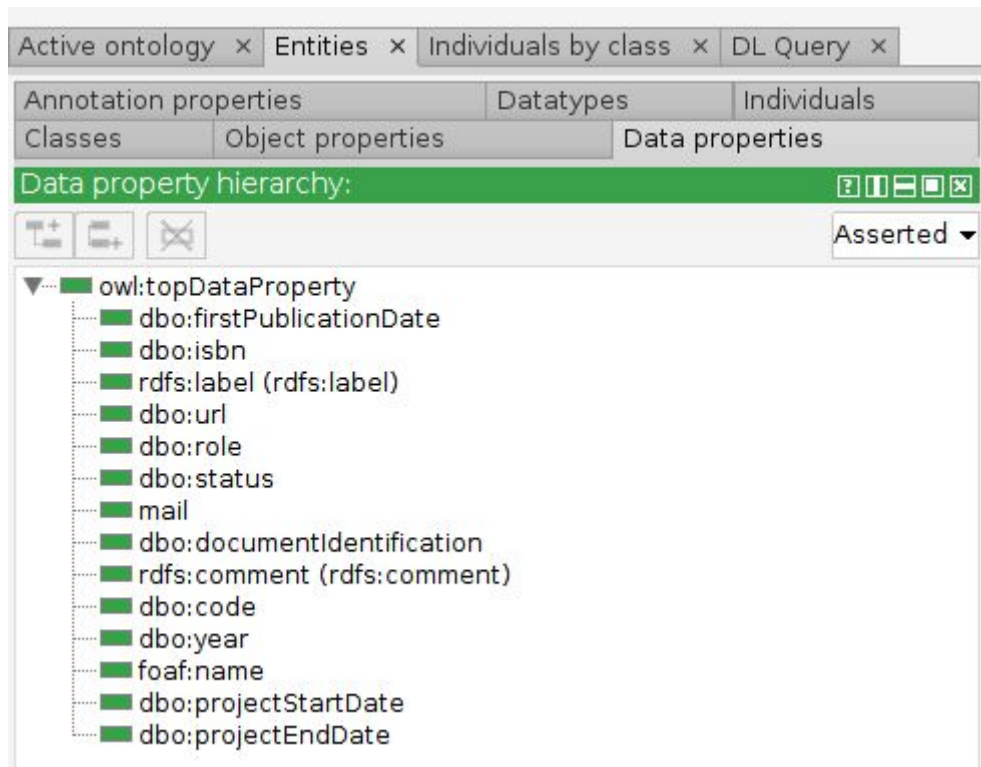
Como segundo paso añadimos los prefijos que irán en nuestra ontología en la parte de **Ontology Prefixes** algunos ya están pero hay que añadir los que vamos a utilizar como propios prefijos.



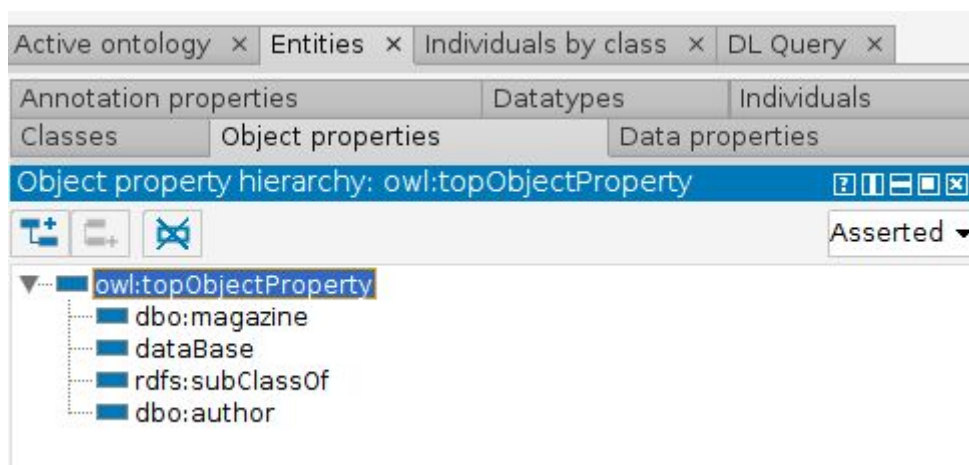
Como tercer paso comenzamos a crear las clases haciendo click en **Entities** y seguido a **Classes**, a continuación se muestra una captura de pantalla. (Se le dio un label y comment pero no se guardaba)



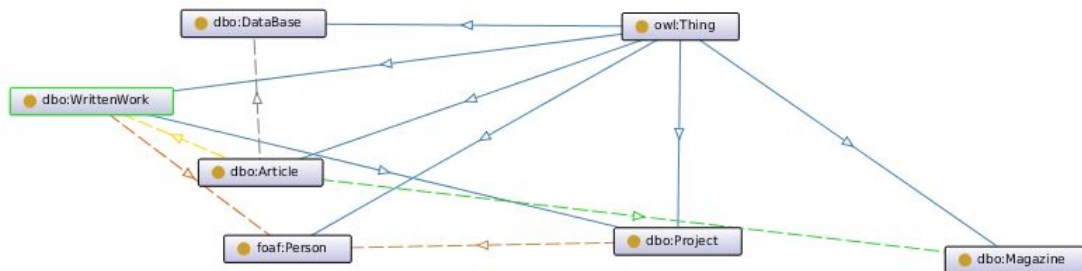
Como cuarto paso ponemos las propiedades de los datos en la parte de **Data property hierarchy**, como se muestra en la siguiente captura. Y se les pone un **Domain** y **Range**.



Como quinto paso ponemos las propiedades de los objetos lo hacemos en **Object properties**, como se adjunta la captura de pantalla. Se le pone un **Domain** y **Range**.



Como sexto paso guardamos todo y generamos un gráfico en la parte de **views** con las clases como se muestra el gráfico resultante.



Generación de datos RDF

De acuerdo al modelo ontológico y una vez que los datos estén listos, implementar en jena los métodos necesarios para leer las fuentes y el modelo ontológico y en base a esto generar las tripletas RDF. Luego subir las tripletas en un repositorio RDF el cual debió haber sido seleccionado en base a algunos criterios.

A continuación se muestran capturas de pantallas del programa utilizando **Jena**

Variables de prefijos

```

Resource WrittenWork = null;
Resource Person = null;
Resource Article = null;
Resource Magazine = null;
Resource Database = null;
Resource Project = null;

//Fijar Prefijo para URI base de dos datos a crear
String dataPrefix = "http://example.org/data/";
model.setNsPrefix("myData", dataPrefix);

//Fijar prefijos de vocabularios incorporados en Jena
String foaf = "http://xmlns.com/foaf/0.1/";
model.setNsPrefix("foaf", foaf);

String dbo = "http://dbpedia.org/ontology/";
model.setNsPrefix("dbo", dbo);
Model dboModel = ModelFactory.createDefaultModel(); // MODELO PARA LA ONTOLOGIA
dboModel.read(dbo);

String owl = "http://www.w3.org/2002/07/owl#";
model.setNsPrefix("owl", owl);

```


Variables de lectura de los datos

```
//DATA ESTADISTICAS MATRICULADOS
br1 = new BufferedReader(new FileReader("/Users/macbookpro/Desktop/sbc/data/article_data.csv"));
br2 = new BufferedReader(new FileReader("/Users/macbookpro/Desktop/sbc/data/proyectocsv.csv"));
System.out.println("llego");
String line1;
String line2;
br1.readLine();
```

Creacion de las clases de la ontologia dentro de Java

```
Person = model.createResource(dataPrefix + prsn)
    .addProperty(RDF.type, (foaf + "Person"))
    .addProperty(FOAF.name, data1[11])
    .addProperty(dboModel.getProperty(dbo + "documentIdentification"), model.createTypedLiteral(data1[10], XSSDDatatype.XSD));

WrittenWork = model.createResource(dataPrefix + ww)
    .addProperty(RDF.type, (dbo + "WrittenWork"))
    .addProperty(FOAF.name, data1[5])
    .addProperty(dboModel.getProperty(dbo + "firstPublicationDate"), model.createTypedLiteral(data1[7], XSSDDatatype.XSD))
    .addProperty(dboModel.getProperty(dbo + "year"), model.createTypedLiteral(data1[3], XSSDDatatype.XSD))
    .addProperty(dboModel.getProperty(dbo + "abstract"), model.createTypedLiteral(data1[1], XSSDDatatype.XSD));

Article = model.createResource(dataPrefix + art)
    .addProperty(RDF.type, (dbo + "Article"))
    .addProperty(dboModel.getProperty(dbo + "url"), model.createTypedLiteral(data1[9], XSSDDatatype.XSD))
    .addProperty(dboModel.getProperty(dbo + "status"), model.createTypedLiteral(data1[6], XSSDDatatype.XSD));
Magazine = model.createResource(dataPrefix + mgzn)
    .addProperty(RDF.type, (dbo + "Magazine"))
    .addProperty(dboModel.getProperty(dbo + "isbn"), model.createTypedLiteral(data1[0], XSSDDatatype.XSD))
    .addProperty(FOAF.name, data1[8]);

Database = model.createResource(dataPrefix + db)
    .addProperty(RDF.type, (dbo + "Database"))
    .addProperty(FOAF.name, data1[4]);
Project = model.createResource(dataPrefix + prjt)
    .addProperty(RDF.type, (dbo + "Project"))
    .addProperty(FOAF.name, data2[3])
    .addProperty(dboModel.getProperty(dbo + "year"), model.createTypedLiteral(data2[9], XSSDDatatype.XSD))
    .addProperty(dboModel.getProperty(dbo + "code"), model.createTypedLiteral(data2[1], XSSDDatatype.XSD))
    .addProperty(dboModel.getProperty(dbo + "status"), model.createTypedLiteral(data2[0], XSSDDatatype.XSD))
    .addProperty(dboModel.getProperty(dbo + "comment"), model.createTypedLiteral(data2[2], XSSDDatatype.XSD))
    .addProperty(dboModel.getProperty(dbo + "projectEndDate"), model.createTypedLiteral(data2[10], XSSDDatatype.XSD))
    .addProperty(dboModel.getProperty(dbo + "projectStartDate"), model.createTypedLiteral(data2[4], XSSDDatatype.XSD));
```

Creacion de las relaciones entre clases de la ontologia

```
model.add(WrittenWork, myModel.getProperty(dbo + "author"), Person);
model.add(Project, myModel.getProperty(dbo + "Project"), Person);
model.add(Article, RDFS.subClassOf, WrittenWork);
model.add(Project, RDFS.subClassOf, WrittenWork);
model.add(Article, myModel.getProperty(dbo + "magazine"), Magazine);
model.add(Article, myModel.getProperty(dbo + "dataBase"), Database);
```

Conexión con virtuoso

```
String service = "http://localhost:8890/sparql";
String defaultGraph = "http://localhost:8890/sbc_proyecto";
```

Consulta de sparsql dentro java

```
public void verconsulta() {  
  
    String prueba = "";  
    prueba = String.valueOf(texto.getText());  
  
    areal.setText(" ");  
    areal.append("RESULTADO DE BUSQUEDA GENERAL\n");  
  
    String queryString = "PREFIX dbo: <http://dbpedia.org/ontology/>"  
        + " PREFIX foaf: <http://xmlns.com/foaf/0.1/>"  
        + "SELECT DISTINCT ?o "  
        + "WHERE {"  
        + "?s foaf:name ?o "  
        + "FILTER regex(?o, '\" + prueba + \"', 'i'))";  
    System.out.println("primera-----" + queryString);  
}
```

Pruebas del programa



Sistema de Información Académica Científica de UTPL

Buscar por Palabra Clave ▼ Buscar

RESULTADO DE BUSQUEDA GENERAL

"Informativo estudiantil "La Nota UTPL"Black/Medianet"

Implementacion de tecnicas de control biologica en cultivos de ciclo corto en la finca UTPL

Fototeca digital colaborativa del patrimonio fotografico de la UTPL

Bibliografía

- [1]B. Francisco Hamidian Fernández and G. Ramón Ospino Sumoza, “¿Por qué los sistemas de información son esenciales?,” 2015.
- [2]J. Hendler, “Agents and the Semantic Web,” *IEEE Intell. Syst.*, vol. 16, no. 2, pp. 30–37, Mar. 2001.
- [3]Emelys Madrid, “Web: Ventajas y Desventajas de la Web Semántica,” 2010. [Online]. Available: <http://websemanticae.blogspot.com/2010/11/ventajas-y-desventajas-de-la-web.html>. [Accessed: 28-Feb-2019].
- [4]Universidad Estatal de Bolívar, “Estudio de la Web,” [Online]. Available: <http://www.ueb.edu.ec/sitio/images/PDF/DIC/NIVELII/TEORICO/1-Evoluci%C3%B3n-Web.pdf>
- [5]Natzir Turrado, “Qué es un buscador semántico,” 2011. [Online]. Available: <https://www.analistaseo.es/posicionamiento-buscadores/buscador-semantico/>. [Accessed: 28-Feb-2019].
- [6]“Ejemplos de buscadores semánticos - Buscadores Semánticos,” 2011. [Online]. Available: <https://sites.google.com/site/buscadoressemanticos/ejemplos-de-buscadores-semanticos>. [Accessed: 28-Feb-2019].