



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
La Universidad Católica de Loja

ÁREA TÉCNICA

**TITULACIÓN DE INGENIERO EN SISTEMAS INFORMÁTICOS Y
COMPUTACIÓN**

**Aplicación de técnicas de Linked Data para la publicación de datos
enlazados de Open Educational Resources encontrados en
OpenCourseWare**

TRABAJO DE FIN DE TITULACIÓN

AUTOR:

Novillo Navarro, Andrea Alejandra

DIRECTOR:

Piedra Pullaguari, Nelson Oswaldo, Ing.

LOJA – ECUADOR

2014

APROBACIÓN DE LA DIRECTORA DEL TRABAJO DE FIN DE TITULACIÓN

Ingeniero.

Nelson Oswaldo Piedra Pullaguari.

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de fin de titulación: Aplicación de técnicas de Linked Data para la publicación de datos enlazados de Open Educational Resources encontrados en OpenCourseWare realizado por Novillo Navarro Andrea Alejandra, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, Diciembre de 2014

f)

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo Novillo Navarro Andrea Alejandra declaro ser autor (a) del presente trabajo de fin de titulación Aplicación de técnicas de Linked Data para la publicación de datos enlazados de Open Educational Resources encontrados en OpenCourseWare, de la Titulación de Ingeniero en Sistemas Informáticos y Computación, siendo Ing. Nelson Oswaldo Piedra Pullaguari director (a) del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 67 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través, o con el apoyo financiero, académico o institucional (operativo) de la Universidad”

f.

Autor: Andrea Alejandra Novillo Navarro

Cédula: 1103936298

DEDICATORIA

El presente trabajo de tesis es una muestra de que con esfuerzo y ganas de seguir adelante se pueden cumplir las metas, que cada caída nos ayuda a fortalecer y saber que lo importante es llegar a la meta propuesta, no importan los obstáculos encontrados en el camino a ella, si no la satisfacción de decir “Lo logre y voy a continuar”.

Dedico este trabajo primeramente a Dios por brindarme cada día de vida para lograr esta meta. A mi pilar fundamental mi familia, mi Madre Violeta que ha sido mi punto de referencia para lograr saltar cada obstáculo, por alimentar mis ganas de seguir con su apoyo incondicional, sus palabras de cariño y aliento Gracias por ser mi “Mejor Amiga”. A mi padre Manuel por ser aquel con su empeño me ha demostrado que la vida está llena de metas y que los principios inculcados en casa son fundamentales durante consecución de estas. A mis hermanos por ser el granito de alegría que tiene mi vida con sus ocurrencias y su forma de ver la vida.

A mis amigos por ser mi apoyo en las buenas y malas, por aquellas palabras de aliento y sobre todo aquellos alones de oreja y por haber compartido aquellas malas noches de estudio risas y ocurrencias.

A las familias que me acogieron en esta ciudad como miembro de ellos y por brindarme su cariño, sus consejos pero sobre todo su apoyo.

Andrea A. Novillo Navarro

AGRADECIMIENTOS

Agradezco a aquellos profesores que han sido parte de mi formación profesional a través de su enseñanza. En especial al Ing. Nelson Piedra por ser mi tutor durante esta etapa culminante de mi vida universitaria, pero sobre todo un gran amigo.

A cada uno de los integrantes del departamento de tecnologías de la Web Avanzada e Inteligencia Artificial por brindarme sus conocimientos y ayuda incondicional.

A mis compañeros tesistas por esa chispa de alegría durante las arduas jornadas de trabajo y brindarme su apoyo.

A todos aquellos que de alguna u otra manera me han brindado palabras de aliento, consejos y fortaleza en esta etapa de mi vida profesional.

INDICE DE CONTENIDOS

APROBACIÓN DE LA DIRECTORA DEL TRABAJO DE FIN DE TITULACIÓN.....	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA.....	iv
AGRADECIMIENTOS	v
INDICE DE CONTENIDOS	vi
RESUMEN.....	1
ABSTRACT	2
INTRODUCCION.....	3
OBJETIVOS	5
OBJETIVOS ESPECIFICOS	5
JUSTIFICACIÓN	6
1. ESTADO DEL ARTE.....	7
1.1. Web.....	8
1.1.1. Evolución actual de la Web ó Web semántica.	8
1.1.2. Web de datos ó Web semántica.	11
1.2. Linked data.....	11
1.2.1. Principios de Linked Data.	12
1.2.2. Beneficios del Linked Data.	13
1.2.3. Tecnologías Linked Data.	13
1.2.4. Proceso de publicación de datos.	20
1.3. Descripción del Ambiente de datos OCW y OER.....	23
1.3.1. Descripción OCW.	23
1.3.2. OCW y OER en Iberoamérica.....	25
1.3.3. Contenido de OCW.....	27
1.3.4. Estructura de los sitios.....	28
1.3.5. Plataformas usadas en OCW	28
1.3.6. Licenciamiento de los OCW.....	31

1.3.7.	Calidad de contenidos	34
1.3.8.	Descripción de OER	36
1.4.	Repositorios OCW.....	37
1.5.	Análisis.....	39
2.	ANÁLISIS DE FUENTES DE DATOS Y VOCABULARIOS	41
2.1.	Introducción.....	42
2.2.	Análisis de datos OCW y OER almacenados en HTML.....	42
2.2.1.	Identificación de fuentes de datos.....	42
2.2.2.	Análisis técnico y descripción del dominio: OERs dentro de los OCW.....	43
2.2.3.	Descripción del Dominio OCW/OER.....	48
2.2.4.	Enriquecimiento de información.....	50
2.3.	Especificación de entorno.....	51
2.4.	Segmentación y categorización de Información.....	52
2.5.	Análisis de Vocabularios.....	53
2.6.	Planteamiento de clases y propiedades a reutilizar	56
2.7.	Conclusiones.....	57
3.	DESARROLLO DE FRAMEWORK PARA LA EXTRACCIÓN DE DATOS DESDE HTML	59
3.1.	Extracción de Información.....	60
3.1.1.	Técnicas de Web scraping.....	60
3.2.	Expresiones regulares o patrones de extracción.....	62
3.3.	Proceso scraping para extracción de información.....	68
3.3.1.	Generación de Código Scrapy.....	70
3.3.2.	Limpieza de Datos.....	72
3.4.	Extracción de metadatos de OER.....	73
3.4.1.	Creación de Crawler para extracción de Metadatos.....	73
3.5.	Conclusiones:.....	75
4.	ESTANDARIZACIÓN DE DATOS PARA PROCESO DE PUBLICACIÓN EN EL AMBIENTE OCW/OER.....	77

4.1.	Definición de Vocabulario.....	78
4.2.	Creación de Ontología en Jena.....	82
4.2.1.	Desarrollo de Schema.....	82
4.2.2.	Verificación Schema OcwOer.....	83
4.3.	Conversión de Datos a RDF.....	84
4.3.1.	Carga de Datos.....	84
4.4.	Publicación de Datos.....	87
4.4.1.	Consultas SPARQL.....	87
4.5.	Conclusiones.....	95
	DISCUSIÓN.....	97
	CONCLUSIONES.....	101
	RECOMENDACIONES.....	103
	BIBLIOGRAFÍA.....	104
	ANEXOS.....	107
	Anexo 1.....	108
	Vocabulario de Clases.....	108
	Vocabulario de Propiedades.....	111
	Anexo 2.....	120
	Creación de ontología OCW/OER.....	120
	Anexo 3.....	121
	Schema vocabulario OcwOer.....	121
	Anexo 4.....	129
	Denominaciones de menús de cursos.....	129
	Anexo 5.....	134
	Manual de Instalación de Scrapy en Windows.....	134
	Anexo 6.....	146
	Datos Obtenidos de sitios OCW.....	146
	Anexo 7.....	149

Plataformas empleadas en sitios OCW..... 149

RESUMEN

El presente proyecto de investigación se desarrolla con la finalidad de publicar datos enlazados en el ámbito Recursos Educativos abiertos que se encuentra en los diferentes sitios Opencourseware de algunas Universidades nivel mundial; para cumplir con el objetivo del proyecto se utilizarán los principios de Linked Data definidos por Tim Berners-Lee.

Para la obtención de información de los distintos Oers se construyó un framework de extracción de información de sitios web mediante la utilización de Scrapy y por medio de la creación de un crawler de extracción de metadatos, se enriquece la información de los OERs obtenidos. Además, se construyó un vocabulario denominado OcwOer el cual representa el ámbito del proyecto, reutilizándose vocabularios como DUBLIN CORE, FOAF, GEO, VCARD; asimismo se implementó un programa diseñado en JAVA que utiliza la librería JENA, para la conversión de información a formato RDF que permite contar con datos estructurados y entendidos por máquinas. Finalmente, se realiza la publicación de datos en Virtuoso mediante el endpoint SPARQL que este posee, realizándose consultas con el fin de resolver las interrogantes planteadas inicialmente.

Palabras claves: Linked Data, OCW, Recursos educativos abiertos, Web Semántica, Jena, Scrapy, RDF.

ABSTRACT

This research project is developed in order to publish linked data in the field open educational resources located in different sites Opencourseware some universities worldwide; to meet the project objective Linked Data principles defined by Tim Berners-Lee will be used.

To obtain information from various Oers a framework for extracting information from websites using Scrapy and through the creation of a crawler metadata extraction was built, the information obtained is enriched OERs. In addition, a vocabulary called OcwOer which represents the scope of the project was built, reused vocabularies like Dublin Core, FOAF, GEO, VCARD; also a program designed using the JENA JAVA library, for converting information to RDF format allows for structured data and understood by machines was implemented. Finally, the publication of data is done by Virtuoso SPARQL endpoint that this pose, carrying out consultations in order to resolve the questions raised initially.

Keywords: Linked Data, OCW, Open Educational Resources, Semantic Web, Jena, Scrapy, RDF.

INTRODUCCION

La información que se encuentra en algunos sitios web suele ser parcial o inexacta, por lo que es necesario alimentar nueva información a las bases de datos que se mantienen en la Web actualizando dicha información, la Web semántica toma importancia pues busca vincular la información que se encuentra en la Web, tomando en consideración formatos entendibles no solo por humanos, sino además por máquinas permitiendo de esta manera una vinculación o enlazado de datos.

Hablar de enlazado de datos es tomar como punto de partida la idea planteada por Tim Berners-Lee sobre su concepto en la Web, en donde su deseo era la interconexión e inferencia semántica de información, la cual debería ser prácticamente automática para ser procesable, entendible y reutilizable por máquinas y permitiendo que sean consultados desde diferentes fuentes (Berners-Lee,T, 2001).

Linked Data surge de la necesidad de vincular datos, siendo el motor de la Web semántica. El enlazado de datos se basa en que la información existente se pueda mostrar, vincular y conectar en formatos RDF por medio de URI`s únicas en la Web.

En el presente proyecto busca generar información en función del principio de Open data, tomando en consideración como principal fuente los sitios OpenCourseWare (OCW) y su contenido. Para ello se ha generado un análisis de los datos existentes en dichos sitios, sus cursos y contenidos, conociendo de esta manera cómo obtener datos que sean actualizados y generen una ayuda inmediata de vinculación de información.

El paso principal para la vinculación datos es obtener la información de la Web por ello se genera como paso inicial el análisis de los datos y la manera de extraerlos directamente desde sus principales fuentes por medio de técnicas de extracción de datos, formando así una base de datos generalizada y actualizada según la estructura que manejan los OCW. Siguiendo con el enlazado de la información a la nube de datos bajo los principios básicos de Linked Data, planteados por el mismo concepto de Berners-Lee.

La publicación de estos datos nos ayudará a conseguir algunos objetivos básicos enfocados en los Principios de Linked Data como son:

- Conocer el estado actual de las universidades que mantienen sitios OCW a nivel mundial, además de su estructura técnica.

- Desarrollar un vocabulario en el ámbito de los Recursos educativos abiertos (OERs) encontrados dentro de los diferentes sitios OCW.
- Aplicar los principios de Linked Data en la publicación de OERs encontrados dentro de los OCW.

El presente trabajo consta de 4 fases; la **primera fase** es una introducción general al ambiente de trabajo denominado Estado de arte, brindando un enfoque general de lo que es la publicación y enlazado de datos; partiendo desde los inicios de la Web, su evolución semántica hasta la aplicación de las técnicas Linked Data, y finalmente una introducción a los OERs conociendo de su estado actual.

La **segunda fase** trata sobre la Preparación y Obtención de datos sobre OCW/OER a partir de contenidos almacenados en HTML, tomando en consideración la parte técnica y la estructura en que se encuentran almacenados dichos datos; básicamente este capítulo es el inicio de la aplicación de las técnicas o principios de Linked Data pues se reconocen las fuentes de datos a tratar y plantean interrogantes que se pretenden responder con la publicación de Datos.

La **tercera fase** es la creación de un framework para extracción de datos desde HTML, mediante la implementación de técnicas open source cómo es Web scraping.

La **fase final** trata el enfoque de Estandarización de datos en formatos de publicación, en el cual se plantea la creación de una ontología OER y se define un vocabulario; previamente se analizan posibles clases y propiedades a reutilizar. Se empleará una librería de Java para la generación de datos en formatos RDF cómo es Jena. Finalmente, se publicarán los datos en un Servidor Virtuoso Open Source y se contestarán aquellas interrogantes planteadas en la segunda fase.

Las conclusiones y recomendaciones del proyecto se realizarán como acotamiento a este proyecto, generando recomendación de proyectos futuros que aportarán importancia al ambiente OCW.

OBJETIVOS

El objetivo general del proyecto es aplicación de técnicas de Linked Data para la publicación de datos enlazados de Open Educational Resources encontrados en OpenCourseWare, generando un almacén de datos mediante la recopilación y extracción de información de los diferentes sitios OCW y los respectivos OERs; logrando accesibilidad e interoperabilidad mediante transformación de información en formato RDF y el consumo de estos a través de consultas SPARQL siendo eminente enlace a la nube de datos abiertos.

OBJETIVOS ESPECIFICOS

- Extraer OERs de los diferentes sitios OCW que se encuentran en formatos HTML; generando un almacén de datos unificado aplicando técnicas de extracción, tomando en consideración la estructura de sus cursos, los recursos almacenados y los metadatos de estos últimos.
- Aplicar técnicas Linked Data en la publicación de datos enlazados de sitios OCW y sus respectivos OERs
- Desarrollar un vocabulario de datos para el ámbito de los sitios OCW y sus contenidos OER.
- Generar datos acerca de OERs en formato RDF.
- Validar la propuesta del proyecto respondiendo las diferentes preguntas planteadas, a partir de consultas SPARQL de la información extraída.

JUSTIFICACIÓN

El obtener y almacenar de información acerca de los Recursos educativos (OERs) abiertos se vuelve un verdadero problema, por la heterogeneidad y diversidad que mantienen los sitios web empleados por los OCW; generándose dicho problema además por el uso de distintas plataformas, estructuras, restricciones de acceso, lenguaje, y formato usados. De esta manera se produce la falta de interoperabilidad, localización, consulta y el reúso de los OERs.

El presente trabajo nace como apoyo a la iniciativa propuesta por Instituciones de educación superior a nivel internacional denominada "Promotion of Open Educational Resources", cuyos resultados hasta inicios del 2012 se presentan en (Tovar, Piedra & otros, 2012). Una de las líneas de trabajo de esta iniciativa se enfoca en mejorar la interoperabilidad y reusabilidad, localización y consulta de Recursos educativos abiertos y de forma concreta cursos OCW. Para conseguir este objetivo, el primer paso se basa en extraer los datos asociados a cada curso y que están disponibles en los sitios del OpenCourseWare Consortium y Universia, dos de las instituciones que en el mundo abanderan la promoción de iniciativas OpenCourseWare.

Conociendo los puntos claves de la publicación mencionada, podemos denotar la importancia de mantener información actualizada y sin restricción de acceso en un mismo lugar, planeando de alguna manera obtener y almacenar la información que hoy en día se encuentra sobre los OCW y principalmente sobre su contenido OER; para esto se plantea la aplicación de técnicas de extracción de datos, además de la implementación de Linked Data para la publicación de datos enlazados en el ambiente educativo como son los OCW y sus respectivos OERs.

1. ESTADO DEL ARTE

1.1. Web.

Hablar de la Web es retornar a sus inicios, cuando esta fue tan solo una interacción en formato sin flexibilidad entre usuario y cliente, sino de una observación pasiva sin posibilidad de editar dicha información.

Hoy en día se puede decir que la evolución y crecimiento de la Web, genera la necesidad de querer compartir información e interoperabilidad entre el usuario y las comunidades virtuales de una manera más dinámica por lo que nace la extensión de Web2.0 como primera instancia, como ejemplo de extensión se encuentran están las redes sociales, servicios Web, servicios de alojamiento, etc. Las mejoras paulatinas de la Web han generado un movimiento social tan amplio entre múltiples aplicaciones, el uso de inteligencia artificial y la denominada Web semántica buscando plasmar la idea de Tim-Berners-Lee como es la comunicación de máquina a máquina con poca intervención humana se le denomina actualmente Web 3.0.

La Web fue creada como un sistema de información global en el cual se pudiera obtener información sin distinción de situación geográfica y sobre todo de plataformas y lenguajes de programación en pocas palabras está basado en tres principios básicos : todos pueden publicar, todos pueden leer, nadie puede restringir (Gutierrez,2008). Paralelamente han ido creciendo aquellas tecnologías que han permitido que su contenido deje de ser estático como en sus inicios entre ellas están HTML, HTTP, XML, JAVA, JavaScript, etc y la creciente avalancha de la búsqueda aparecen aquellas tecnologías que ayudaran a mantener la semántica dentro de la Web y automatizar la información que esta comparte como son OWL, SPARQL, RDF, etc.

1.1.1. Evolución actual de la Web ó Web semántica.

En la Web mantenemos datos que crecen de una manera potencial, siendo el principal problema el saber que significan estos datos o como interpretarlos, pues buscadores aun no tienen la capacidad de interpretar la información de una manera semántica sino sintáctica.

Como cita (Berners-Lee,T, 2001) “La idea es tener datos en la red definidos y unidos de manera que puedan ser usados por computadoras, no solo con el propósito de visualización, sino para la automatización, integración y uso repetido de los datos en múltiples aplicaciones.”

Hablar de la evolución de la Web fue la introducción a lo que se detalla actualmente como Web semántica, que no es un cambio de lo que se conoció en sus inicios, sino más bien es el avance que en principios definía Tim Berner-Lee (Berners-Lee,T, 2001) ;esta extensión de la Web involucra lo que es la semántica y la inteligencia artificial considerando metadatos y vocabularios comunes en los que vincula aplicaciones y herramientas como XML,RDF'S, ontologías, pudiendo de esta manera tener un mínimo esfuerzo y tiempo requerido para la búsqueda de información con nivel. Al hablar de esto nos referimos a buscar información contundente, por ejemplo: *Librerías que vendan libros de Linked data en Loja*; al poner esta frase en algún buscador sintáctico tan solo nos arrojará nombres de librerías y algunos títulos de libros pero no la información específica que solicitamos. Pues bien aquí aplica el tener una inmensa base de datos bien estructurada, obteniendo de una manera interpretativa y semántica la información enlazada en la Web.

La infraestructura, tecnologías y lenguajes de programación que necesita la Web semántica para su desarrollo según Tim Berners-Lee se aclara en la siguiente imagen:

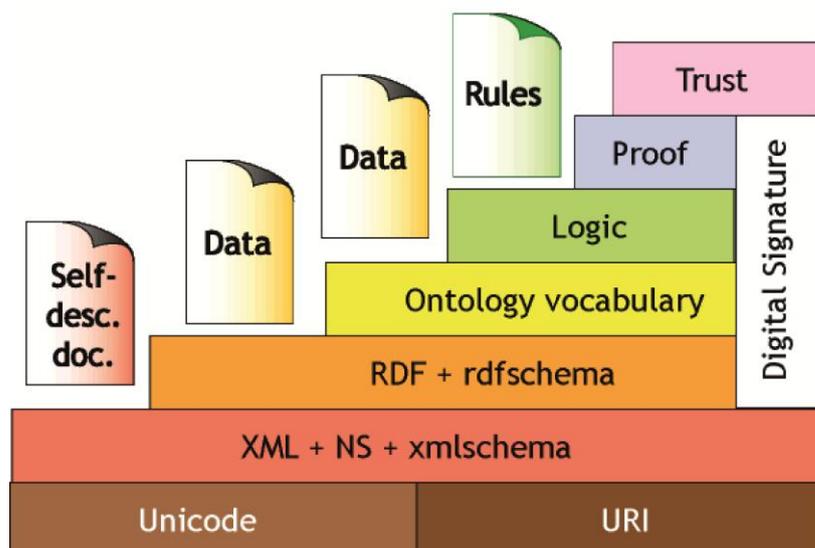


Figura 1. Arquitectura Web Semántica
Fuente: <http://www.w3.org/2001/09/21-orf/hagino-sw/slide8-0.html>

A continuación se detalla cada uno de los componentes que conforman la torre semántica según las definiciones de Tim Berners-Lee:

Unicode:

Es un estándar cuyo objetivo es proporcionar el medio por el cual un texto en cualquier forma e idioma pueda ser codificado sin aparición de símbolos extraños.

URI:

Son cadenas que permiten acceder a cualquier recurso de la Web, siendo estos identificadores únicos.

Según Berners-Lee (Berners-Lee, T, 2005) un "Uniform Resource Identifier" (URI) es una secuencia compacta de caracteres que identifica un recurso abstracto o físico, teniendo una sintaxis y un proceso para resolver estos identificadores.

Se debe considerar la diferencia del término URI que se mencionó anteriormente y el término "Uniform Resource Locator" (URL); este último encapsula un subconjunto de URIs, además de la identificación es un medio de localización y acceso de los recursos de la Web, mediante la descripción de su mecanismo de acceso primario. Tomando en consideración además que no siempre un URI es un URL.

XML+NS+xmlschema:

Capa técnica que agrupa diferentes tecnologías, que tienen un formato común para intercambiar atributos por medios nombres propios como especificación y referenciándolos en la Web.

RDF+RDFschema:

Está basada en la capa anterior, genera el lenguaje universal definiendo sintaxis y modelos de datos para la representación semántica de estos en la Web, por medio de la descripción de un vocabulario. Mantiene una sintaxis basada en XML.

Ontology (Ontologías):

Nos permite clasificar y modelar un entorno de información. Esta capa permite extender la funcionalidad de la Web Semántica agregando nuevas clases y propiedades para describir los recursos.

Logic (Lógica):

En esta capa se establecen reglas de inferencia dentro de las ontologías, pues permiten determinar si un razonamiento se sigue de sus premisas.

Proof (Pruebas):

Se realizan demostraciones matemáticas, mediante “pruebas” escritas en un lenguaje unificador de la Web semántica.

Trust (Confianza):

Basada en la confiabilidad de la información, comprobado de forma exhaustiva las fuentes de datos, generando fiabilidad de los recursos para los usuarios; se ejecuta conjuntamente con la capa de pruebas.

Digital Signature (Firma digital):

Utilizada por los ordenadores y agentes para verificar que la información ha sido ofrecida por una fuente de confianza.

1.1.2. Web de datos ó Web semántica.

También se le conoce como Web 3.0, en la que se describe la evolución y la interacción entre las personas y el internet mediante diferentes formas, generando así la transformación de la red en una gran base de datos de información, Bajo el enfoque de (W3C, 2010) la Web de datos es la vinculación de información de los datos en la Web pero de una manera más semántica, a diferencia de la Web del hipertexto. Es decir la Web de hipertexto aún está en auge, relaciona documentos escritos en formatos HTML; en cambio la Web de datos enlaza información descrita en RDF y es exactamente el último punto donde aplican los principios de Linked data.

La Web de los datos también se puede rastrear siguiendo los enlaces RDF, así como el documento de Web tradicional se rastrea siguiendo los enlaces de hipertexto. Basándose en los datos rastreados, motores de búsqueda puede proporcionar consultas sofisticadas, similar a las capacidades de consulta proporcionados por bases de datos relacionales hoy. (W3C, 2010)

1.2. Linked data.

Según el enfoque general de la Web semántica dado por TimBerness-Lee no solo es necesario tener información en la Web, sino enlazar estos datos siendo refinada la información que los usuarios como personas y máquinas puedan obtener.

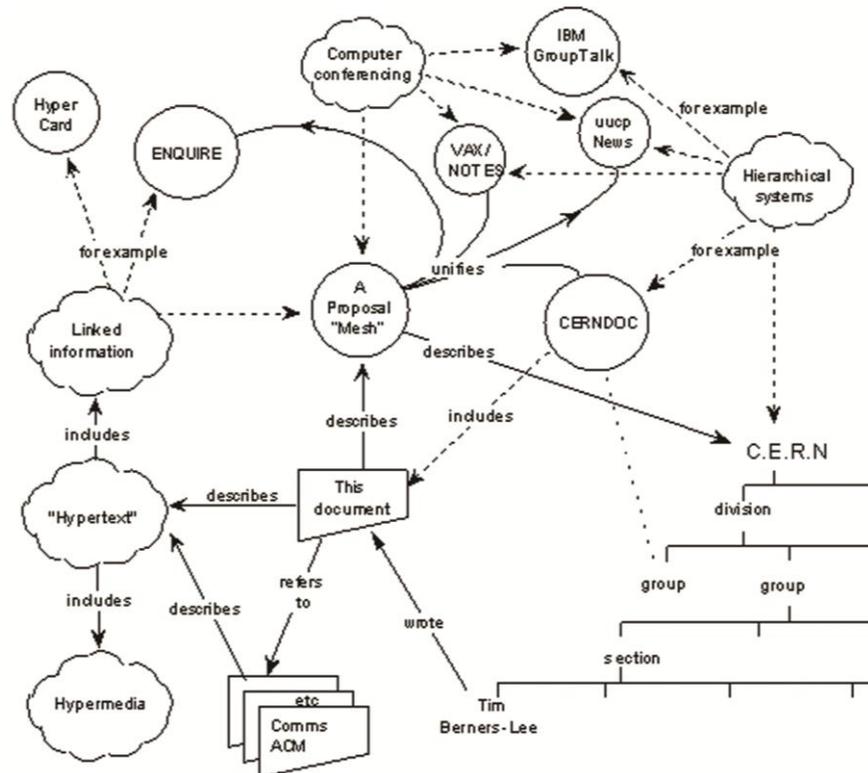


Figura 2. Modelo de gestión de información en el CERN

Fuente: Tim Berners-Lee, CERN March 1989, May 1990. Tomada de: <http://www.w3c.es/Presentaciones/2009/1111-introduccionUM-MA/>

1.2.1. Principios de Linked Data.

Tim Berners-Lee menciona (Berners-Lee, T, 2006) en una de sus presentaciones para la W3C¹ cuatro principios que caracterizan los datos vinculados detallados a continuación:

1. Utilizar URIs: Porque serán los identificadores únicos de los recursos publicados, identificándose así en la Web.
2. Usar URIs HTTP: No sería una publicación si no hay un medio por el cual mostrar dichos recursos; de esta manera asegurar que cualquier recurso pueda ser buscado y accedido en la Web.
3. Información útil de los URIs: Pudiendo ser empleado de esta manera propiedades y las clases que se encuentra en los datos y obtener información de descripciones estándares como RDF, RDFS y ontologías OWL, incluyendo las relaciones entre los términos de la ontología del recurso.

¹ <http://www.w3.org/>

4. Enlazar los URIs a Otros URIs: conformando de esta manera una red más amplia el conocimiento con fuentes externas.

1.2.2. Beneficios del Linked Data.

Tomando como referencia lo que menciona (W3C) podemos citar algunos de los beneficios de los datos enlazados.

- La Web de datos puede ser rastreada por los contenidos RFD.
- Contiene mecanismos de acceso único y estandarizado.
- Facilita el trabajo de los motores de búsqueda.
- Vinculación de los datos de diferentes fuentes.

1.2.3. Tecnologías Linked Data.

1.2.3.1. Representación o modelado de información.

La representación de datos se realiza mediante la transformación de los datos en formato RDF, para ello se deberá previamente analizar ontologías, taxonomías y vocabularios para identificar el ámbito de trabajo.

1.2.3.1.1. Ontologías y Vocabularios

Existen varias formas de representar ó modelar la información de un ambiente y cuando se habla de Web semántica, las ontologías son el mejor medio para expresar formalmente el significado de las cosas.

Existen un sin número de definiciones de Ontologías, según (Gruber ,1993) su definición abstracta es la “especificación explícita de una conceptualización” siendo ésta “una visión abstracta y clara del mundo que queremos representar con algún propósito”.

La utilidad de una ontología según (Noy & McGuinness, 2000) se basa en:

- Compartir una comprensión común de la estructura de la información entre personas y agentes de software.
- Permitir la reutilización del conocimiento propio de un dominio.
- Hacer las asunciones propias de un dominio, explícitas para terceras partes ajenas a él.
- Separar el conocimiento del dominio, del conocimiento operacional.

- Analizar el conocimiento del dominio.

Existen algunas herramientas para la creación de ontologías como son:

- Protege²
- Cmap Tools COE³

Para la generación de ontologías se debe considerar, la reutilización de vocabularios existente, cumpliéndose así con uno de los principios de Linked Data, y evitado además la redundancia de términos; entre estos vocabularios tenemos:

- **Friend-of-a-Friend (FOAF)**⁴ , el vocabulario para describir a las personas.
- **Dublin Core (DC)**⁵ define los atributos generales de metadatos. Ver también sus nuevos dominios y rangos de proyectos.
- **(SIOC)**⁶ , el vocabulario para la representación de las comunidades en línea.
- **Doap** vocabulario para describir los proyectos.
- **SKOS**⁷ es el vocabulario para representar taxonomías y conocimientos poco estructurados.
- **Vocabulario revisión**⁸ , el vocabulario para representar opiniones.
- **Creative Commons (CC)**⁹ , vocabulario para describir los términos de la licencia.

1.2.3.2. Representación de sintaxis RDF.

El Resource Description Framework (RDF) es un estándar de la W3C, utilizado para representar la información de los recursos en la Web entre máquinas o plataformas; empleando además los vocabularios como los definidos en el punto anterior, para hacer más representativo el ámbito conocimiento.

Según cita (Lmarca, 2012) “En el caso de RDF es fundamental utilizar palabras que transmitan un significado inequívoco, con el fin de que las aplicaciones entiendan el enunciado para un procesamiento correcto. En RDF, este significado se expresa a

² <http://protege.stanford.edu/>

³ <http://www.ihmc.us/groups/coe/>

⁴ <http://www.foaf-project.org/>

⁵ <http://dublincore.org/documents/dcmes-xml/>

⁶ <http://sioc-project.org/>

⁷ <http://www.w3.org/2004/02/skos/>

⁸ <http://vocab.org/review/terms.html>

⁹ <http://creativecommons.org/ns>

través de un esquema. Podemos pensar en un esquema como una especie de diccionario que define los términos que se utilizarán en una declaración o sentencia RDF para otorgarle significados específicos. Con RDF se pueden utilizar una gran variedad de formas de esquema, incluyendo la definida en RDFS que posee características especiales para automatizar tareas utilizando RDF, pero también otras muchas formas”.

La W3C (Graham & Carroll, 2004) también utiliza las siguientes características que debería contener un RDF:

- Grafo de modelo de datos
- URI basado en vocabularios.
- Tipos de datos.
- Vinculación
- Sintaxis XML.

La estructura de un RDF maneja la información en triplas siendo así:

- *Sujeto*: Es el recurso al cual nos estamos refiriendo ó también conocido como concepto.
- *Predicado*: Es el recurso que indica lo que estamos definiendo ó denominado propiedad.
- *Objeto*: Puede ser un recurso que puede considerarse el valor definido ó contenido del sujeto.

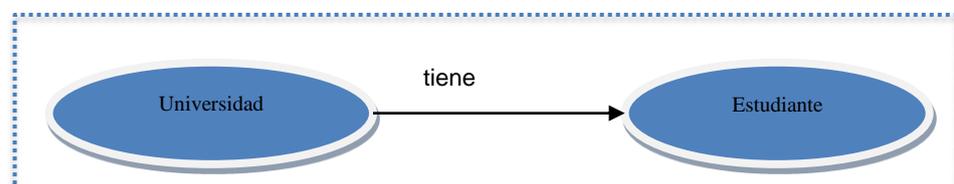


Figura. 3 Modelado en tripleta
Autor: Andrea Novillo

Sujeto(URI)	Universidad
Predicado(URI)	Tiene
Objeto(URI / LITERAL)	Estudiante

Figura 4. Expresión RDF

Autor: Andrea Novillo

RDF también emplea sintaxis basada en XML, en el siguiente ejemplo podemos visualizar su estructura.

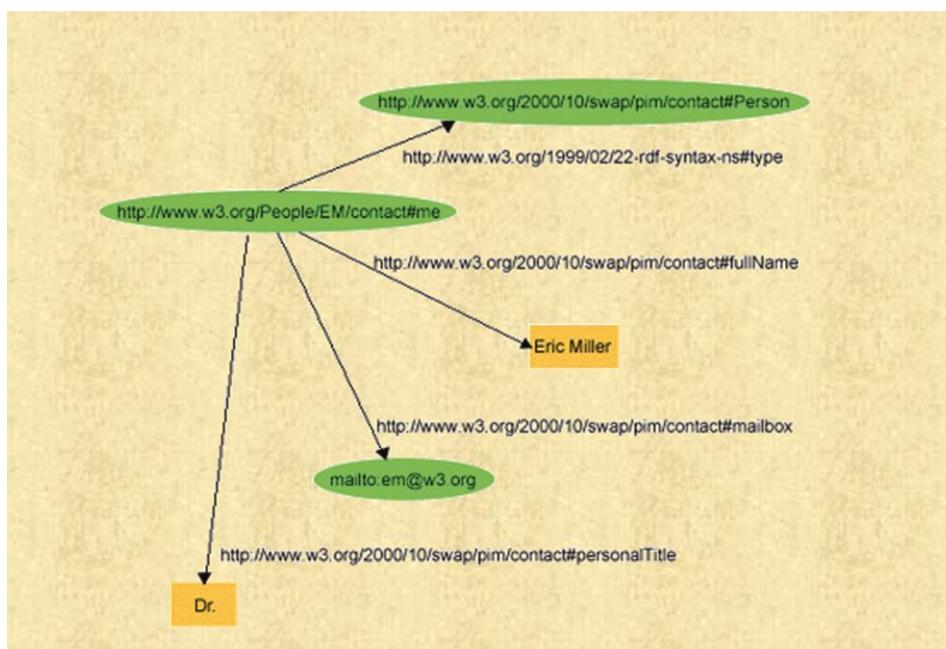


Figura 5. Grafo que describe a Erick Miller

Fuente: <http://www.hipertexto.info/documentos/rdf.htm>

```
<?xml version="1.0"?>
<RDF:RDF xmlns:RDF="http://www.w3.org/1999/02/22-RDF-syntax-ns#"
  xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">

  <contact:Person RDF:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox RDF:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</RDF:RDF>
```

Figura 6. Descripción de información en XML

Fuente: María Jesús Lamarca Lapuente, *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*. Encontrado en <http://www.hipertexto.info/documentos/rdf.htm>

La W3C (Manola, Miller&McBride, 2004), cita algunas características que se deben tener en consideración para un buen diseño RDF como son:

- *Independencia:* Cualquier usuario puede desarrollar.
- *Intercambio:* Una de las características de este estándar es que permite el intercambio de información de forma masiva, esto gracias a la utilización del lenguaje de marcas XML.
- *Escalabilidad:* RDF nos permite representar expresiones basadas en un modelo de tripletas, facilitando el manejo y uso de objetos en la búsqueda en grandes volúmenes de información.
- *Las propiedades son recursos:* las propiedades pueden tener sus propias propiedades a la vez, lo que permite ser manipuladas como cualquier otro recurso.
- *Los valores pueden ser recursos:* los valores de las propiedades debe incluir recursos.
- *Las sentencias pueden ser recursos:* Las sentencias deben tener sus propias propiedades.

El esquema RDF mantiene algunas clases y propiedades como:

Tabla 1. Clases del esquema RDF y RDFS.

Nombre de a Clase	Comentario
rdfs:Resource	La clase de recurso, cada uno.
rdfs:Literal	La clase del valor literal, por ejemplo, cadenas de texto y números enteros.
rdf:XMLLiteral	La clase de los valores literales de los valores literales XML.
rdfs:Class	La clase de las clases.
rdf:Property	La clase de las propiedades RDF.
rdfs:Datatype	La clase de los tipos de datos RDF.
rdf:Statement	La clase de las declaraciones RDF.
rdf:Bag	La clase de los contenedores desordenados.
rdf:Seq	La clase de los contenedores ordenados.
rdf:Alt	La clase de los contenedores de alternativas.
rdfs:Container	La clase de los contenedores RDF.
rdfs:ContainerMembershipProperty	La clase de las propiedades de los miembros contenedores, <i>rdf_1</i> , <i>rdf_2</i> , ..., todas ellas son subpropiedades de 'miembro'.
rdf:List	La clase de las listas RDF.

Fuente: http://www.hipertexto.info/documentos/esquemas_rdf.htm

Tabla 2. Propiedades del esquema RDF y RDFS.

Nombre de la Propiedad	Comentario	Domain (Dominio)	Range (Rango)
rdf:type	El sujeto es una instancia de una clase.	rdfs:Resource	rdfs:Class
rdfs:subClassOf	El sujeto es una subclase de una clase.	rdfs:Class	rdfs:Class
rdfs:subPropertyOf	El sujeto es una subpropiedad de una propiedad.	rdf:Property	rdf:Property
rdfs:domain	Un dominio de la propiedad del sujeto.	rdf:Property	rdfs:Class
rdfs:range	Un rango de la propiedad del sujeto.	rdf:Property	rdfs:Class
rdfs:label	Un nombre para el sujeto legible por seres humanos.	rdfs:Resource	rdfs:Literal
rdfs:comment	Una descripción del recurso sujeto.	rdfs:Resource	rdfs:Literal
rdfs:member	Un miembro del recurso sujeto.	rdfs:Resource	rdfs:Resource
rdf:first	El primer item en la lista RDF del sujeto.	rdf:List	rdfs:Resource
rdf:rest	El resto de la lista RDF del sujeto después del primer item.	rdf:List	rdf:List
rdfs:seeAlso	Más información sobre el recurso sujeto.	rdfs:Resource	rdfs:Resource
rdfs:isDefinedBy	La definición del recurso sujeto.	rdfs:Resource	rdfs:Resource
rdfs:value	Propiedad idiomática usada para valores estructurados .	rdfs:Resource	rdfs:Resource
rdf:subject	El sujeto de la declaración RDF del sujeto.	rdf:Statement	rdfs:Resource
rdf:predicate	El predicado de la declaración RDF del sujeto.	rdf:Statement	rdfs:Resource
rdf:object	El objeto de la declaración RDF del sujeto.	rdf:Statement	rdfs:Resource

Fuente: http://www.hipertexto.info/documentos/esquemas_rdf.htm

Además existen herramientas para trabajar en la creación del formato RDF, entre estas como principal se encuentra:

- **JENA¹⁰**: Es Framework creado en java por HP Labs permite la manipulación de meta datos para la Web Semántica, cuenta con la capacidad de implementar los modelos RDF (Resource Description Framework) y RDQL (RDF Data Query Language). Además permite leer, recorrer y modificar grafos basados en OWL para la realización de aplicaciones OWL.

1.2.3.3. Formato de explotación de información.

Al hablar de un formato de explotación de Información no tan solo es hablar de poder enlazar los datos en la Web y poder ser hallados por medio de buscadores, es hablar de como recuperar los datos en la Web semántica tomando en consideración que:

- Datos en RDF no servirían de nada si no se pueden re-utilizar.
- Los lenguajes de la Web semántica necesitan interactuar con los datos almacenados en la “base de datos” RDF.
- Necesidad parecida al lenguaje SQL de bases de datos relacionales.

1.2.3.3.1. Almacenamiento de información.

Existen aplicaciones y frameworks para almacenamiento y consulta de datos en formato RDF, creados con el fin de tener un almacén de datos semánticos como:

¹⁰ <http://jena.apache.org/>

- 4 Store¹¹
- Sesame¹²
- Virtuoso Open Source¹³
- AllegroGraph RDF Store¹⁴

1.2.3.4. Consumo y visualización de información.

Según el ciclo de vida de Linked Data la información obtenida en formato RDF, debe de ser consumida y explotada; es ahí donde nace la necesidad de consultar los almacenes de datos, que anterior mente se generaron por ello se necesita un lenguaje de consulta en el ámbito semántico, como es:

SPARQL¹⁵

Simple Protocol and RDF Query Language es un Lenguaje de consulta del ámbito de la Web semántica de W3C. Es decir define la sintaxis y la semántica necesarias para una expresión de consulta sobre un grafo RDF y las diferentes formas de resultados obtenidos.

Su misión es devolver todas las tripletas o componentes solicitados basándose en la comparación de una triplete pasada como parámetro de la consulta (grafo básico) con todas las tripletas que componen el grafo RDF. Cabe recalcar que su sintaxis es similar al estándar SQL de bases de datos relacionales.

Las consultas SPARQL cubren tres objetivos:

- Extraer información en forma de URIs y literales.
- Extraer sub-estructuras RDF.
- Construir nuevas estructuras RDF partiendo de resultados de consultas.

Siguiendo con el proceso de publicación el tener la información por medio de consultas SPARQL es necesario observar dicha información, por ello se obtienen aplicaciones o navegadores; cuyo objetivo principal es permitir a los usuarios navegar entre los diferentes datos que contienen enlaces RDF, entre ellos se encuentran:

- Tabulador¹⁶

¹¹ <http://4store.org/>

¹² <http://www.w3.org/2001/sw/wiki/Sesame>

¹³ <http://virtuoso.openlinksw.com/dataspace/dav/wiki/Main/>

¹⁴ <http://www.w3.org/2001/sw/wiki/AllegroGraph>

¹⁵ <http://xspqrql.deri.org/>

- Disco¹⁷
- OpenLink¹⁸
- LinkSailor¹⁹
- LOD Browser²⁰
- The Marbles Linked Data²¹

También se pueden utilizar servicios para encontrar URIs en la Web o enlaces RDF como:

- URIqr²²
- Sindice²³
- Sig.ma²⁴
- Falcons²⁵
- SWSE²⁶

1.2.4. Proceso de publicación de datos.

Bajo el enfoque de los principios de Linked Data es importante tomar en consideración un modelo de datos RDF, estandarizado para publicarlo en la Web y el otro punto es utilizar enlaces RDF para interconectar los elementos de datos con otras fuentes, además de los principios ya mencionados en el punto anterior.

Para tomar en consideración la forma de publicación de datos es necesario que se planteen las siguientes preguntas como cita (W3C, 2010) ¿Con que cantidad de datos se cuenta, ¿Cómo se almacenaran los datos?, ¿Cómo actualmente se encuentran los datos a vincular?; teniendo en consideración las preguntas planteadas, se tomar en cuenta que es necesario saber que se va a publicar conociendo el ámbito de trabajo, segundo hacia quienes van dirigidos estos datos y tercero quienes van a reutilizar dicha información. Luego de analizar los diferentes ambientes de la información podemos considerar las respuestas a las preguntas citadas, dando así una idea del ámbito en el que nos encontramos y que aplicar para lograr el objetivos deseado; aplicando algunas

¹⁶ <http://www.w3.org/2005/ajar/tab>

¹⁷ <http://wifo5-03.informatik.uni-mannheim.de/bizer/ng4j/disco/>

¹⁸ <http://www.openlinksw.com/>

¹⁹ <http://linksailor.com/nav>

²⁰ <http://browse.semanticweb.org/>

²¹ <http://marbles.sourceforge.net/>

²² <http://URIqr.com>

²³ <http://www.sindice.com/>

²⁴ <http://sig.ma/>

²⁵ <http://iws.seu.edu.cn/services/falcons/documentsearch>

²⁶ <http://www.swse.org/>

fórmulas o recetas como ellos les denominan a los pasos de enlazado de datos y el uso de posibles herramientas.

1.2.4.1. Ciclo de Vida del proceso de publicación de Linked data.

Conocer el método de enlazado de datos implica tener una noción de que procedimiento se debe seguir, por ello se necesita plantear un ciclo de vida para conocer paso a paso el trabajo realizado como lo mencionan (Corcho & Gómez, 2010) a continuación se detalla:

- Especificar: Se necesitará conocer el entorno de trabajo a utilizar.
- Modelar: Para ello se generará una representación grafica del entorno de trabajo denominadas ontologías y aplicando el uso de vocabularios.
- Generar RDF: Para que puedan ser reconocidas por máquinas deben estar descrita la información en formatos que estas entiendan.
- Enlazar: Para generar un enlace de datos se debe guardar la información en almacenes de datos y consumir fuentes de datos que tengan similitud o manejen el mismo entorno.
- Publicar: Publicar podría significar hacer pública la información de los nuevos metadatos bajo el concepto de vocabularios.
- Explotar: Para este último paso se deberá consumir la información, mediante consultas que den un toque semántico a la información que los usuarios estarán buscando. Además de vincular la información a la *nube de datos abiertos*.

Es un poco desordenado no conocer a fondo cada uno de las etapas del proceso de publicación de información, por ello (Corcho & Gómez, 2010) nos brindan un enfoque más detallado de este proceso como podemos ver en figura 7.

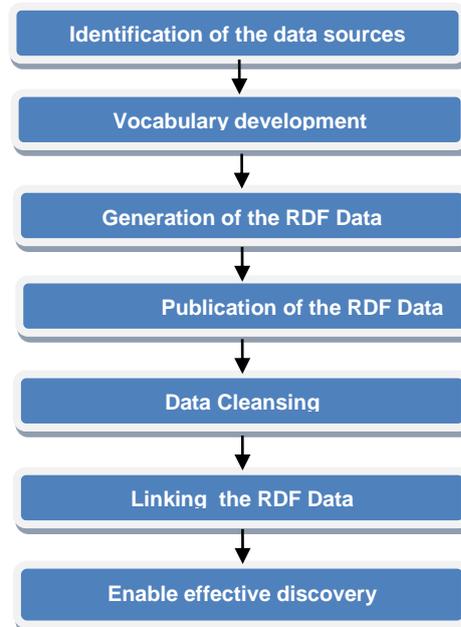


Figura 7. Proceso de Publicación de Linked data en la Web

Fuente: Corcho, O. & Gómez, A... Universidad Politécnica de Madrid
 Florianópolis, *Herramientas de publicación de datos para portales de datos
 abiertos*, September 1st 2010 (3º OntoBras 2010). Sitio web:
[http://datos.gob.es/datos/sites/default/files/files/%2004.V3_Tecnologia_Portal
 es_OpenData.pdf](http://datos.gob.es/datos/sites/default/files/files/%2004.V3_Tecnologia_Portal

 es_OpenData.pdf)

Sin embargo un proceso de publicación de datos puede mejorarse o adaptarse a cada uno de los proyectos con los que se trabaje, existen algunos procesos más entre ellos está el planteado por (Bizer, y otros, 2009), este proceso de publicación consta de tres pasos:

1. Asignar URIs a las entidades descritas por el conjunto de datos, y luego utilizar el protocolo HTTP para cambiar estas referencias por representaciones RDF.
2. Establecer enlaces RDF sobre las fuentes de datos de la Web, para que los clientes puedan navegar por la Web de datos siguiendo los enlaces RDF.
3. Proporcionar metadatos sobre los datos publicados, para que los clientes puedan evaluar la calidad de los datos publicados y elegir entre diferentes medios de acceso.

Se han comparado dichos procesos de publicación, en los cuales se ha descubierto la similitud de desarrollo, como principal punto ambos procesos mantienen el modelado de las fuentes de datos por medio de implementación de ontologías y la revisión de vocabularios como FOAF, SIOC, DublinCore, AIISO, etc basándose en el concepto de Reutilización, además del uso del protocolo HTTP, siendo el citado por (Corcho&Gomez,2010) un poco más completo al detallar cada una de las actividades a desarrollar é incluyendo la limpieza de los datos para evitar ambigüedades.

Se han comparado características y usos de las herramientas empleadas dentro de la Web semántica para la publicación de información; llegando a tener un marco claro de que se va a emplear en el presente proyecto y según la metodología empleada por (Corcho&Gomez,2010) denotándose así:

1. Para Identificación de datos y *modelamiento de Información* se empleará CmapTools, pues tiene un fácil uso, permitiendo al usuario manipular clases, propiedades y un sinfín de metadatos. Además de emplear vocabularios existentes evitando la redundancia.
2. Para la generación de los datos en formato RDF, se empleará el framework de java denominado JENA en el cual se provee un soporte de modelos RDF y es fácil generar la salida de información en formatos de serialización RDF como son Xml/RDF, Turtle, N3, etc.
3. Para el almacenamiento y publicación de datos, según la comparación de rendimiento y facilidad de uso se empleará Virtuoso Open Source y para su consumo se maneja el endpoint SPARQL que este incluye.

1.3. Descripción del Ambiente de datos OCW y OER.

1.3.1. Descripción OCW.

OpenCourseWare (OCW) es una publicación digital abierta y gratuita de materiales educativos alta calidad. Están clasificados desde cursos completos hasta asignaturas. Mantiene una licencia abierta accesible a cualquiera y en cualquier lugar de la red así lo menciona (OE CONSORTIUM, 2012), puesta en marcha en Abril del 2001. Basada en internet y fundada conjuntamente por el Instituto Tecnológico de Massachusetts (MIT) en colaboración con la Fundación William and Flora Hewlett y la Fundación Andrew w. Mellon. (UNIVERSIA OCW, 2012).

Esta iniciativa nació de la necesidad de consumir y distribuir de una forma libre los recursos educativos, que cada centro universitario tiene en sus propios repositorios y que en ocasiones tan solo sirve para sus alumnos, docentes, etc; transformándose en una mejora de la Web semántica basada en la idea de Tim Berners-Lee.

En base a la necesidad mencionada es que aparece OpenCourseWare Consortium (OE CONSORTIUM, 2012), que es una comunidad mundial de cientos de instituciones de educación superior y organizaciones asociadas comprometidas a promover OpenCourseWare y su impacto en la educación global, siendo un órgano de

coordinación para el movimiento a escala mundial y como un foro para el intercambio de ideas y planificación futura.

Estos son algunos de los beneficios que brinda el consorcio:

Consulta: Acceso toda la información que mantienen los diferentes sitios OCW.

Colaboración: Puede obtener y brindar información OCW.

Visibilidad colectiva: Aumentar la posibilidad colaborativa en la red, por medio de intercambio de recursos educativos abiertos.

Un sitio OCW es según (UNIVERSIA OCW, 2012):

- Un espacio web que contiene materiales docentes creados por profesores para la formación superior.
- Una representación de un conjunto de recursos (documentos, programas, calendario, etc) utilizados en el proceso de enseñanza-aprendizaje de las asignaturas que imparten los profesores.
- Ofrecidos libremente y son accesibles universalmente en la red.
- Son sitios de difusión respetando la propiedad intelectual.
- Se permite el uso, la reutilización, la adaptación y la distribución por otros, con ciertas restricciones.
- Un sitio OWC no es un centro de educación a distancia ni de interacción entre alumno y profesor.

Según (Universia OCW, 2012) entre las características de utilizar OWC y algunos inconvenientes que se derivan, están:

Beneficios:

- Apoyo al conocimiento al liberar recursos didácticos.
- Estimula la innovación y perfeccionamiento de los recursos docentes utilizados por los profesores. Generando calidad de recursos.
- Favorece el reconocimiento de los docentes por medio de la propiedad intelectual y el reconocimiento de la autoría.

- Apoya al principio de casa sitio OCW, que es generar recursos educativos de calidad y libre consumo.

Inconvenientes:

Entre los inconvenientes que se encuentra en el uso OCW tenemos:

- Obliga al profesor a usar la estructura definida por cada sitio OCW.
- Limita la comercialización de materiales docentes.
- Posible duplicidad en la gestión de recursos al atender dos escenarios diferentes.

Según lo mencionado en (MitOCW Rep, 2011), actualmente existen aproximadamente 2,083 cursos con acceso libre publicados. En la figura 8 podemos observar el tipo de usuarios y el porcentaje de uso del proyecto OCW.

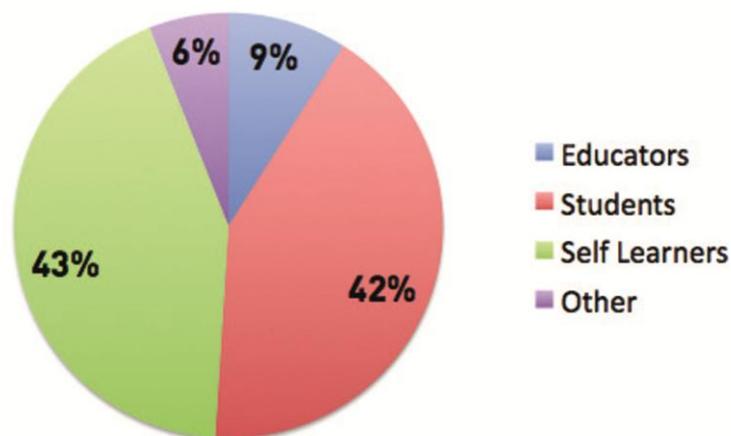


Figura. 8 Tipo de usuarios OWC
Fuente: <http://OCW.mit.edu/about/site-statistics/>

1.3.2. OCW y OER en Iberoamérica.

No solo los países desarrollados son aquellos que se han adentrado en la ola de los OCW; Iberoamérica es un grupo de países que están buscando el desarrollo y no solo en la parte industrial sino además en la parte educativa, puesto que la educación en estos países esta siendo el principal aporte luego de sustentabilidad económica. Para ello será necesario obtener y colaborar con recursos basados en el principio del Consorcio OCW, por lo cual nace Universia que es quien dirige de manera alguna estos países iberoamericanos.

Estos son algunos países que están siguiendo el enfoque de OPEN DATA ²⁷(OpenData, 2012) y la compartición de estos por medio de las redes OCW como es el proyecto Universia (Universia OCW, 2012) entre ellos están:

Tabla 3. Países Sudamericanos con sitios OCW

Países Sudamericanos	
México	República Dominicana
Venezuela	Uruguay
Brasil	Portugal
Chile	Ecuador
Colombia	Argentina

Autor: Andrea Novillo

La inclusión de cada uno de los países y sus respectivas universidades como ya mencionamos, resulta de las necesidades de compartir información enfocándose en áreas nuevas de conocimiento y tratando de afianzar los conocimiento en algunos países por medio de recursos educativos abiertos ,dentro del ámbito de los países Sudamericanos podemos hablar de Ecuador, siendo la Universidad técnica particular de Loja la primera en pertenecer al grupo de Universia²⁸ (Universia OCW, 2012) y al Consorcio OCW ²⁹ (OE CONSORTIUM, 2012) por medio de su sitio OCWUTPL³⁰, a partir del año 2009 basándose en los principios de OCW y con el propósito de:

- Proporcionar **acceso libre**, sencillo y coherente en los materiales de los cursos se ofrecen, tanto en sus modalidades de estudio presencial como a distancia; dirigido para educadores, estudiantes y autodidactas de todo el mundo.
- Crear un **modelo eficiente** basado en estándares generados de OCW Consortium y UNIVERSIA, así como de otros organismos afines.

²⁷ <http://www.opendatafoundation.org/>

²⁸ <http://www.universia.net/>

²⁹ <http://www.ocwconsortium.org/>

³⁰ <http://ocw.utpl.edu.ec>

1.3.3. Contenido de OCW.

Los contenidos de OCW están estructurados en lo general de acuerdo a las necesidades de cada uno de los sitios. En la mayoría de estos sitios se clasifica la información bajo el nombre de cursos y cada uno de estos pertenece a un área técnica, biológica, comunicativa, etc, dichos cursos además contiene un menús como pueden ser guías didácticas, ejemplos, áreas de estudios, tipos de materias, etc. Esto no se puede estandarizar de manera general para los sitios OCW, puesto que las necesidades son diferentes y las palabras ó clasificaciones utilizadas son propias de cada uno de ellos.

El lenguaje de los recursos ó cursos básicamente, mantienen idiomas como español, inglés y catalán a nivel de Latinoamérica. En consorcios occidentales se manejan lenguajes básicos como japonés, chino tradicional é inglés.

Tomaremos como punto de referencia la UTPL a nivel de Iberoamérica para demostrar como esta estructurada la información en su sitio OCW.

OpenCourseWare UTPL

OpenCourseWare (OCW) es un ejemplo de las iniciativas que desde el 2001 han emergido para promover el acceso libre y sin restricciones al conocimiento. Desde el 2009, la [Universidad Técnica Particular de Loja](#), Ecuador, (UTPL) ha instalado una oficina de OCW y se ha unido al consorcio OCWC y a Universia. La red de universidades iberoamericanas, [Universia](#), desde el principio se ha comprometido con estos objetivos, y de acuerdo con universidades españolas promueve el consorcio [OCW Universia](#) al que [UTPL](#) pertenece.

Social UTPL OER

- [OerProject](#)
- [OpenUTPL](#)
- [Wiki ECC](#)
- [Blog ECC](#)

La UTPL ha decidido establecer su sitio OpenCourseWare en <http://ocw.utpl.edu.ec/>

Figura 9. Portada del sitio OCW de la Universidad Técnica Particular de Loja
Fuente: <http://ocw.utpl.edu.ec/>

1.3.4. Estructura de los sitios

La estructura general de cada curso a nivel de los consorcios depende de las tecnologías empleadas para su desarrollo, pero además cuenta con estructuras similares como se muestra a continuación:

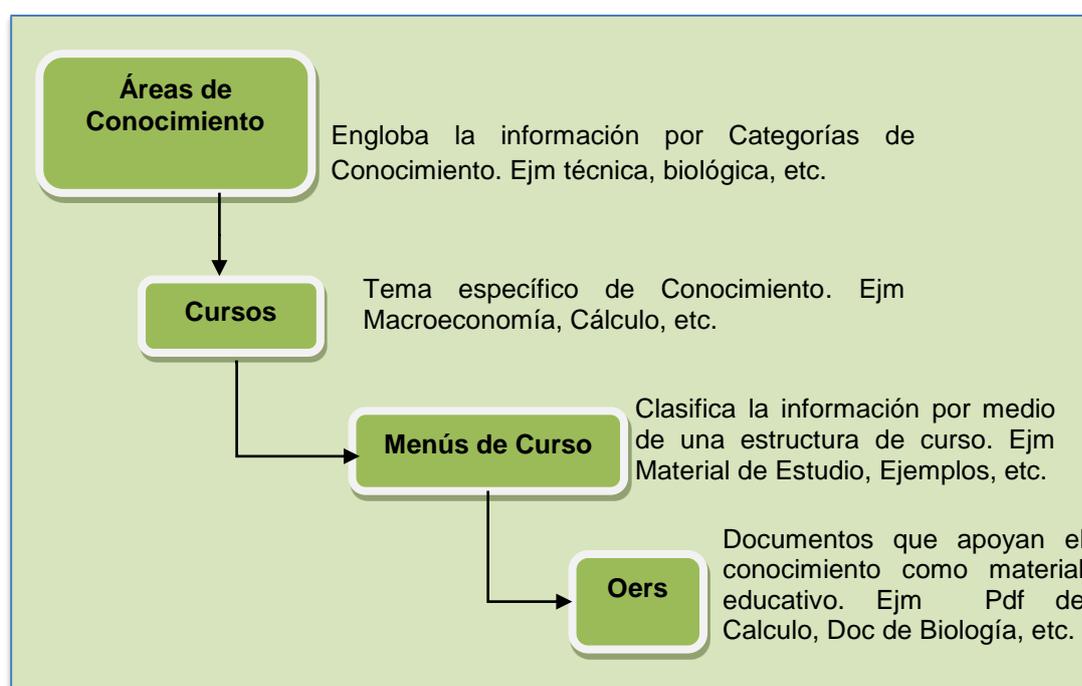


Figura. 10 Estructura generalizada de los sitios OCW
Fuente: Andrea Novillo.

1.3.5. Plataformas usadas en OCW

Como se ha mencionado anteriormente para poder tener en línea la información de los OCW, se necesitan tecnologías o plataformas de desarrollo. En el **anexo 7**, Borra realiza un estudio de las plataformas utilizadas entre sitios OCW y de aquellos sitios que se encuentran activos e inactivos.

Los sitios OCW en su mayoría emplean plataformas como Educommons y páginas en HTML; la primera fue propiamente creada para el apoyo de estos sitios siendo la más usada a nivel de Sudamérica, y la segunda es empleada en la mayoría de consorcios orientales usada para acoplarse a un estándar según la necesidad de la institución. No podemos dejar de mencionar además Moodle cuya plataforma fue creada prácticamente con fines educativos, pero el principal detalle de esta es que contiene mecanismos de acceso, como son los login de ingreso.

En las siguientes gráficas se puede observar las diferentes plataformas y la estadística de uso de estas:

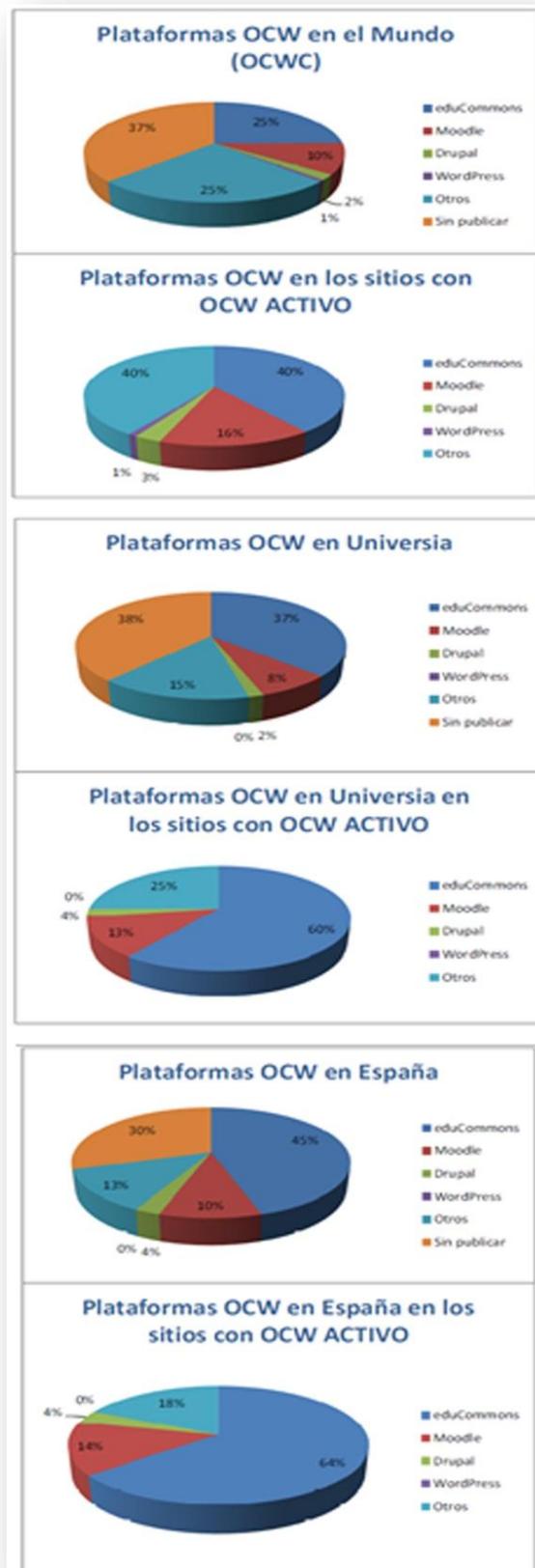


Figura 11. Estadística de Plataformas OCW.

Fuente: Borra, O. (2010), Observatorio de Plataformas para OCW (2012): Madrid.

1.3.6. Licenciamiento de los OCW.

A pesar de que son recursos compartidos libremente, estos no deberían dejar de lado la parte legal, puesto que el esfuerzo de las personas debe ser reconocido y la retroalimentación de estos debería quedar plasmada como propiedad intelectual de quienes apoyan.

A continuación se pueden observar las licencias usadas por los sitios OCW, así como también el reglamento que utilizan para que cada uno de estos recursos pertenezca a un sitio. Cabe recalcar que la licencia mayormente empleada a nivel de los sitios OCW es la de Creative Commons.

GPL³¹

Esta es la licencia más conocida en nuestro medio, fue creada por la “Free Software Foundation” cuyo propósito general fue el proteger al software de intentos de apropiación protegiendo así las libertades de los usuarios y de las aportaciones de otro.

Como se cita en (GNU, 2010) las libertades que esta licencia presta a los usuarios en general son:

- Libertad de utilizar el programa con cualquier propósito.
- Libertad de estudiar cómo funciona el programa, modificarlo y adaptarlo a sus necesidades.
- Libertad de distribuir copias, con lo que se puede ayudar al prójimo.
- Libertad de mejorar el programa y hacer públicas las mejoras a los demás, de modo que toda la comunidad se beneficie.

Mit License³²

Mit license es también una licencia de software libre esta licencia está incluida en X Windows System que es originario del Mit en los años 1980.

Según cita (MIT license, 2009) esta licencia consta de condiciones, derechos y limitaciones de responsabilidad, siendo muy similar a la licencia BSD.

³¹ <http://www.gnu.org/copyleft/gpl.html>

³² <http://www.opensource.org/licenses/MIT>

“Copyright (c) <año> <copyright holders>

Se autoriza, de forma gratuita, a cualquier persona que obtenga una copia de este software y archivos de documentación asociados (el "Software"), para trabajar con el Software sin restricción, incluyendo sin limitación los derechos para usar, copiar, modificar, fusionar, publicar, distribuir, su licenciar, y / o vender copias del Software, y para permitir a las personas a quienes se suministra el software para hacerlo, con sujeción a las siguientes condiciones:

Creative Commons ³³

Toda obra del tipo que esta sea cuenta con Copyright³⁴ que son derechos de propiedad Intelectual generados automáticamente. Creative Commons es una entidad sin fines de lucro y nace como una alternativa a la propiedad intelectual, en el cual se liberaran algunos derechos y mientras que otros se mantienen y así las obras no serán obsoletas.

La visión de Creative Commons esta basada en explotar todo lo que hay en el internet, teniendo un acceso libre a todo recurso como es en educación y cultura aportando de esta manera a un crecimiento por medio de la retroalimentación.

Se debe indicar que Creative Commons Corporation no es un despacho de abogados, ni ofrece servicios legales. La distribución de esta licencia no crea una relación abogado-cliente. Creative Commons no ofrece garantías sobre la información suministrada, ni asume responsabilidad por los daños causados por su uso.

Creative Commons ofrece distintas combinaciones de propiedades como:

- **Atribución (By):** Obliga a citar fuentes y autores de la obra.
- **No comercial (NC):** La obra no podrá ser comercializada bajo ningún contexto.
- **Sin obra derivada (ND):** Obliga a que esa obra sea distribuida inalterada, sin cambios.
- **Compartir igual (SA):** Obliga a que si la obra tiene un mejoramiento esta sea compartida bajo las mismas licencias que la original.

³³ <http://creativecommons.org/>

³⁴ http://es.wikipedia.org/wiki/Derecho_de_autor

En la **figura 12** se observan otras combinaciones posibles:

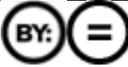
COMBINACIÓN	PROPIEDAD
	By
	By+ND
	By+ND+NC
	By+NC
	By+NC+SA
	By+ SA

Figura 12. Combinación de licencias Creative Commons
 Fuente: <https://creativecommons.org/>

1.3.6.1. Términos de licencias y uso de contenidos OCW.

Conociendo el concepto de las diversas licencias dentro de la mayoría de sitios OCW originalmente se utilizaron licencias **MIT**, conforme fueron avanzando las necesidades de cada sitio se aplicaron licencias GPL hasta el 2004. Actualmente en la mayoría de sitios Web y en especial los principales consorcios como Universia⁵ y OCW Consortium⁶ utilizan licencias Creative Commons Attribution 3.0 License (CC 3.0License) que como citan dentro de su información utilizan propiedades como Reconocimiento, No comercial y compartir igual.

En la siguiente gráfica se puede observar de una manera detallada la licencia Creative Commons 3.0, que actualmente es empleada por la mayoría de sitios OCW.

Usted es libre de:	
	Compartir – copiar, distribuir y compartir públicamente la obra.
	Mezclar - adaptar el trabajo
Bajo las siguientes condiciones:	
	Reconocimiento – Debe reconocer el trabajo de la manera especificada por el autor (de la manera que sugiera que se apoye su obra)
Con Entendimiento de que:	
<p>Renuncia - Cualquiera de estas condiciones puede ser renunciada si se obtiene permiso del propietario del copyright.</p> <p>Dominio público – Cuando una obra o cualquiera de sus elementos se encuentran bajo el dominio público según la ley aplicable, dicho estado no es de ninguna manera afectado por la licencia.</p> <p>Otros derechos – De ninguna manera cualquiera de los siguientes derechos puede ser afectado por la licencia: Trato justo o uso correcto de sus derechos o de otros derechos de autor aplicables a las excepciones y limitaciones/ Moral del autor. Derechos de otras personas que puedan tener sobre la obra, o sobre la manera en que la misma pueda ser utilizada, como publicidad o derechos de privacidad.</p> <p>Aviso – Para cualquier reutilización o distribución, se deben dejar claros los derechos de la obra. La mejor manera de hacerlo es mediante un enlace a la página web (http://creativecommons.org/licenses/by/3.0)</p>	

Figura 13. Resumen de licencia de OCW

Fuente: <http://creativecommons.org/licenses/by/3.0/>

Cada sitio OCW mantiene un reglamento interno, políticas de seguridad y códigos de ética para aportar a las licencias utilizadas en general aquellas universidades que están vinculadas a alguna plataforma OCW. En el caso de Universia cuenta con avisos legales, código de ética y en OCW Consortium términos y Condiciones de uso.

1.3.7. Calidad de contenidos.

Conociendo que los datos están inmersos en la red, pero la mayoría de recursos educativos aunque siguen normas de publicación que la Universidad cree convenientes; por ello es necesario mejorar la calidad de los recursos OCW para que sean recursos de calidad certificados.

Según (*Piedra & Romero, 2012*) este es un modelo de calidad que encierra el concepto de OCW para evaluar cada uno de los recursos.



Figura 14. Áreas que soportan modelo de calidad de contenidos OCW

Fuente: Piedra, N. & Romero, A... (2012). *calidad de contenidos en OCW*.

Disponible en

http://repositoral.cuaed.unam.mx:8080/jspui/bitstream/123456789/2706/1/romero_piedra_calidad_de_contenidos_OCW.pdf

El modelo antes mencionado se basa en la propuesta del modelo de calidad de recursos educacionales, obteniendo también un bosquejo de algunos criterios para su evaluación (Piedra & Romero, 2012).

AREA DEL MODELO	CRITERIOS
Modelos de desarrollo, distribución y licenciamiento	<ul style="list-style-type: none"> Debe usarse contenido bajo licencia Creative Commons . Debe prestarse atención al renombre de la Universidad que ofrece el recurso.
Alcance académico	<ul style="list-style-type: none"> El contenido está correctamente explicado, no presenta errores de redacción que puedan generar confusión. Cubre con los contenidos que se pretende cubrir de una forma conveniente para el estudiante al que está dirigido. Permite al estudiante lograr los objetivos de estudio del material o al docente explicar un contenido.
Presentación al usuario	<ul style="list-style-type: none"> El recurso tiene una navegación amigable y es fácil de usar. La disposición de los elementos de presentación es adecuada para los usuarios hacia los que va dirigido. Contiene muy pocos elementos distractores. En el caso de elementos multimedia debe considerarse la calidad del recurso y que el formato del archivo sea asequible para el estudiante al que va dirigido. En el caso de datos en bruto, el profesor adicionar un documento que explique el propósito del recurso y cómo trabajarlo, caso contrario no debe utilizarse.
Requerimientos tecnológicos e interoperabilidad	<ul style="list-style-type: none"> El recurso funciona sin problemas en la plataforma adecuada. Considere que los recursos contengan tags que permitan asociarlos con un área de conocimiento.
Accesibilidad	<ul style="list-style-type: none"> Los recursos deben encontrarse en diferentes formatos que sean accesibles. Disponible en medios que incluyan texto alternativo o subtítulos. Incluir alternativas al uso de texto como imágenes. En videos incluir subtítulos o etiquetas. Para páginas que siga las recomendaciones de W3C

Figura 15. Criterios de calidad.

Fuente: Piedra, N. & Romero, A... (2012). *calidad de contenidos en OCW*.

Disponible en

http://repositorial.cuaed.unam.mx:8080/jspui/bitstream/123456789/2706/1/romero_piedra_calidad_de_contenidos_OCW.pdf

1.3.8. Descripción de OER

Los Recursos Educativos Abiertos (REA) o sus siglas en inglés (OER), son recursos que han sido creados para usarlos como parte de la enseñanza y aprendizaje, están disponibles al público en general de una manera libre. Estos pueden ser libros, imágenes, videos, etc.

Según Cita (UNESCO 2002) son recursos para la enseñanza, el aprendizaje y la investigación, que residen en el dominio público o han sido publicados bajo una licencia de propiedad intelectual, que permite que su uso sea libre para otras personas. Incluyen: cursos completos, materiales para cursos, módulos, libros de texto, vídeos, pruebas, software y cualquier otra herramienta, materiales o técnicas utilizadas para apoyar el acceso al conocimiento.

En base a la investigación realizada se puede afirmar que son recursos educativos, que se encuentran en la Web al alcance de todo tipo de personas y en cualquier lugar del mundo; pudiendo ser utilizados para crear una educación de calidad para todos como respaldo académico, sin dejar de pasar por alto el reconocimiento de la propiedad intelectual de sus autores y contribuyentes siendo además un repositorio de objetos de aprendizaje.

Se deben considerar tres elementos fundamentales de REA:

- **Contenidos Educativos:** Recursos educativos como libros, programas educativos completos, módulos de contenido, publicaciones, etc.
- **Herramientas:** Software para apoyar la creación, entrega (acceso), uso y mejoramiento de contenidos educativos abiertos
- **Recursos de Implementación:** Parte de legal de recursos, Diseño, adaptaciones del contenido, materiales y técnicas para apoyar al acceso del conocimiento.

1.4. Repositorios OCW.

En el presente proyecto se nombran algunos repositorios que están siendo centro de compartición de recursos educativos libres y de calidad.

MitOpenCourseWare (Massachusetts Institute of Technology)³⁵

Nace como un primer Proyecto editorial OER/OWC en el instituto de Tecnología de Massachusetts en el año 2001 con vocación internacional y de creación de un movimiento de difusión libre de materiales docentes, con el fomento del trabajo cooperativo.

Universia³⁶

³⁵ <http://ocw.mit.edu>

³⁶ <http://www.fundacionuniversia.net>

Fundación Universia persigue favorecer la inclusión educativa y laboral de las personas con discapacidad, con especial atención en el ámbito universitario y en el empleo de calidad, utilizando como instrumento las Nuevas Tecnologías de la Información y las Comunicaciones. Basándose en el OpenCourseWare (OCW) que es una iniciativa editorial electrónica a gran escala y vincula recursos por medio de la cooperación de universidades a nivel de Latinoamérica. Convirtiéndose así en un gran almacén de información libre.

OCWConsortium³⁷

Este proyecto nace en el 2006 como una comunidad mundial de cientos de instituciones de educación superior y organizaciones asociadas se comprometieron a avanzar en OpenCourseWare y su impacto en la educación global. Servimos como un recurso para iniciar y mantener proyectos OCW, como un órgano de coordinación del movimiento a escala mundial, y como un foro para el intercambio de ideas y la planificación futura.

A continuación podemos citar algunos proyectos que se han formado a partir de la necesidad de vinculación de datos libres a través de la Web, sirviendo de esta manera como punto de encuentro de dichos recursos.

Open Data³⁸

Es el mejor de los ejemplos de vinculación de datos por medio de Linked Data, objetivo es impulsar la Web de datos mediante la identificación de conjuntos de datos existentes que estén disponibles bajo licencias abiertas, convirtiéndolas en RDF bajo los principios de Linked Data y publicarlos en la Web

Course Ware³⁹

Es un repositorio semántico que contiene y publica datos vinculados RDF, los datos presentes son provistos desde el proyecto ReSIST, el mismo que posee más de 50 millones de ítems, recomendados para todos los involucrados con la enseñanza de los temas relacionados.

Se pueden encontrar también algunos sitios de interés que aporten información acerca de calidad y contenidos OER/OCW como los citados a continuación.

Opal⁴⁰

³⁷ <http://www.ocwconsortium.org/>

³⁸ www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

³⁹ courseware.rkbexplorer.com

Nace por la necesidad de obtener una mejora en cantidad y calidad de recursos educativos abiertos que pueden ser incorporados a la educación superior.

En 2012, la Iniciativa OPAL se han establecido las bases para el concepto emergente de las prácticas educativas abiertas. Se han construido un fácil de utilizar, marco de calidad dinámico y viable para la OEP, que será validado por un auto-sostenible, auto-organizada de la comunidad

5. Plataforma OER⁴¹

A partir de noviembre del 2011 esta puesta en marcha la Plataforma REA es una nueva e innovadora plataforma en línea que ofrecerá una selección de publicaciones de la UNESCO como REA plenamente autorizados.

La Plataforma se ha creado con software libre de la Universidad de Witwatersrand (Sudáfrica) en el marco del consorcio AVOIR (Africa Virtual Open Initiatives and Resources) integrado por once universidades africanas.

1.5. Análisis.

En el presente Estado de arte, se analizan aquellas directrices para aplicar técnicas de enlazado de datos en el ambiente de trabajo Ocw y sus respectivos Oers. Primeramente se debe conocer en que ambiente y para que se desarrolló Linked Data, por lo cual se inicia conociendo el desarrollo de la web desde sus inicios hasta lo que hoy conocemos como Web 3.0, obteniendo así la premisa de que Linked data es el corazón de la Web semántica la cual tiene como idea principal interconectar datos, generando interoperabilidad entre máquinas sin la necesidad de operadores e inferencia humana.

Uno de los principales pasos para desarrollar el proyecto es la extracción de información, pues al momento de analizar los diferentes sitios Ocw podemos darnos cuenta que la información es tan amplia que necesitaremos realizar un proceso rápido, fácil , eficaz y automático para la obtención de información; es ahí donde se indaga sobre técnicas de extracción que apoyen principio fundamental por el que se rige este proyecto como es Open source; conociendo que nuestra información se encuentra almacenada en sitios web se analizan técnicas de Web scraping, entre estas analizamos como prioridad el framework Scrapy desarrollado en lenguaje Python ,por ser simple y 100% Open source.

⁴⁰ <http://www.oer-quality.org/>

⁴¹ <http://www.wsis-community.org>

Para continuar con el desarrollo de proyecto se investiga Linked data, el cual es el proceso para la publicación de datos estructurados que permite la interconexión de datos y da mayor utilidad a la información sobre la cual se trabaja. Se conocen además los principios bajo los cuales se maneja, los beneficios de su uso y las tecnologías aplicables en el proceso de publicación. Uno de los procesos de publicación de datos mas confiable y entendible es el citado por Corcho & Gómez, pues el proceso es mas claro y detallado, pudiendo ser verificado en el transcurso del ciclo de vida que se emplea.

Analizando cada uno de los pasos del ciclo de vida de la publicación de datos se analizan tecnologías que afianzaran al correcto desarrollo del proceso como son:

- Especificar: Para afianzar dicho paso se debe conocer a fondo el entorno de trabajo, generándose como primer paso la extracción de información, de la cual se hablo previamente; analizándose por tal motivo las diferentes técnicas de extracción poniéndole mayor cuidado al framework Scrapy.
- Modelar: Para conocer y generar el diseño de nuestro entorno de trabajo como son sitios Ocw y sus respectivos Oers, se analizan herramientas de diseño bajo la connotación semántica entre ellos tenemos Cmaptools pues es accesible y de uso fácil; además se analizan los distintos vocabularios que se pueden reutilizar para generar un vocabulario robusto para nuestro entorno de trabajo.
- Generar RDF: Existen distintas formas de generar formatos RDF, pues va desde la creación manual hasta la creación mecánica. Por lo cual se analiza con mayor cuidado la librería Jena, que es fácil y mucho más robusta para el desarrollo de formatos RDF.
- Enlazar, publicar, explotar: Se analizan varios almacenes de datos, analizando mas a fondo Virtuoso, por ser un completo servidor que soporta almacenar modelos RDF y publicación de datos mediante el endpoint de SPARQL.

Un análisis de la situación actual de los Ocw y sus respectivos Oers, nos da una pauta de que es lo que se quiere publicar y hacia quienes va dirigido el presente proyecto. Afianzando así a las propuestas educativas que hoy en día apuntan a la interoperabilidad y reutilización de recursos, que aporten conocimiento a las personas y sobre todo el uso libre de dicha información bajo el contexto Open source.

2. ANÁLISIS DE FUENTES DE DATOS Y VOCABULARIOS

2.1. Introducción.

Como se menciona en la introducción, la identificación de fuentes de datos es el punto inicial según el proceso de publicación, en el cual se establecerá el dominio de datos a publicar, listando conceptos, clases y propiedades en el ámbito de trabajo.

En contexto de trabajo la información se encuentra vinculada a los sitios OCW que mantienen una estructura dependiendo de la plataforma en que estos se encuentren desarrollados.

2.2. Análisis de datos OCW y OER almacenados en HTML.

El ambiente de trabajo con el que se maneja el presente proyecto, es la extracción de información desde sitios OCW almacenados en HTML, conteniendo información de la Universidad a la que pertenece, áreas de conocimiento, menús de información y contenido, cursos; dentro de este último se encuentran los recursos educativos que se desean publicar mediante los principios de Linked Data.



Figura 16. Estructura de sitios Ocw y contenido OER
Autor: Andrea Novillo

2.2.1. Identificación de fuentes de datos.

Analizando la información que se desea obtener y conociendo el concepto de este proyecto, se han verificado sitios OCW que se encontraron en los diversos consorcios como OCW Consortium, Universia, entre otros; el primer paso para verificar la información de estos sitios es la verificación de su estructura.

Se pueden mencionar aquellas universidades que tienen sitios OCW activos, para esta preparación de datos se ha trabajado bajo la observación de información que contienen, entre ellos tenemos la Universidad Politécnica de Madrid y la Universidad Técnica Particular de Loja, puesto que estas instituciones educativas buscan apoyar este proyecto.

Conociendo que los consorcios manejan mucha mas información de diversas universidades, se procedió analizar el resto para concluir en el formato estándar de la información que se detallada y obtener una idea general de la estructuración. Para ello se deben analizar varios factores como plataformas utilizadas, información expuesta, formato de información, lenguajes, segmentación de áreas de conocimiento, etc.

2.2.2. Análisis técnico y descripción del dominio: OERs dentro de los OCW.

2.2.2.1. Análisis Técnico del Domino

2.2.2.1.1. Plataformas OCW

Al referirse a un análisis técnico de desarrollo se esta hablando del uso de plataformas que los diferentes sitios OCW han empleado para la exposición de su información; conociendo la estructura de la información expuesta en dichas plataformas y la facilidad ó dificultad de extracción que permiten.

Tabla 4. Plataformas usadas en los diferentes sitios OCW

PLATAFORMAS	CONSORCIOS			
	OCW CONSORTIUM	UNIVERSIA	CONSORCIOS OCCIDENTALES	OTROS
EDUCOMMONS	✓	✓		✓
DRUPAL	✓	✓		
MOODLE	✓		✓	✓
WORDPRESS				✓
HTML	✓		✓	✓
DESARROLLOS PROPIOS	✓		✓	

Autor: Andrea Novillo

2.2.2.1.2. Lenguajes en OCW

Es importante conocer como la información esta expuesta en estos sitios, el saber que lenguaje esta utilizando cada sitio OCW para exponer su información y tomado en cuenta lo que se va a almacenar; genera investigar que codificación de caracteres va a ser empleado en la base de datos para generar el almacén de información. A continuación se detallan algunos lenguajes empleados en estos sitios:

Tabla 5. Lenguajes empleados en cursos OCW

LENGUAJES	CONSORCIOS			
	OCW CONSORTIUM	UNIVERSIA	CONSORCIOS OCCIDENTALES	OTROS
INGLES	✓	✓	✓	✓
ESPAÑOL	✓	✓		
CATALAN		✓		
PORTUGUES		✓		
JAPONES	✓		✓	
CHINO	✓		✓	
COREANO	✓		✓	
ARABE	✓		✓	
VIETNAMITA	✓		✓	

Autor: Andrea Novillo

2.2.2.2. Estructura de sitios OCW

La estructura de cada sitio por lo general viene dada por la plataforma empleada y diseño ó estructura que cada sitio emplee para exponer su información. Analizando y observando la mayoría de sitios OCW y los datos contenidos se puede exponer las siguientes estructuras de información.

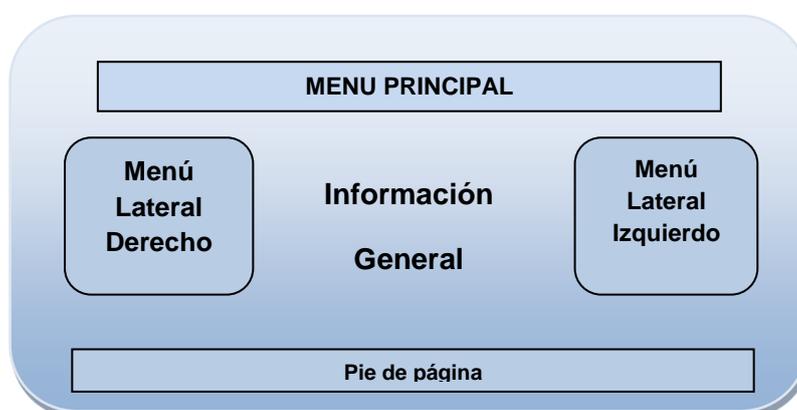


Figura 17. Estructura general de los sitios OCW

Autor: Andrea Novillo.

La **Figura 17** muestra una estructura generalizada de los diferentes sitios OCW, a continuación se detalla cada uno de los componentes que conforman dicho sitios; conociendo además en donde exactamente se encuentra la información que se necesita extraer para el presente proyecto.

a. **Menú Principal Superior:** La mayoría de los sitios cuentan con este tipo de menú para ayudar al usuario a ubicar la información correspondiente como es: Información del sitio OCW (Home), Listado de Cursos ó directorio actual de la navegación.

Cabe indicar que este módulo desaparece al ingresar a los respectivos cursos.

b. **Menú Lateral Derecho:** Se muestran las categorías en la que los cursos han sido clasificadas como Áreas de Conocimiento. Además dentro de la portada de los cursos muestra el menú ó contenido de estos. Se puede detallar a continuación:

▪ Áreas de Conocimiento: A continuación se detallan algunas áreas de conocimiento que emplean los sitios analizados, para ello se ha clasificado la información en cuatro áreas generales. Cabe recalcar que la información que se muestra en estas áreas de conocimiento depende de las necesidades de la universidad ó en su defecto la estructura curricular que maneja.

- Área Técnica
 - Termodinámica.
 - Computación.
 - Sistemas geográficos.
 - Inteligencia artificial.
 - Física, etc
- Área Biológica.
 - Ciencias
 - Química.
 - Ciencias de la Salud, etc.
- Área Humanística.
 - Ciencias Humanas.
 - Ciencias de derecho.
 - Educación física o deportiva, etc.
- Administrativa.
 - Procesos.
 - Banca.

- Administración general, etc.

- Menú o Contenido del Curso: Aportan el programa de trabajo al curso por medio de archivos en HTML, documentos en varios formatos, Videos e imágenes. A continuación algunas denominaciones del contenido:
 - Home del Curso
 - Programa o Syllabus
 - Recursos
 - Bibliografía, etc.

- c. **Menú lateral izquierdo**: Este módulo muestra sitios de interés ó publicidad, en algunos casos por el uso de plataformas como Moodle, HTML, desarrollos propios tiende a cambiar la información contenida y actúa como el modulo lateral derecho.

- d. **Información general**: Como su nombre lo describe presenta información principalmente del menú superior, cuando se ingresa a las áreas de conocimiento ó cursos muestra el detalle de información de estos.

- e. **Pie de pagina**: Describe información del sitio OCW como es dirección, licencia de uso en algunas ocasiones muestra la plataforma bajo la cual ha sido publicada su información.

A continuación podemos observar un ejemplo real de la estructura de información del sitio OCW de la Universidad Tecnica Particular de Loja (<http://ocw.utpl.edu.ec/>).

Guía de Aprendizaje

Contenidos del curso

- Cálculo I
- Home
- Información del Curso
- Programa de Contenidos
- Guía de Aprendizaje
- Material de Estudio
- Evaluaciones
- Bibliografía
- Equipo Docente

Menú Curso

Guía de Aprendizaje

Unidad 1: Funciones y límites

Unidad 2: Derivada

Unidad 3: Aplicaciones adicionales a las derivadas

Unidad 4: Funciones exponenciales y logarítmicas.

OBJETIVOS DE APRENDIZAJE	CONTENIDOS Unidades/Temas	CRONOGRAMA ORIENTATIVOS Tiempo Estimado	ACTIVIDADES DE APRENDIZAJE	RECURSOS DIDÁCTICOS	EVALUACIÓN
1. Identificar las diferentes clases de funciones y aplicar sus propiedades. 2. Examinar la función de límite y continuidad de una función.	<p style="color: blue; font-weight: bold;">Unidad 1: Funciones y límites</p> <ul style="list-style-type: none"> ■ Funciones ■ Modelos Funcionales ■ Límites 	<p style="color: blue; font-weight: bold;">Semana 1, 2 y 3</p> 12 horas de autoestudio 12 de interacción.	<ul style="list-style-type: none"> ■ Lectura comprensiva de los conceptos, nomenclatura y teorema del texto base. ■ Resolución de ejercicios planteados. 	<ul style="list-style-type: none"> ■ Material de Estudio ■ Bibliografía Básica ■ Texto básico ■ Guía Didáctica ■ Lectura 	Desarrollo de las autoevaluaciones 1-3 Evaluaciones a distancia: <ul style="list-style-type: none"> ■ Parte Objetiva ■ Parte de

Figura 18. Estructura de Información Contenido y OERs.
 Fuente: <http://ocw.utpl.edu.ec/>

2.2.2.3. Formatos de archivos OERs

Los documentos encontrados en los diferentes cursos, cuentan con variados formatos entre los principales se encuentran:

.jqz; .doc; .pdf; .zip; .ppt; .jsp; .pptx; .avi; .html; .shtml; .pps; .dfw; .mth; .swf; .flv; .htm; .odt

2.2.3. Descripción del Dominio OCW/OER.

El modelo que maneja los diferentes sitios OCW esta centrado en la exposición libre de su información educativa. La segmentación ó estructura de cada uno de ellos esta generalizada por cursos, dentro de los cuales contienen un menú de información como son Syllabus, Introducción, Autores, Bibliografías, etc. La especificación del contenido que se encuentra en los sitios OCW pueden ser vistos de manera general en el **anexo 4**.

Conociendo las fuentes de datos de donde se va a obtener la información necesaria para enriquecer y generar la base de datos OCW/OER; es necesario analizar y clasificar la información a obtener, para ellos se explican algunos menús que mantienen según su estructura la mayoría de cursos OCW:

Generalidades: A esta característica la podemos denominar como la portada en general del curso en la que se muestra la siguiente información:

Portada: en la que se muestra información del curso como es Título, Autor, Área a la que pertenece, fecha de revisión o modificación, numero de créditos de la materia.

Recomendaciones y conocimientos previos recomendados: Esto viene dado por cada profesor según las necesidades de la materia.

Descripción general de la Asignatura: Es decir de que trata la materia.

Objetivos: Conocimientos y Capacidades: Dependen de la materia y de lo que el docente considere necesario.

Material Docente: Es la descripción de lo que se utilizará en el curso.

Tareas Prácticas: Se le puede denominar al área de ejercicios prácticos a realizar.

Actividades de Evaluación: Consta de evaluaciones, exámenes, etc para conocer el nivel del alumno luego de la instrucción brindada.

Bibliografía/Bibliography: Muestra las fuentes de información que se consultaron para la creación del curso.

Lecturas Obligatorias: Son lecturas que aunque no se mencionaron en el programa general de estudio, afianzarán el conocimiento del interesado.

Lecturas Complementarias: Ayudarán a enriquecer el conocimiento del interesado.

Ejercicios, proyectos y caso, Prácticas: Demostración de algunos casos de Pruebas, Presentación de algunos proyectos y ejercicios para incrementar habilidades.

Pruebas de Evaluación, Exámenes que pueden ser por capítulo o finales, dependiendo de lo que explique la guía de estudio.

Software: Almacena algunos enlaces de software en los que los ejercicios o tareas podrían realizarse.

Guías de aprendizaje, Plan de Trabajo, Programa Syllabus: Generan el encapsulamiento del trabajo que van desarrollar los interesados, estipulando el tiempo de las diferentes tareas en el transcurso del curso, además muestra información de los temas a estudiar dentro de cada curso.

Notas de Curso: Información en resumen de lo visto por clase ó capítulo. Citando aspectos importantes del Curso.

Laboratorios: Similares ejercicios.

Material de Clases: Documentación pertinente de cada curso en diferentes formatos. Principalmente aquí se encuentran almacenados los OERs.

Glosarios: Términos que ayudan a definir la información que no ha mencionado a fondo o también se pueden tratar como palabras claves.

Autores del material ó Información de profesores: Genera la información necesaria de los docentes que han generado los documentos, y deja como constancia datos personales por medio de los cuales algún usuario pueda consultar interrogantes con los profesores u autores.

De acuerdo a la información expuesta algunos almacenan Recursos educativos abiertos, como lo son archivos y documentos que aportan conocimiento individual al curso. Estos OERs son expuestos en diferentes formatos como se mencionó en el punto anterior.

2.2.4. Enriquecimiento de información.

En este punto se analizará el aporte de los datos verificados para el enriquecimiento de la información en una base de datos, saber con qué información se cuenta ayuda a identificar el grado de información que se quiere ofrecer a los usuarios finales.

En el estado de arte se analizó la calidad de OCW, los puntos o áreas que debería soportar un OCW con calidad de contenidos según (Piedra & Romero, 2012) es la presentación del usuario y la accesibilidad. Esto no solo identifica al sitio sino la calidad de información que se presente en estos.

El tener un sitio OCW en la Web no implica que estos al ser buscados aparezcan absolutamente todos, pues cuando el usuario necesita la información acerca de un tema en específico se vuelve tedioso buscar respuesta por respuesta que ofrece como resultado dicho buscador; en el probablemente la información almacenada es vaga y repetitiva. Para solucionar esta pérdida de tiempo del usuario es que se crea la BD_OCW, almacenando información unificada que genere una confianza de elección y comparación.

Tomando en consideración que ya existen consorcios que mantienen almacenada la información de los sitios OCW, y tomando en consideración el tiempo que el usuario pierde al buscar información; nace la inquietud de que información se debería extraer y como estructurarla en una base de datos para mantener la información de todos los sitios OCW sea unificada.

En la siguiente tabla se puede observar que información mantienen los consorcios en sus sitios Web acerca de los OCW y cual es la información que el presente proyecto

pretende obtener y unificar, almacenando la información en la base de datos a la cual se le ha denominado BD_OCWOER.

Tabla. 6 Estructura de información en base de datos OcwOer.

	Datos	Descripción
Datos Almacenados en la consorcios actualmente	Entidad Educativa	
	Nombre de Universidad	Presenta el enlace de la Universidad a la que pertenece el OCW Site.
	OCW	
	OCW_Site	Muestra el enlace que tiene el sitio OCW, de la universidad antes mencionada.
Datos necesarios para unificar en BD_OCWOER	Contenido del Curso	
	Cursos	Dentro de cada uno de los enlaces de sitios OCW se encuentra información de los distintos cursos de los cuales se obtendrán: <ul style="list-style-type: none"> Área de conocimiento. Nombre del curso.
	Menú	Es la estructura que maneja cada curso, para representar su contenido.
	OER	Son archivos que aportan conocimiento dentro de cada uno de los menús del curso.
	Metadatos OER	Es la información que mantiene cada archivo según el formato en el que se encuentra.

Autor: Andrea Novillo

2.3. Especificación de entorno.

Para conocer el dominio de trabajo e identificar clases y propiedades que construirán la estructura en la base de datos, se deben plantear algunas interrogantes como son:

- ¿Cuáles son la Áreas de conocimiento que mantienen los OCW?
- ¿Cuáles son los tipos de menús que existen en los diferentes cursos?
- ¿Que cursos se encuentran en los diferentes sitios OCW?
- ¿Qué Universidades mantienen un sitio OCW?
- ¿Cuáles son las licencias empleadas en los Sitios OCW?
- ¿Cuáles son los autores por Universidades?

- ¿Cuáles son los países con representación de sitios OCW?
- ¿Qué formatos mantienen los OERs?
- ¿Qué OERs existen?
- ¿Cuáles son los metadatos encontrados en los OERs?
- ¿Qué lenguajes se emplean en los OERs?
- ¿Qué Universidades contienen OERs?

2.4. Segmentación y categorización de Información.

En este punto se procederá a identificar conceptos en base al análisis de posibles proveedores de contenido Ocw como son:

Tabla 7. Proveedores de Contenido OCW

PROVEEDOR	CONTENIDO APORTADO
UNIVERSIA	Sitios OCW a nivel Latinoamericano.
OCWConsortium	Sitios OCW a nivel Global
Sitios OCW	Cursos, OER, Descripciones.

Autor: Andrea Novillo

Las interrogantes mencionadas anteriormente son aquellas que dan pie a seleccionar la información que se necesita en el presente proyecto siendo el contenido OER dentro de todos los OCW.

En cuanto al esquema de Información que se pretende obtener se han analizado algunos conceptos taxonómicos como se plantean en la imagen a continuación:

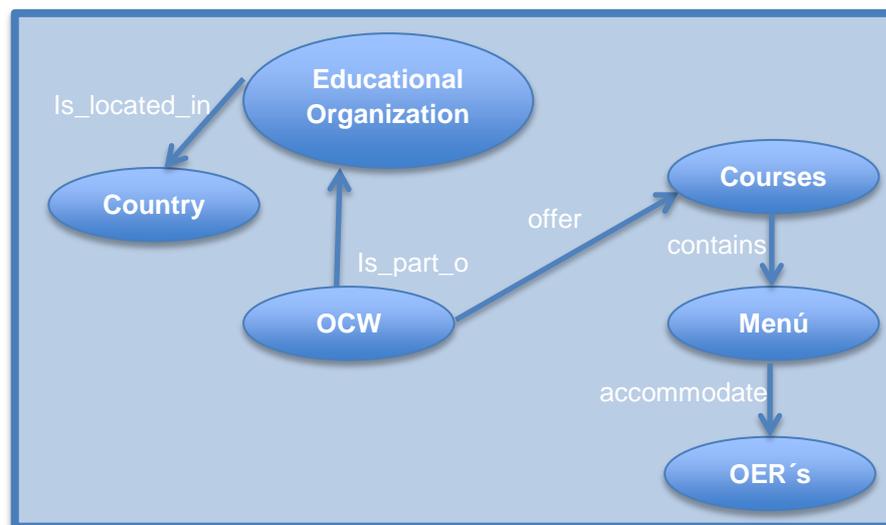


Figura 19. Taxonomía de la Información Ocw/Oer
 Autor: Andrea Novillo

2.5. Análisis de Vocabularios

Un vocabulario es la representación de un entorno por medio de propiedades y/o clases que expresan significado del contexto. Hay varios esquemas en RDF que definen conceptos de mundos reales tales como Educación, Geografía, Recursos electrónicos, Personas, etc.

Como se menciona en (Heath & Bizer, 2011) se deben considerar algunos criterios al momento de seleccionar los vocabularios para su reutilización, citando algunas preguntas como :

- **Uso y consumo:** ¿Este vocabulario es de uso generalizado?, ¿El uso de este vocabulario, establece un conjunto de datos más o menos accesible a las aplicaciones existentes de Linked Data?
- **Mantenimiento y gestión:** ¿Este vocabulario se mantiene actualizado de acuerdo a un claro proceso de gestión?, ¿Cuándo y en qué se basan las actualizaciones realizadas?
- **Cobertura:** ¿El vocabulario cubre la cantidad suficiente de datos para justificar la adopción de sus términos y compromisos ontológicos?
- **Expresividad:** ¿El grado de expresividad en el vocabulario es apropiado para el conjunto de datos y para el escenario de la aplicación?, ¿Es lo suficientemente expresivo o no?

Pues tomando en cuenta estos criterios, se dispuso a verificar que vocabularios respondían estas incógnitas y se acoplan a las necesidades del presente proyecto.

- **Dublin Core**

“Dublin Core Metadata Initiative” (DCMI), es un vocabulario de metadatos empleado para la definición de recursos. Entre las actividades que se desarrollan mediante este vocabulario están:

- Facilitar el acceso abierto a recursos educativos, capacitación y documentación relacionada con el diseño innovador y las mejores prácticas de metadatos.
- Apoyo a una comunidad mundial de personas que trabajan con metadatos para compartir experiencias y encontrar soluciones comunes a través de herramientas de colaboración, publicaciones y reuniones.
- Promover la cooperación y la interoperabilidad entre los estándares y vocabularios, mediante la participación de otras organizaciones y comunidades.
- Desarrollo y mantenimiento de los vocabularios de metadatos DCMI, y promover su uso en conjunto con otros vocabularios para describir los recursos.

Tabla 8. Vocabulario de Términos DCM.

Clases	Agent, AgentClass, BibliographicResource, FileFormat, Frequency, JURIsdiction, LicenseDocument, LinguisticSystem, Location, LocationPeriodOrJURIsdiction, MediaType, MediaTypeOrExtent, MethodOfAccrual, MethodOfInstruction, PeriodOfTime, PhysicalMedium, PhysicalResource, Policy, ProvenanceStatement, RightsStatement, SizeOrDuration, Standard
Propiedades	abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo, contributor, coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted, description, educationLevel, extent, format, hasFormat, hasPart, hasVersion, identifier, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, issued, isVersionOf, language, license, mediator, medium, modified, provenance, publisher, references, relation, replaces, requires, rights, rightsHolder, source, spatial, subject, tableOfContents, temporal, title, type, valid

Fuente: <http://dublincore.org/documents/dcmi-terms/>

- **FOAF**

“The Friend of a Friend” (FOAF), es una tecnología que busca simplificar compartir y utilizar información sobre personas y su actividades como fotos, calendarios, trabajos,

etc , permitiendo así transferir información entre sitios Web, y extender de forma automática, combinar y volver a utilizar en línea.

Tabla 9. Especificación de vocabulario FOAF

Clases	Agent , Document ,Group , Image , LabelProperty ,OnlineAccount ,OnlineChatAccount ,OnlineEcommerceAccount ,OnlineGamingAccount ,Organization , Person ,PersonalProfileDocument, Project
Propiedades	account ,accountName, accountServiceHomepage , age ,aimChatID , based_near , birthday , currentProject , depiction , depicts ,dnaChecksum , familyName ,family_name , firstName , focus ,fundedBy , geekcode , gender ,givenName , givename ,holdsAccount , homepage ,icqChatID , img , interest ,isPrimaryTopicOf , jabberID ,knows , lastName , logo , made ,maker , mbox , mbox_sha1sum ,member , membershipClass ,msnChatID , myersBriggs , name ,nick, openid , page , pastProject ,phone , plan , primaryTopic ,publications , schoolHomepage ,sha1 , skypeID , status , surname ,theme , thumbnail , tipjar , title ,topic , topic_interest , Weblog ,workInfoHomepage ,workplaceHomepage ,yahooChatID

Fuente: <http://xmlns.com/foaf/spec/>

- **Geographis Ontology - Geo**

“Basic Geo” se trata de un voabulario RDF que proporciona la comunidad Web Semántica con un espacio de nombres para la representacion geografica como son latitud y longitud.

- **Address**

“Talis Address Vocabulary”, es un vocabulario sencillo para representar direcciones.

Tabla 10. Esquema de Propiedades Address

PROPIEDAD	DESCRIPCION
countryName	Nombre del país mencionado.
Fax	Número de fax empleado por el recurso, persona, etc.
localityName	Provincia ó estado en que se encuentra un recurso, persona, institución, etc.
postalCode	Código postal del sitio en que se encuentra el recurso, persona, etc.
regionName	Nombre de región ó estado donde se encuentra el recurso, persona, etc.
streetAddress	Calles en que se encuentra ubicado el recurso, persona, etc.
Tel	Número telefónico del sitio, recurso, institución, etc,

Autor: Andrea Novillo

2.6. Planteamiento de clases y propiedades a reutilizar

Se especificaran los términos que son relevantes para el ámbito de trabajo de los OCW, para ello en base a los vocabularios citados se escogerán los conceptos que aporten sentido al ambiente de trabajo del presente proyecto.

A continuación se ilustran los términos que reutiliza el presente proyecto.

Tabla 11. Términos reutilizados de los diferentes vocabularios

Vocabulario	Metadato	Descripcion
FOAF	foaf:name	Representará el nombre de los sitios OCW. Nombre de recursos Educativos. Universidades y Centros educativos.
Dublin Core	dc:source	Enlace ó link del recurso OCW, universidad, OER, etc.
	dc:subject	Es la palabra clave o denominacion de tag que se encuentran dentro de los OCW.
	dc:description	Resumen o descripcion de la clase de la que se hablara como universidades, recursos, paises.
	dc:language	Lenguaje en el que fue descrito el recurso o sitio OCW.
	dc:creator	Es la persona o entidad que creo el recurso, curso, sitio Web, etc.
	dc:date	Periodo en que se creo el recurso.
	dc:title	Nombre q se le da al Curso, el tipo de menu y el OER.
	dc:type	Es el formato con que se manejan los OERs
	dc:contributor	Aquellos autores secundarios de los Cursos y OERs.
	dc:rights	Informacion de derechos de los Sitos OCW, Cursos y Recursos.
Geo	geo:lat	Latitud de los paises en que se encuentran situadas las Universidades.
	Geo:long	Longitud de los paises en que se encuentran situadas las Universidades.

Fuente: Andrea Novillo

2.7. Conclusiones

- El conocer el ambiente de trabajo en el que se basa el presente proyecto por medio de análisis técnicos de los diferentes sitios Ocw, como son plataformas ,lenguajes y

estructuras en que se encuentran desarrollados; facilita la visión general de que técnicas podemos emplear para obtener la información necesaria.

- Enfocándose en el primer punto, el reusó de los vocabularios existentes genera un vocabulario consensuado. Evitando así la duplicación de términos dentro de un mismo ambiente.
- Como primera instancia se obtiene una taxonomía inicial de los sitios Ocw, puesto que aun no se definen clases ni conceptos específicos

3. DESARROLLO DE FRAMEWORK PARA LA EXTRACCIÓN DE DATOS DESDE HTML

3.1. Extracción de Información.

El primer paso para crear un framework para la extracción de información, es conocer que el objetivo se basa en obtener datos que generen de alguna manera riqueza de información. Es por ello que durante el desarrollo de este proyecto se busca analizar las diferentes fuentes de información como son los sitios OCW.

Según el enfoque de los miembros del GICAC (Tovar, Piedra & otros, 2012) los datos OCW debe ser actualizados, confiables y además accesible sin ningún tipo de restricciones, puesto que la información que estos sitios mantienen son la respuesta directa a la desprivatización del conocimiento; además de ser difícil encontrar material que disponga la sociedad de las necesidades específicas de información. Siendo posible la interoperabilidad y reusabilidad de recuperar información sobre materiales educativos, implementando la Web Semántica.

Se debe diferenciar la extracción de datos, de minería de datos; puesto que la primera se refiere a obtener información de la web por medio de herramientas disponibles y la segunda es la inferencia sobre almacenes grandes de información analizando y generando conocimiento a partir de datos.

Por consiguiente la información debe ser extraída y no inferida directamente de los sitios OCW por medio de herramientas, librerías o tecnologías que trabajen bajo el enfoque de ser de código abierto, para mantenernos en la línea de trabajo de los OCW y generar un almacén de datos.

La extracción de información reside en obtener y analizar el código HTML, patrones de similitud la información deseada. La dificultad de este objetivo radica en que las páginas HTML no cuentan con una a estructura similar ni estandarizada.

3.1.1. Técnicas de Web scraping.

Se le conoce como *Web scrapy* a las técnicas utilizadas para extraer información de sitios Web. Se enfoca más en la transformación de datos sin estructura en la Web (formato HTML) en datos estructurados que pueden ser almacenados y analizados en una base de datos central.

3.1.1.1. JSOUP⁴²

Es una librería bajo el estándar de java. Trabaja con HTML proporcionando una extracción y manipulación de datos.

⁴² <http://jsoup.org/>

Características:

- Raspar y analizar código HTML de una URL, un archivo o cadena de
- Encontrar y extraer los datos, utilizando DOM de recorrido o de selectores CSS
- Manipular los elementos HTML, atributos y texto
- Limpia el contenido enviado por los usuarios en contra de una lista blanca segura, para evitar XSS ataques
- Salida ordenada HTML

3.1.1.2. Framework Scrapy⁴³

Esta es una técnica de scraping de alto nivel que no solo trabaja en base etiquetas HTML, sino en estructuras generalizadas propias de Python, además puede trabajarse en plataformas como Linux, Windows, Mac .Se puede utilizar para la extracción de datos, seguimiento y comprobación automática de la información.

Características

Simple

Scrapy fue diseñado pensando en la simplicidad, proporcionando las características que usted necesita sin ponerse en tu camino

Productivo

Sólo tiene que escribir las reglas para extraer los datos de las páginas Web y dejar Scrapy rastrear todo el sitio Web para usted

Rápido

Scrapy se utiliza en la producción de rastreadores para raspar por completo más de 500 sitios de proveedores de todos los días, todo ello en un servidor

Extensible

Scrapy fue diseñado pensando en la extensibilidad y por lo que ofrece varios mecanismos para conectar el nuevo código sin tener que tocar el núcleo marco

⁴³ <http://scrapy.org/>

Portátil

Scrapy se ejecuta en Linux, Windows, Mac y BSD

Open Source y el 100% desarrollado en el lenguaje de programación Python

Scrapy está completamente escrito en Python, lo que hace que sea muy fácil de hackear.

3.2. Expresiones regulares o patrones de extracción.

Para poder realizar la extracción y teniendo un previo conocimiento de que información se debe obtener de las páginas en formato HTML, se deberá generar estructuras o bien llamadas también expresiones regulares.

Estas expresiones regulares dependen de la estructura que mantienen las diferentes plataforma, empleadas en los por los sitios OCW, para ello se seguirá un proceso de obtención de información; según el análisis del código fuente se observara la ubicación del dato a obtener, como se muestra en la *figura 20*:



Figura 20. Análisis de información de código fuente.

Fuente: <http://ocw.utpl.edu.ec/>

Este punto se enfoca en el análisis de patrones de las diferentes estructuras encontradas a nivel de los cursos brindados en los OCW y sus contenidos denominados *Open Education Resources* (OER); los cuáles se extraerán mediante técnicas de extracción de la información requerida para almacenarla en la base de datos .A partir de

esta información múltiples aplicaciones serán desarrolladas para generar un conocimiento para estudiantes, profesores y autodidactas de todo el mundo; de forma específica se está trabajando en la recomendación de este tipo de recursos para acercar el conocimiento al público en general.

Una de las líneas de trabajo que citan (Tovar, Piedra & otros, 2012) se enfoca en mejorar la interoperabilidad, localización y consulta de Recursos Educativos Abiertos y de forma concreta cursos OpenCourseWare (OCW).

Para la identificación de patrones se ha trabajado con el complemento de Firefox como es *Firebug*⁴⁴, analizando cada una de las plataformas que mantienen los sitios OCW como son:

1. Educommons

Se ha mencionado en el capítulo anterior, que esta plataforma fue creada expresamente para adaptarla a sitios OCW para que la información sea estandarizada, por ello la obtención de patrones tiene un grado de facilidad mayor que las restantes, puesto de alguna manera la información se encuentra estandarizada.

A continuación se detallan algunas de las estructuras básicas obtenidas de algunos sitios OCW que mantiene esta plataforma:

Tabla 12. Patrones de expresiones de plataformas Educommons

Formato	Estructura	Interpretación
Formato 1	<code>(//div[@class="visualPadding"]/div/dl[@id="portlet-simple-nav"]/dd/a')</code>	La información requerida se encuentra en 2 niveles de etiquetas div y bajo el ámbito de la etiqueta dd.
Formato 2	<code>(//div[@class="visualPadding"]/div[@id="portlet-eduCommonsNavigation"]/div/a')</code>	La información requerida se encuentra bajo 3 niveles de etiquetas div
Formato 3	<code>(//div[@class="content"]/div[@class="item-list"]/ul[@class="course-navigation links inline"]/li/a')</code>	Trabaja con 2 niveles de etiquetas div este ultimo nivel contiene etiqueta ul y finalmente la información bajo la etiqueta li
Formato 4	<code>(//div[@id="main"]/div[@id="local_navigation"]/div[@c</code>	Es similar al formato 3 con la variante de

⁴⁴ <http://getfirebug.com/>

	<code>lass="get_started"]/div/ul/li/a')</code>	4 niveles de etiquetas div.
Formato 5	<code>('//td[@class="menu-principal"]/table/tr/td/a')</code>	La información requerida se encuentra en el interior de una tabla bajo el dominio de una etiqueta td
Formato 6	<code>('//div[@class="visualPadding"]/div/div/a')</code>	La información requerida se encuentra bajo 3 niveles de la etiqueta div.

Autor: Andrea Novillo

En la figura 21, se puede observar la estructura de los sitios OCW construidos bajo la plataforma en mención:

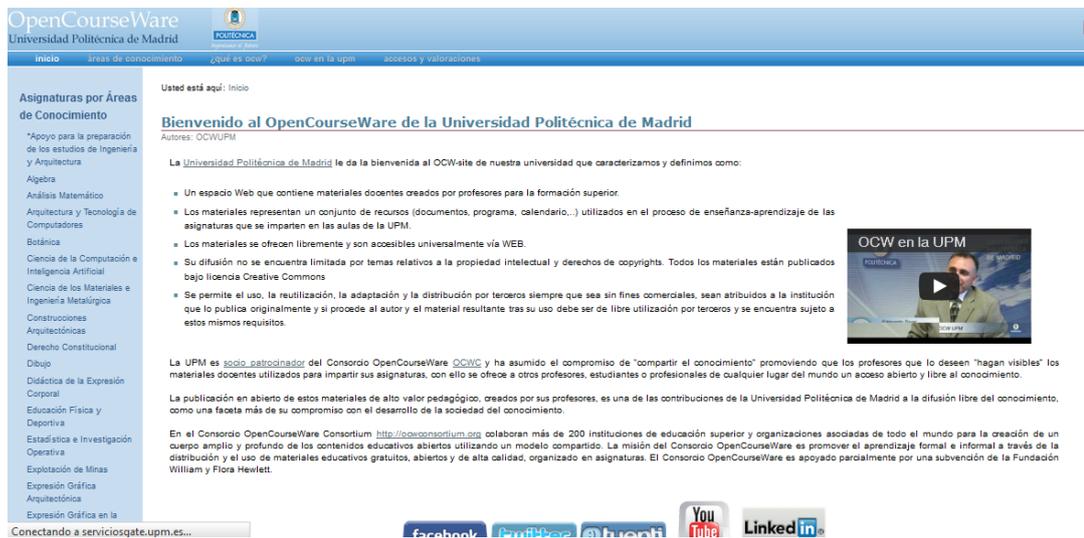


Figura. 21. Sitio creado bajo plataforma Educommons

Fuente : <http://ocw.upm.es/>

2. Moodle

La estructura con la que Moodle trabaja es generar por cada recurso un espacio en el que se pueda almacenar su contenido. En algunos sitios puede ser necesario obtener un registro de acceso como usuario, en general se tiene una clave de invitado para cumplir con el requisito de OCW.

A continuación se detallarán algunos de los sitios OCW que han empleado dicha plataformas y sus respectivas estructuras de almacenamiento de información:

a) ocw.ugr.es⁴⁵

“Es un centro de enseñanza virtual de la universidad de Granada.”

La información de cada curso esta almacenada en OCW.ugr.es/course cada uno de estos están incluidos en una categoría con un **Id** determinado.

La información en este sitio proporciona se encuentra estructurada por áreas de conocimiento, mediante el análisis del código se obtienen los siguientes patrones:

```
<Div class: info >  
  
  <ul class= name>.....este especifica el nombre de la materia  
  
  <ul class= teacher>.....especifica el nombre de los maestros
```

La información por curso y su contenido está estructurada de la siguiente manera:

```
<Div class: (nombre de la clase) >  
  
  <table class>  
  
    <tbody>  
  
      <tr class= content>...contienen un metadato denominado summary  
  
      <ul class= (nombre de clase)>  
  
        por cada curso y su contenido se genera :  
  
          <li> y un <href>
```

b) ocw.unex.es⁴⁶

Este sitio pertenece a la *Universidad de Extremadura* en España; a diferencia de la mayoría pues contiene las categorías en el lado derecho de la página y el contenido se encuentra estructurado en manera de árbol. No es necesario el registro de acceso, pero al igual que los anteriores tiene una cuenta de invitado.

Su estructura viene dada por:

⁴⁵ <http://ocw.ugr.es/>

⁴⁶ <http://ocw.unex.es/>

```

<li class= (expanded last)>..... por categorías
<href= /node/(mun del nodo)>
<h2>....nombre de la material
  <div class="content">
    <div id=(num del nodo)>
      <ul>
        <li>
          <href="user"(nodo de porfes)>

```

c) ocw.upct.es⁴⁷

Este sitio OCW pertenece a la *Universidad Politécnica de Cartagena*; contiene un RSS general por medio del cual se accede a los curso en general cuenta con un vocabulario y metadatos mas usados.

Contiene una cuenta de registro de Usuario pero no es necesario el logging. Internamente para ver el contenido del curso su estructura esta dado por:

```

<div class="content">
  <href=(pág del sitio)/course/view.php? id=(curso especifico)>

```

La mayoría de los cursos cuentan con categorías dadas por cada uno de los sitios, cuando tiene un RSS no genera el URL con el nombre de éste, sino la categoría en la que se encuentra.

⁴⁷ <http://ocw.bib.upct.es/>

Universidad Politécnica de Cartagena

OCW UPCT

CAMPUS MARE NOSTRUM
Viernes 18 Enero 2012

UPCT Admisión Estudios I + D + I Servicios Estudiantes

Categorías

- Escuela Arquitectura e Ing de la Edificación
- Escuela de Ingeniería de Caminos y de Minas
- ETS Ingeniería Agronómica
- ETS Ingeniería de Telecomunicación
- ETS Ingeniería Industrial
- ETS Ingeniería Naval y Océánica
- Facultad de Ciencias de la Empresa

Todos los cursos ...

Accesos directos

- Repositorio Digital UPCT
- PoliMedia
- Servicio de Documentación de la UPCT
- Universidad Politécnica de Cartagena

Información

- Preguntas frecuentes
- Contacto

Bienvenido al OpenCourseWare de la Universidad Politécnica de Cartagena

La Universidad Politécnica de Cartagena le da la bienvenida a su portal OCW UPCT cuyo principal objetivo es proporcionar acceso libre, sencillo y coherente a los materiales docentes elaborados por los profesores de nuestra universidad, con el fin de que puedan ser compartidos y utilizados por docentes y estudiantes de otras universidades o por cualquier persona que quiera acceder a ellos.

OCW es una iniciativa editorial electrónica a gran escala puesta en marcha en abril del 2001, basada en Internet y fundada conjuntamente por el Instituto Tecnológico de Massachusetts (MIT), la Fundación William and Flora Hewlett y la Fundación Andrew W. Mellon. OCW es coordinado en España y América Latina por Univerisat.

OCW es un ejemplo más de las iniciativas que en los últimos tiempos han emergido para promover el acceso libre y sin restricciones a la información, en este caso a los materiales docentes.

VI Edición de los Premios MECD - UNIVERSIA a la iniciativa OCW 2012

Universidad Politécnica de Cartagena

OPEN COURSEWARE CONSORTIUM

Open Course Ware

UNIVERSIA

DOS ASIGNATURAS DE LA UPCT OBTIENEN DOS ACCÉSITS EN LOS PREMIOS ME-UNIVERSIA A LA INICIATIVA OCW 2011

Abastecimiento de aguas de Francisco Javier Pérez de la Cruz, Juan García Bermejo y Mario Andrés Urrea Mallebrera (Departamento de Ingeniería Civil-Área de Ingeniería Hidráulica-Grado en Ingeniería Civil)

Química general de Beatriz Miguel Hernández, Mercedes Alacid Carceles, José Antonio Fernández López, Gerardo León Albert, M^a José Martínez García y Juan Ignacio Moreno Sánchez (Departamento de Ingeniería Química y Ambiental-Área de Ingeniería Química-Grado en Ingeniería en Tecnologías Industriales)

Fotos de la entrega de premios

Figura 22. Sitio OCW UPCT desarrollado en Moodle

Fuente: <http://ocw.upct.es>

3. HTML

Este tipo de plataforma presentó los mayores inconvenientes al momento de obtener una estructura similar, puesto que como no tiene una estandarización en sus diseños el análisis se debió generar un patrón ó expresión por cada uno de los sitios OCW que manejan esta estructura.

A continuación se detallan algunas estructuras obtenidas:

Tabla 13. Patrones de expresiones de plataformas HTML

Formato	Estructura
1	<code>(//td[@class="menú principal"]/table/tr/td/a')</code>
2	<code>(//div[@id="wrapper"]/div[@id=menu]/div[@id="inlineMenu"]/div[@class="grpMenuDespeglabe"]/li/div/a')</code>
3	<code>(//div[@class=main]/div[@class="left_course"]/ul[@class="course_categories"]/li/a')</code>
4	<code>(//div[@id="inst37"]/div[@class="content"]/ul[@class="list"]/li/div[@class="column c1"]/a')</code>
5	<code>(//div[@class="content"]/div[@class="item-list"]/ul[@class="course-navigation links inline"]/li/a')</code>
6	<code>(//div[@id="content"]/div[@id="contentLeft"]/div[@id="contentMenu"]/ul/li/ul/li/a')</code>
7	<code>(//div[@id="templatemo_left_column"]/div[@class="templatemo_leftmenu"]/a')</code>
8	<code>(//div[@class="left menu"]/div[@class="left_menu_item"]/a')</code>
9	<code>(//body/table/tr/td/p/a')</code>
10	<code>(//div[@id="block-OCW-0"]/div[@class="inner clearfix"]/div[@class="content clearfix"]/ul/li/a')</code>
11	<code>(//div[@class="insidenav"]/ul/li/a')</code>
12	<code>(//div[@class="viewa-fiel-body"]/div[@class=field-content]/ul/li/ul/li/span/a')</code>
13	<code>(//div[@id="tubular"]/div[@id="tubular2"]/a')</code>
14	<code>(//ul[@class="course_categories"]/li/a')</code>
15	<code>(//div[@id="sectionLinks"]/ul/li/div)</code>
16	<code>(//div[@class="contents"]/ul/li/a')</code>
17	<code>(//div[@class="viewa-fiel-body"]/div[@class=field-content]/ul/li/span/a')</code>
18	<code>(//div[@class="block_navigation"]/ul[@class="block_tree list"]/li[@class="type_structure depth_2 contains_branch"]/ul/li/p/a')</code>

Autor: Andrea Novillo

3.3. Proceso scraping para extracción de información.

El proceso de extracción scrapy es necesario para la obtención de información de los diferentes sitios OCW, en el cual se maneja de una manera ordenada cada paso para poder extraer y almacenar la información en una base de datos, que luego será consumida para trabajar bajo los principios de Linked Data.

a) A partir de los datos obtenidos de los cursos de OpenCourseWare, mediante el URL de estos se procede al análisis de las estructuras bajo el concepto verificación de menús y OERs.

- b)** A partir del primer punto se obtienen estructuras estándares y se procede al scraping de datos. El analizar la validez de las URL por medio de la librería TinyUrl ⁴⁸ evitará algunos conflictos de visualización.
- c)** Se vincularán los datos de tal manera que formen una base de datos (BD) estructurada en forma de tripletas, para la facilitación de lectura y formación de vocabularios RDF.
- d)** A partir de la información obtenida de los OERs y su estructura se construirá un agente de descarga que almacenará físicamente los documentos ó recursos en un directorio especificado por el usuario.
- e)** Luego de extraer los recursos en un solo lugar, el analizador de metadatos empezara a realizar su trabajo por medio de las librerías empleadas en el crawler.
- f)** Finalmente el conjunto de datos de las diferentes BD se organizará de tal forma que crezcan verticalmente y muestren una estructura de datos enlazados siendo Sujeto, Predicado y Objeto siendo esta una forma fácil de almacenamiento y acceso la información.

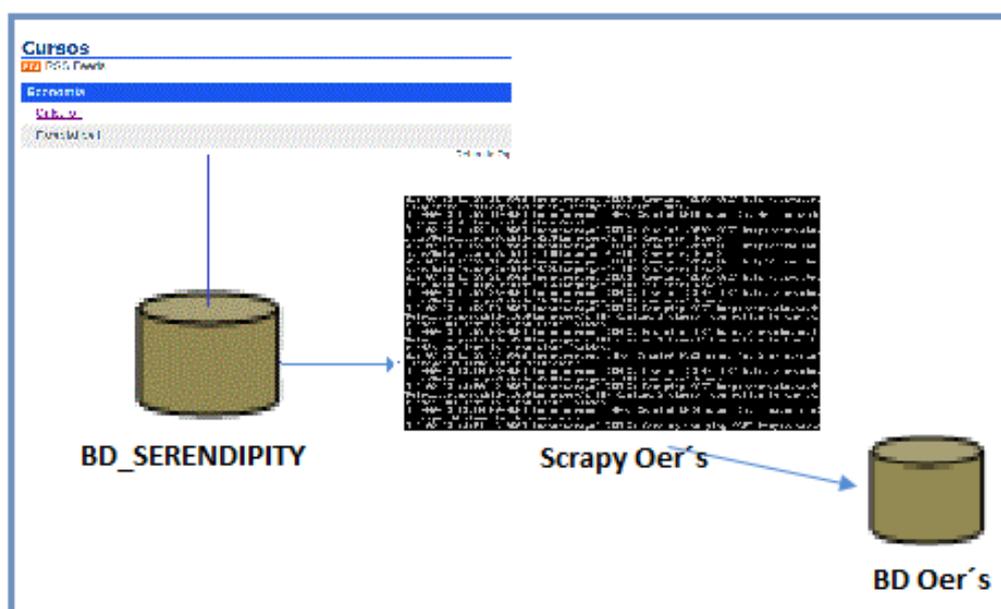


Figura.23. a). *Proceso de extracción de datos OCWOER*

Autor: Andrea Novillo

⁴⁸ <http://tinyurl.com/>



Figura 23. b). Proceso análisis de metadatos en OERs
 Autor: Andrea Novillo

3.3.1. Generación de Código Scrapy.

En la figura 23 a) del proceso scrapy se implementó el framework propio de Python, para la extracción de información por medio de estructuras o patrones regulares.

Framework Scrapy

Es un framework propio de Python que hace más fácil la extracción de información de páginas HTML, o también denomina Web scraping.

Para trabajar con este framework, se realizó la instalación de Scrapy en Windows para mayor información revisar **anexo 5**.

Codificación Scrapy

En este punto se observará la creación del script que emplea el framework Scrapy y la manera de obtener información de menús y contenidos OERs.

1. Previo a la extracción de información se deberá crear una base de datos (BD) para el almacenamiento de información.

```

def urls_oers(self, tabla):
    """
    ingreso a bd;
    """
    lista = []

    db=self.conectar()
    cursor=db.cursor()
    cursor.execute('Select Cursos from link_cursos;')
    datos = cursor.fetchall()
    for d in datos:
        lista.append(d[0])
    self.cerrar(db)
    return lista

def insertar_datos(self, sujeto, predicado, objeto):
    print "guardando datos"
    db=self.conectar()
    cursor=db.cursor()
    cursor.execute('INSERT INTO menu_cursos_1 (sujeto, predicado, objeto) VALUES (%s, %s, %s)'
        (sujeto,
        predicado,
        objeto))
    db.commit()
    self.cerrar(db)

```

Figura 24. Conexión BD en proceso Scrapy
 Autor: Andrea Novillo.

2. Se agregan los patrones de las diferentes plataformas empleadas, analizados y formados en el capítulo II para analizar los diferentes sitios OCW, asignándoles un **hxs** que es una llamada a la conexión de respuesta de un URL empleado para verificar si el sitio se encuentra activo ó no.

```

class tipomenu(BaseSpider):
    name = "menuscursos"
    b = BDdatos()
    start_urls = b.urls_oers("")
    tinyurl.create_one(start_urls)
    def parse(self, response):
        diccionario = {}
        hxs = HtmlXPathSelector(response)
        #Patrones de Formatos
        educomons_f1 = hxs.select('//div[@class="visualPadding"]/div/dl[@id="portlet-simple-nav"]/dd/a')
        formato_educomons_f1 = 'formato1'
        educomons_f2 = hxs.select('//div[@class="visualPadding"]/div[@id="portlet-eduCommonsNavigation"]/div/a')
        formato_educomons_f2 = 'formato2'
        educomons_f3 = hxs.select('//div[@class="content"]/div[@class="item-list"]/ul[@class="course-navigation 1']
        formato_educomons_f3 = 'formato3'
        educomons_f4 = hxs.select('//div[@id="main"]/div[@id="local_navigation"]/div[@class="get_started"]/div/ul
        formato_educomons_f4 = 'formato4'
        educomons_f5= hxs.select('//td[@class="menu-principal"]/table/tr/td/a')
        formato_educomons_f5 = 'formato5'
        educomons_f6 = hxs.select('//div[@class="visualPadding"]/div/div/a')
        formato_educomons_f6 = 'formato6'
        #Moodle
        moodle_m1 = hxs.select('//div[@id="inst37"]/div[@class="content"]/ul[@class="List"]/li/div[@class="column
        formato_moodle_m1='formato_moodle1'
        moodle_m2 = hxs.select('//td[@id="left-column"]/div/div[@id="inst40"]/div[@class="content"]/ul/li/div/a')
        formato_moodle_m2='formato_moodle2'
        moodle_m3 = hxs.select('//div[@class="i3"]/table/tr/td/ul/li/a')
        formato_moodle_m3='formato_moodle3'

```

Figura 25. Estructuras o expresiones regulares en script
 Autor:Andrea Novillo

3. Se realiza el análisis de las URL y la extracción de información, al comparar los diferentes patrones tratando de obtener los enlaces y el texto que le acompaña.

```

diccionario[url] = []
if len(educomons_f1)>0:
    for datos in educomons_f1:
        if len(datos.select("@href").extract())>0 and len(datos.select("text()").extract())>0:
            texto=datos.select("text()").extract()[0]
            enlace=datos.select("@href").extract()[0]
            diccionario[url].append(("tipo_menu", texto))
            diccionario[url].append(("link", enlace))
    else:
        if len(educomons_f2)>0:
            for datos in educomons_f2:
                if len(datos.select("@href").extract())>0 and len(datos.select("text()").extract())>0:
                    texto=datos.select("text()").extract()[0]
                    enlace=datos.select("@href").extract()[0]
                    diccionario[url].append(("tipo_menu", texto))
                    diccionario[url].append(("link", enlace))
            else:
                if len(educomons_f3)>0:
                    for datos in educomons_f3:
                        if len(datos.select("@href").extract())>0 and len(datos.select("text()").extract())>0:
                            texto=datos.select("text()").extract()[0]
                            enlace=datos.select("@href").extract()[0]
                            diccionario[url].append(("tipo_menu", texto))
                            diccionario[url].append(("link", enlace))

```

Figura 26. Análisis de enlaces según estructuras.

Autor: Andrea Novillo

4. Luego del análisis de cada URL a través de cada patrón, la información extraída se almacenará en la BD previamente creada.

```

for d in diccionario.keys():
    if diccionario[d].__class__ == list:
        for valor in diccionario[d]:
            predicado = valor[0]
            objeto = valor[1]
            self.b.insertar_datos(url, predicado, objeto)
    else:

```

Figura 27. Almacenamiento en BD

Autor: Andrea Novillo

Cuando finalice el proceso de extracción de datos de OERs, inician otras fases como son el pre-procesamiento y la limpieza de datos, de manera los datos pueden ser utilizados por diferentes aplicaciones, enfocadas a dar mayor visibilidad y uso de este tipo de recursos; pudiendo aplicarse como siguiente paso el uso de los principios de Linked Data para enlazar y reusar datos OER a través de Comunidades Web.

3.3.2. Limpieza de Datos.

El obtener los datos por medio de técnicas de extracción, no significa que la información es totalmente completa y fiable; por lo cual se realizará el análisis de la información obtenida, dicho análisis comprende verificar la información y compararla con el sitio Ocw de donde se consiguió, y si es necesario completar la información por medio de scripts en la BD, dejándolos listos para el próximo proceso.

A continuación se pueden citas casos de información no válida:

- URLs acortadas, es decir que al empezar de esta no se tenía la parte inicial de un sitio e identificativa de un url en la web, como es `http:// [.]`.
- Información repetida.
- URLs que no pertenecen al ciclo de extracción que se realizaba, Ejm un primer ciclo era extraer los cursos de un lugar tomando en cuenta su URL y el texto; en ocasiones y por el número de patrones la información se repetía ó pertenecía a otra parte de la página. Por lo que se procedió a refinar las estructuras que mantenían este caso.

3.4. Extracción de metadatos de OER.

Extraer la información de los diferentes documentos denominados OERs, ayudará a completar la información de la BD.

Un crawler es una pieza pequeña de software, la cual trabaja directamente en internet tratando de tomar información y llevarla a un servidor.

El crawler que se presenta en el proyecto emplea librerías que extraen y analizan los metadatos ó información de recursos ó archivos con diferentes extensiones, como podemos denotar a continuación:

PDFTextStream⁴⁹: Esta librería fue implementada para la lectura de los metadatos que contiene un archivo ó recurso con extensión pdf.

JSOUP⁵⁰: Librería empleada para analizar metadatos de recursos en formatos HTML.

3.4.1. Creación de Crawler para extracción de Metadatos.

Existen varias herramientas y servicios en línea de extracción de metadatos, que contradicen al enfoque de open source; puesto que algunas son pagadas y otras en su defecto se adueñan de la información obtenida. Pensando en que la información aumenta constantemente y se necesita alimentar con nuevos metadatos, se crea un propio crawler de extracción de metadatos en base a la implementación de librerías como jsoup, PDFxStream entre otras,

El analizador funciona de la siguiente manera:

⁴⁹ <http://snowtide.com/>

⁵⁰ <http://jsoup.org/>



Figura 28. Interfaz de analizador de metadatos

Autor: Andrea Novillo

1. De la base de datos en la que se almacenó la información de los OERs, se extrae el URL de cada uno y se genera la descarga de estos en una dirección del computador.

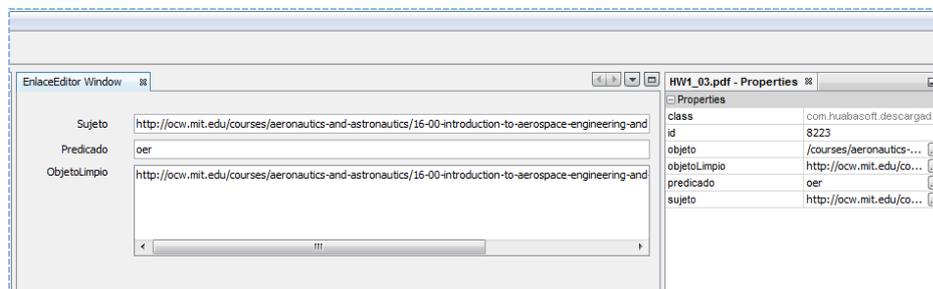


Figura 29. Descarga de archivos u OERs

Autor: Andrea Novillo

2. Por cada recurso descargado se procede a analizar mediante las librerías del crawler. Obteniendo metadatos como: autores, fechas, tipo de extensiones, etc.

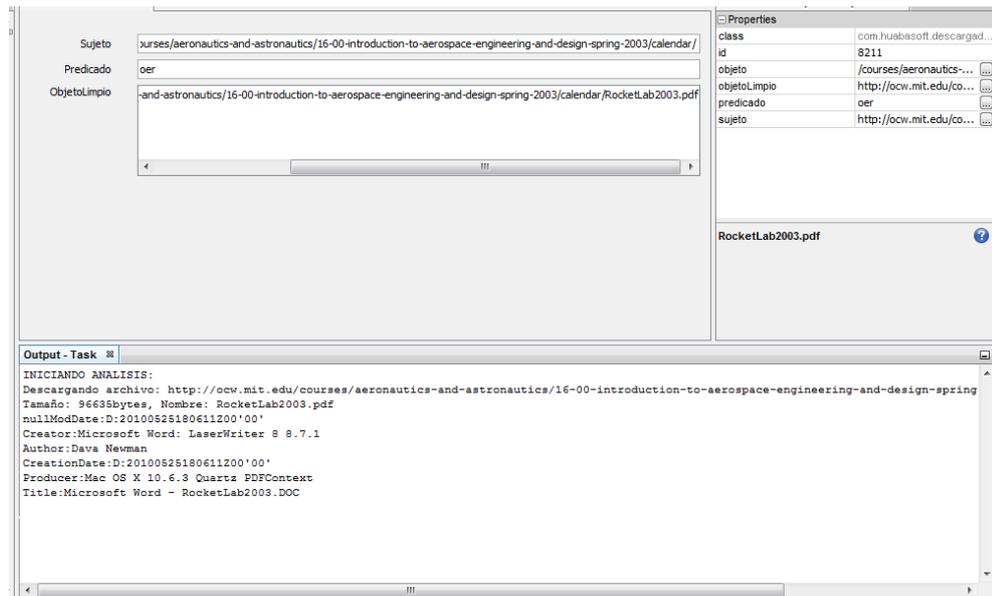


Figura.30. Análisis de metadatos

Autor: Andrea Novillo

3. Finalmente luego de obtener los metadatos estos serán almacenados en la base de datos en que se almaceno previamente la información de los Oers.

	MetadatoOer
	Author
	CreationDate
	Producer
	Title
	SourceModified
	Company
	Subject
	CreationDate--Text
	generator
	X-UA-Compatible
	PTEX.Fullbanner
	Trapped
	MTWinEqns
▶	Version
	_AuthorEmail

Figura 31. Metadatos extraídos

Autor: Andrea Novillo

3.5. Conclusiones:

- El uso de Web scrapy tiene como objetivo general extraer información estructurada de páginas Web, lo cual aporta a nuestro proyecto con el ahorro de tiempo y recurso

humano; puesto que se pueden automatizar dichos procesos de extracción en base a expresiones regulares que mantienen las estructuras de los diferentes sitios OCW.

- El uso de librerías múltiples para la extracción de metadatos, que contienen los diferentes formatos ó extensiones de los OERs aporta información específica y clave de cada uno de los recursos obtenidos.

4. ESTANDARIZACIÓN DE DATOS PARA PROCESO DE PUBLICACIÓN EN EL AMBIENTE OCW/OER

Luego de extraer la información necesaria y conociendo a fondo el ambiente de trabajo, se procede a estandarizar los datos para su respectiva publicación, siguiendo el proceso planteado por (Corcho&Gomez, 2010).

4.1. Definición de Vocabulario.

En el capítulo II, se verificaron algunos vocabularios que aportarían información y contextualización en nuestro ambiente de trabajo. Hay aspectos que son considerados como buenas prácticas que se toman en cuenta para definir la creación de un vocabulario consensuado según lo menciona (Heath y otros, 2011):

- Complementar los vocabularios existentes en lugar de reinventar sus términos.
- Sólo definir nuevos términos en un dominio que usted controle.
- Utilice RDFS y OWL para relacionar los nuevos términos con los vocabularios existentes.
- Aplique los principios de Linked Data, equitativa y rigurosamente para los vocabularios, como para los datasets, es decir, la URI de los términos debe ser desreferenciable para que las aplicaciones de Linked Data puedan buscar su definición.
- Documentar cada término nuevo con etiquetas y comentarios entendibles para el ser humano, *RDFs: label* y *RDFs: comment* están diseñados para este propósito.
- Sólo definir las cosas que importan, por ejemplo, la definición de dominios y rangos ayuda a clarificar cómo las propiedades deben ser utilizadas, pero el exceso de la especificación de un vocabulario también puede producir conclusiones inesperadas cuando los datos son consumidos. Por lo tanto es mejor definir los términos ligeramente.

A partir de estas consideraciones se procede a desarrollar un modelo conceptual del proyecto; verificando así que vocabularios y términos se emplearán para el presente proyecto.

Cumpliendo con el principio de reutilización de vocabularios a continuación se puede observar que vocabularios y términos, serán empleados en el desarrollo del vocabulario denominado OcwOer.

Tabla 14. Términos de Vocabularios empleados

Vocabulario	Término	Descripción
Dublin Core	dc:Description	En el contexto del proyecto es empleado este termino para generar un a descripción de alguna clase.
Dublin Core	dc:source	Especifica el enlace ó URL que contiene una clase.
FOAF	foaf:name	Empleado para declarar el nombre de una clase.
VCard	VCARD:street	Utilizado para generar información de una dirección, en este proyecto es empleado específicamente para definir la ubicación de una clase mediante calles.
GEO	geo:latitude	Empleado para generar la situación geográfica de una clase. Especificando la latitud de esta.
GEO	geo:longitude	Empleado para generar la situación geográfica de una clase. Especificando la longitud de esta.
RSS	rss:link	Empleado para representar un enlace de los RSS que mantiene cada sitio OCW.

Autor: Andrea Novillo

Cada una de las especificaciones que se emplean en el vocabulario representa la relación que contiene el ambiente de trabajo, tomando como punto general la Organización Educacional a la que pertenece cierto sitio OCW, además la situación geográfica a la que pertenece dicha institución, como al contenido que se encuentra dentro de los diferentes sitios hasta culminar en los distintos OERs existentes y sus características.

En base al modelo del ambiente de trabajo y tomando en consideración los conceptos que marcan importancia se muestran de manera general los conceptos empleados a implementarse en nuestro vocabulario. Para ver a detalle cada uno de estos conceptos puede revisarse el **anexo 1**.

Tabla 15. Especificación de clases y propiedades del vocabulario OcwOer

TÉRMINO	TIPO
EducationalOrganization	Clase
City	Clase
Country	Clase
Continent	Clase
Consortium	Clase
OCWSite	Clase
Course	Clase
KnowledArea	Clase
Menu	Clase
OER	Clase
Plataform	Propiedad
Offer	Propiedad
organized_by	Propiedad
Stored	Propiedad
Provides	Propiedad
Accomodate	Propiedad
classified_by	Propiedad
Licenselink	Propiedad
is_locate_in	Propiedad
corresponds_to	Propiedad
belonging_to	Propiedad
is_member_of	Propiedad
Contains	Propiedad
Dbpedialink	Propiedad
Licensename	Propiedad
dbepediaLanguage	Propiedad

language_alternative	Propiedad
State	Propiedad
dbpediaCapital	Propiedad
Capital	Propiedad
Logo	Propiedad
metadata_type	Propiedad
metadata_value	Propiedad

Autor: Andrea Novillo

4.2. Creación de Ontología en Jena.

Luego de analizar el modelo conceptual, finalizada la búsqueda y creación de vocabulario adaptable al presente trabajo, se procede a crear una ontología en CmapTools Coe⁵¹ denominada OcwOer cuyo propósito es tener un enfoque del ambiente de trabajo en el que se recopilan clases y términos se emplearían para representar los sitios OCW y los OERs. Esta creación se realiza a manera únicamente gráfica ya que su implementación se la hace en Jena.

Se trabajo directamente en la creación de la ontología sobre la librería Jena, pues facilita el agregar individuos y transformar la información directamente en formato RDF.

4.2.1. Desarrollo de Schema.

A continuación se detalla como se ha estructurado el schema del entorno de trabajo en base a la librería Jena:

1. La especificación de URI's para cada una de los conceptos es uno de los principios de Linked Data. Por ello se procede a crear un prefijo para este vocabulario como se muestra a continuación.

⁵¹ CmapTools Coe (<http://cmap.ihmc.us>)

```

OntModel model = ModelFactory.createOntologyModel(OntModelSpec.OWL_MEM);
String NS = "ocw";
model.setNsPrefix(NS, "http://www.utpl.edu.ec/oervocab#");

```

Figura 33. Creación de prefijo para vocabulario OCW/OER

Autor: Andrea Novillo

2. Luego se procede a crear las respectivas clases y propiedades.

```

OntClass Country = model.createClass(NS+ ":"+"Country");
OntClass Continent = model.createClass(NS+ ":"+"Continent");
OntClass Consortium = model.createClass(NS+ ":"+"Consortium");
OntClass OcwSite = model.createClass(NS+ ":"+"OcwSite");
OntClass Course = model.createClass(NS+ ":"+"Course");
OntClass KnowledArea = model.createClass(NS+ ":"+"KnowledArea");
OntClass Menu= model.createClass(NS+ ":"+"Menu");
OntClass Oer = model.createClass(NS+ ":"+"Oer");

ObjectProperty is_locate_in= model.createObjectProperty(NS+ ":"+"is_locate_in");
ObjectProperty belonging_to=model.createObjectProperty(NS+ ":"+"belonging_to");
ObjectProperty is_member_of=model.createObjectProperty(NS+ ":"+"is_member_of");
ObjectProperty is_located=model.createObjectProperty(NS+ ":"+"is_located");
ObjectProperty corresponds_to=model.createObjectProperty(NS+ ":"+"corresponds_to");
ObjectProperty contains=model.createObjectProperty(NS+ ":"+"contains");
ObjectProperty contains_course=model.createObjectProperty(NS+ ":"+"contains_course");
ObjectProperty has=model.createObjectProperty(NS+ ":"+"has");

is_locate_in.addProperty(RDFS.domain, Organization);
is_locate_in.addProperty(RDFS.range, City);
belonging_to.addDomain(OcwSite);
belonging_to.addRange(Organization);
is_member_of.addDomain(OcwSite);
is_member_of.addRange(Consortium);
is_located.addDomain(Consortium);
is_located.addRange(Country);
corresponds_to.addDomain(Country);
corresponds_to.addRange(Continent);
contains.addDomain(Course);
contains.addRange(Oer);
contains_course.addDomain(OcwSite);

```

Figura 34. Creación de Clases y Propiedades – Vocabulario OCW/OER

Autor: Andrea Novillo

Para poder observar la ontología creada a mayor detalle del presente proyecto puede dirigirse al **anexo 2**, en el **anexo 3** se podrá revisar el schema del vocabulario creado en formato RDFs.

4.2.2. Verificación Schema OcwOer.

Existen gran cantidad de herramientas que verifican la sintaxis de un archivo RDF. En este caso se utiliza el Validador de la W3C⁵² es de fácil acceso y uso, valida archivos en formatos RDF/XML representando su información en tripletas y grafico presentándose de la siguiente manera:

⁵² Validador W3C <http://www.w3.org/RDF/Validator/>

Triples of the Data Model			
Number	Subject	Predicate	Object
1	ocw:accommodate	http://www.w3.org/2000/01/rdf-schema#range	ocw:Oer
2	ocw:accommodate	http://www.w3.org/2000/01/rdf-schema#domain	ocw:Menu
3	ocw:accommodate	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#ObjectProperty
4	ocw:is_locate_in	http://www.w3.org/2000/01/rdf-schema#range	ocw:City
5	ocw:is_locate_in	http://www.w3.org/2000/01/rdf-schema#domain	ocw:EducationalOrganization
6	ocw:is_locate_in	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#ObjectProperty
7	ocw:stored	http://www.w3.org/2000/01/rdf-schema#range	ocw:Course
8	ocw:stored	http://www.w3.org/2000/01/rdf-schema#domain	ocw:KnowledArea
9	ocw:stored	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#ObjectProperty

Figura 35. Validación W3C – Tripletas del vocabulario Ocw/Oer

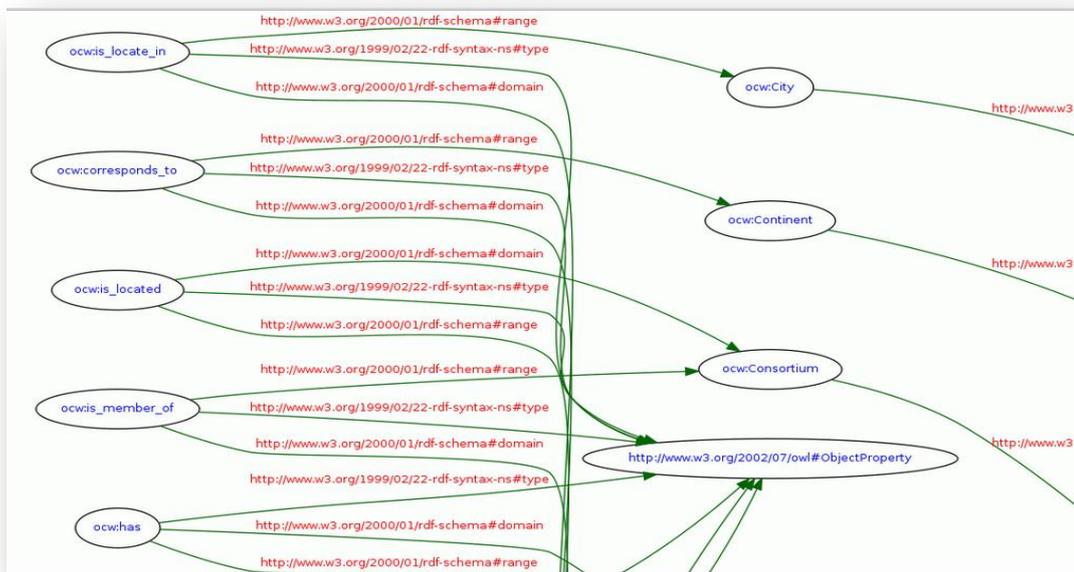


Figura 36. Validación W3C Grafo Ocw/Oer

Autor: Andrea Novillo

4.3. Conversión de Datos a RDF.

En la Fase III se realizó la extracción de información y se especificó que información se almacenaría en la base de datos. Para la conversión de estos datos en formato RDF/XML se emplea Jena; tomando en cuenta el vocabulario ya definido se procede a la carga de información desde la base de datos.

4.3.1. Carga de Datos.

Compartir la información, es lo que se busca bajo el enfoque de Linked Data, para ello la información del ambiente de trabajo deberá estar en formato RDF/XML y así poder continuar con el siguiente punto que es finalmente la publicación de información y las consultas para la solución de las incógnitas planteadas al inicio de este proyecto.

A continuación se explican los pasos para realizar la carga de información al schema OcwOer:

1. Se crea una conexión a la Base de Datos que contiene la información necesaria, la base de Datos se encuentra en un servidor denominado lord.utpl.edu.ec.

```
conexion mysql = new conexion();
if (mysql.Conectar() != null) {
    System.out.println("Ingresast");
} else {
    System.out.println("Conexion Incorrecta");
}

select = "SELECT DISTINCT UniversityName,UniversityLogo,UniversitySubjectEn,UniversityOCWURL,UniversityOCWRSS"
+ ",UniversityOCWPlataform,license,LicenseURI,UniversitySubjectEs,UniversitySubject,Language,UniversityURL, "
+ "UniversityAddress from cursosmetadatas3 gRoup By UniversityName ";
```

Figura 37. Conexión a base de datos Ocw/Oer
Fuente: Andrea Novillo

2. Se asignan variables para la extracción de la información, que son aquellos campos empleados en la base de datos y servirán para presentar la información que se desea obtener en formato RDF según lo diseñado en el vocabulario OcwOer a los cuales también se les denomina Individuos.

```

String uname = rs.getString("UniversityName");
String uurl = rs.getString("UniversityURL") + "\n";
String udesc = rs.getString("UniversityDescription") + "\n";
String cons = rs.getString("MemberOFConsortium") + "\n";
String urlcons = rs.getString("ConsortiumURL") + "\n";
String consdesc = rs.getString("ConsortiumDescription") + "\n";
String uocw = rs.getString("UniversityOCWURL") + "\n";
String uocwrss = rs.getString("UniversityOCWRSS") + "\n";
String ocwplat = rs.getString("UniversityOCWPlataform") + "\n";
String ulogo = rs.getString("UniversityLogo") + "\n";
String udbpedia = rs.getString("UniversityDBPEDIA") + "\n";
String usubj = rs.getString("UniversitySubject") + "\n";
String usuben = rs.getString("UniversitySubjectEn") + "\n";
String usubes = rs.getString("UniversitySubjectEs") + "\n";
String ulat = rs.getString("UniversityLat") + "\n";
String ulong = rs.getString("UniversityLong") + "\n";
String uadress = rs.getString("UniversityAddress") + "\n";
String knowledgeArea_o = rs.getString("KnowledgeArea_o") + "\n";
String knowledgeArea = rs.getString("KnowledgeArea") + "\n";
String tagc = rs.getString("Tags") + "\n";
String areac = rs.getString("Areas") + "\n";
String titlec = rs.getString("Title") + "\n";
String descriptionc = rs.getString("Description") + "\n";
String authores = rs.getString("Authores") + "\n";
String authoresc = rs.getString("Authoresc") + "\n";
String html = rs.getString("html") + "\n";
String imagen = rs.getString("image") + "\n";
String ucontinent = rs.getString("UniversityContinent") + "\n";
String ucountry = rs.getString("UniversityCountry") + "\n";
String ucity = rs.getString("UniversityCity");
String uCapitalDBPEDIA = rs.getString("UniversityCapitalDBPEDIA");
String uCountryDBPEDIA = rs.getString("UniversityCountryDBPEDIA");

```

Figura 38. Extracción de Información de base de datos OCW/OER
Autor: *Andrea Novillo*

4. En base al schema creado anteriormente se le asignarán los individuos, para ello se crean recursos que para nuestro caso serán las clases y se le asigna la información obtenida de la base de datos mediante las propiedades del vocabulario OcwOer. Finalmente la salida generará un archivo en formato RDF/XML el cual pasa a ser revisado en el validador W3C, este puede ser revisado en el **anexo 6**.

```

String course_id = Course + "/" + idc;
String city_id = City + "/" + ucity;
String country_id = Country + "/" + ucountry;
String continent_id = Continent + "/" + ucontinent;
String knowledarea_id = KnowledgeArea + "/" + knowledgeArea;

organizacion = model.createResource(organizacion_id)
    .addProperty(DC.source, uurl).addProperty(FOAF.name, uname).addProperty(ocw.street, address)
    .addProperty(DC.subject, usubj).addProperty(vocw.subject_en, usuben)
    .addProperty(vocw.subject_es, usubes).addProperty(vocw.url_img, ulogo)
    .addProperty(RDF.type, Organization);
consorcio = model.createResource(consorcio_id)
    .addProperty(FOAF.name, cons).addProperty(DC.source, urlcons)
    .addProperty(DC.description, consdesc).addProperty(RDF.type, Consortium);
ocwsite = model.createResource(ocwsite_id)
    .addProperty(FOAF.name, uname).addProperty(DC.source, uocw)
    .addProperty(vocw.license_uri, license_uri).addProperty(vocw.url_img, imagen)
    .addProperty(vocw.plataform, ocwplat).addProperty(RDF.type, OcwSite);
course = model.createResource(course_id)
    .addProperty(DC.identifier, idc)
    .addProperty(DC.title, titlec).addProperty(DC.source, linkc)
    .addProperty(DC.description, descriptionc).addProperty(vocw.author, authoresc)
    .addProperty(vocw.url_img, imagen).addProperty(DC.subject, tagc).addProperty(RDF.type, Course);
city = model.createResource(city_id)
    .addProperty(FOAF.name, ucity)
    .addProperty(DC.source, uCityDBPEDIA).addProperty(RDFS.type, City);
country = model.createResource(country_id)
    .addProperty(FOAF.name, ucountry).addProperty(DC.source, uCountryDBPEDIA)

```

Figura 39. Asignación de propiedades e individuos mediante vocabulario OcwOer.

Fuente: Andrea Novillo

4.4. Publicación de Datos.

El archivo generado en tripletas en el punto anterior es finalmente almacenado en una base de datos propia para RDF, en el cual se podrán visualizar los datos y contestar las interrogantes planteadas en este proyecto por medio de consultas SPARQL.

Para el almacenamiento del grafo RDF se emplea Virtuoso Open Source⁵³, para pruebas iniciales se instaló localmente en Linux.

4.4.1. Consultas SPARQL.

Utilizando el Editor de Virtuoso Open Source, se realizaron algunas consultas que responderán a las preguntas formuladas en el punto 2 del presente proyecto. Afianzando así los principios de Linked Data.

Si las preguntas se realizan en un buscador normal como Google que genera búsquedas de manera sintáctica, se encontrarían varios resultados, tediosos y repetitivos que harán que el usuario pierda el tiempo. Teniendo en cuenta el inconveniente con lo mencionado se procederán a comparar respuestas en base a un buscador normal contra la información almacenada en el endpoint de

⁵³ <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/>

virtuoso después de seguir todo el proceso de publicación de Linked Data del presente proyecto.

Así tenemos:

1er Consulta:

a) Consulta en Google:

Pregunta: *¿Cuáles son los diferentes menús que ofrecen los cursos en los OCWSite?*

Rpta: *Cerca 2850 resultados, ninguna de ellas muestran directamente el contenido del Curso y su respectivo menu.*

Palabras de Búsqueda: *Como es un buscador por coincidencias busca información por coincidencia de palabras como por ejemplo: menú, ofrece, diferentes, curso, OCWSite, etc.*

b) Consulta Virtuoso Endpoint:

Pregunta: *¿Cuáles son los diferentes menús que ofrecen los cursos en los OCWSite?*

Para obtener la respuesta concreta se recurre a una consulta SPARQL en base al Schema del presente proyecto y navegando entre la ontología realizada se pide obtener toda la información que tenga como tipo de dato **Menú**.

Rpta: La información es única y concisa.

Búsqueda: Se basa en el URI que es de tipo Menú.

`</RDF:resource="http://www.utpl.edu.ec/OERvocab#Menu"/>`

Default Graph IRI

Query

```
prefix ocwo: <http://www.utpl.edu.ec/oervocab#>
select distinct ?TituloMenu where {?Enlace rdf:type
<http://www.utpl.edu.ec/oervocab#Menu>;
<http://xmlns.com/foaf/0.1/name> ?TituloMenu }
```

Execute Save Load Clear

TituloMenu
Lecture Notes
Assignments
Systems/Labs 06
Comps./Programming
Fluid Mechanics
Signals/Systems
Syllabus
Systems/Labs 04

Además podemos encontrar los enlaces de los distintos menús:

Curso	Menu
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/assignments/
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/assignments/
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/systems-labs-06/
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/comps-programming/
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/fluid-mechanics/
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/signals-systems/
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-01-unified-engineering-i-ii-iii-iv-fall-2005-spring-2006/systems-labs-04/

2da Consulta

Pregunta: ¿En qué plataforma esta desarrollado el OCWsite?

a) Consulta en Google

Rpta: Cerca de 65.700 resultados, unicamente un enlace da un panorama mas o menos claro de lo que se busca.

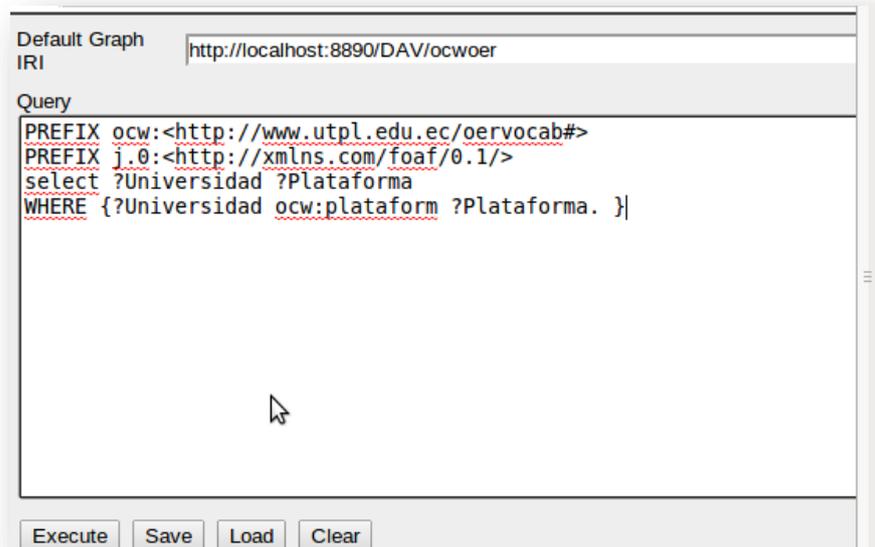
Palabras de Búsqueda: Los resultados obtenidos en la respuesta se basan en palabras como: Plataformas, desarrollo. Sitios OCW.

b) Consulta en Virtuoso EndPoint

Rpta: Nos envia Informacion de plataformas como educommons, moodle,propias, html.

Búsqueda: Se realiza en base a toda aquella información que sea de tipo propiedad y se denomine Plataforma.

RDF:about="http://www.utpl.edu.ec/OERvocab#plataform"



Universidad	Plataforma
http://www.utpl.edu.ec/oervocab#OcwSite	http://www.w3.org/2001/XMLSchema#string
http://ocw.mit.edu/index.htm	Desarrollo Propio

3era Consulta

Pregunta: ¿ Cuáles son los OERs que contienen los diferentes cursos?

a) Consulta en Google

Rpta: Cerca de 2.350.000 resultados , muchos de estos son conceptos de lo que es un OER.

Palabras de Búsqueda: La búsqueda se basa en palabras como: OERs, cursos,cotienen,etc.

b) Consulta en Virtuoso EndPoint

Rpta:

The screenshot shows a web interface for a Virtuoso EndPoint. At the top, there is a text input field for the "Default Graph IRI" containing the URL "http://localhost:8890/DAV/ocwoer". Below this is a section labeled "Query" containing a SPARQL query:

```
PREFIX ocw:<http://www.utpl.edu.ec/oervocab#>
PREFIX j.0:<http://xmlns.com/foaf/0.1/>
select ?Curso ?0er
WHERE {?Curso ocw:provides ?0er. }
```

At the bottom of the interface, there are four buttons: "Execute", "Save", "Load", and "Clear".

Curso	Oer
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/behavior_of_separated.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_10.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_3.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/assignments/ps5_solution.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_17.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_9.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_14.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_1.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_6.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_20.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_5.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/assignments/ps6_solution.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_30.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_12.pdf
http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/	http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-13-aerodynamics-of-viscous-fluids-fall-2003/lecture-notes/lecture_11.pdf

Búsqueda: La búsqueda se basa en el URI

RDF:resource="http://www.utpl.edu.ec/OERvocab#OER"

4ta. Consulta

Pregunta: ¿Cuáles son los consorcios existentes bajo los principios OCW?

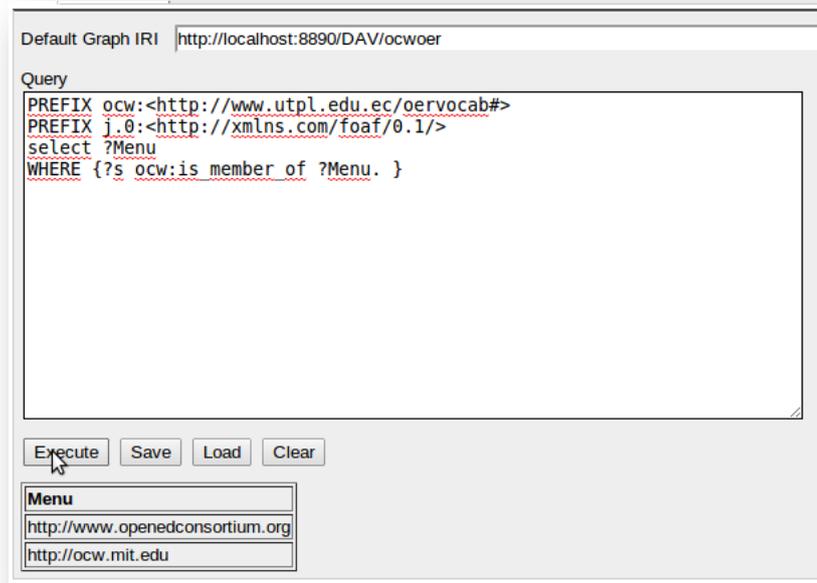
a) Consulta Google

Respuesta: Cerca de 20.800 resultados, de los cuales un enlace en la pagina principal habla del Consorcio Universia.

Palabras de Búsqueda: Consorcio, OCW.

b) Consulta Virtuoso EndPoint

Respuesta:



Búsqueda: La búsqueda se realiza en base al URI

<RDF:about="http://www.utpl.edu.ec/OERvocab#Consortium/">

5ta. Consulta

Pregunta: ¿Cuáles son las Universidades que contienen sitios OCW?

a) Consultas en Google

Respuesta: Cerca de 1420.000 resultados y la mayoría de ellos hablan de universidades que se encuentran en España y dirige a la página principal de lo que es un OCW para estas.

Palabras de Búsqueda: Por coincidencia las búsquedas se basan en Universidad, OCW, sitios, contienen, listas, etc.

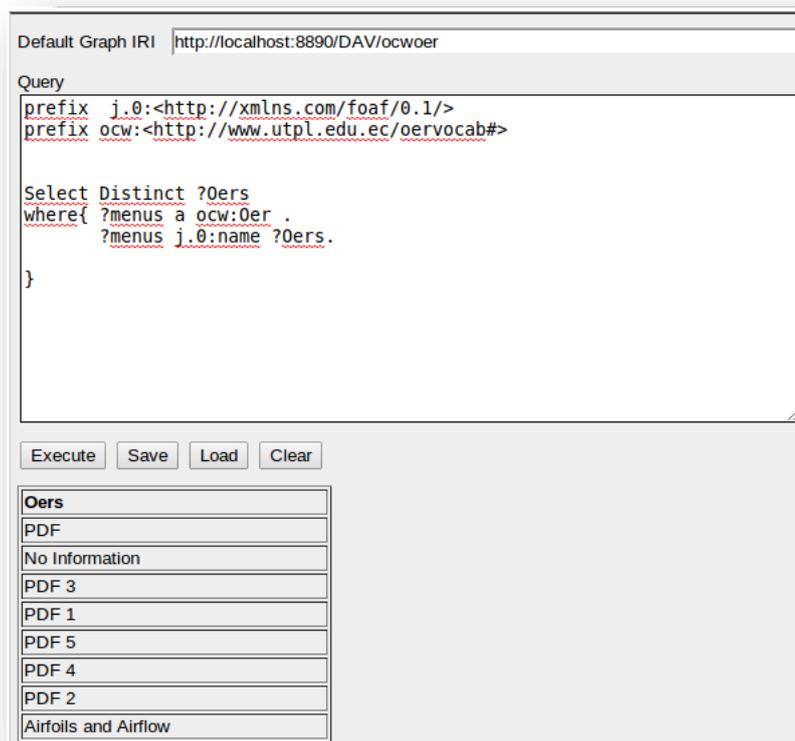
b) Consultas en Virtuoso EndPoint

Default Graph IRI

Query

```
prefix ocw: <http://www.utpl.edu.ec/oervocab#>
select * where {?EnlaceCurso rdf:type
<http://www.utpl.edu.ec/oervocab#EducationalOrganization>;
<http://xmlns.com/foaf/0.1/name> ?NombreUniversidad
}
```

EnlaceCurso	NombreUniversidad
http://web.mit.edu/	MIT - Massachusetts Institute of Technology
http://www.unican.es/index.html	Universidad de Cantabria
http://www.nd.edu/	University of Notre Dame
http://www.utpl.edu.ec/	Universidad Tecnica Particular de Loja - U T P L Ecuador
http://www.upm.es/institucional	Universidad Politécnica de Madrid
http://uc3m.es/	UC3M



Búsqueda: Se basa en la URI

RDF:resource="http://www.utpl.edu.ec/OERvocab#EducationalOrganization

4.5. Conclusiones

- La implementación de herramientas que permitan el modelado de información como es CmapTools es útil al plantear una visión más interactiva y gráfica del ambiente de trabajo.
- El reutilizar vocabularios existentes analizados y aceptados nos ayuda a evitar redundancia de información, además de apegarnos a la aplicación de Linked Data.
- Utilizar RDF Validator and Converter o W3C RDF Validation Service para validar la sintaxis del RDF
- Jena es un framework empleado para la transformación de modelos de información a formatos de lenguaje único. Su uso evita que se tengan que implementar nuevas herramientas para transformar directamente la información

en formatos OWL, RDF, etc. Pero para ello es muy importante tener muy claro el entorno de trabajo.

- Aplicar los principios de Linked Data en la publicación de OERs dentro de los OCW aporta en la iniciativa de mejorar la interoperabilidad, localización y consulta de dichos recursos. Convirtiendo dicha información en verdaderas fuentes de consumo al estar en un formato accesible desde cualquier parte a través de la Web.
- Mediante la implementación de procesos para extracción de información y su posterior transformación a formato RDF, permite establecer un modelo de aplicación para la generación de un almacén de datos de OERs que pueda ser consumido mediante consultas SPARQL, además de ser consumible en la web libremente para trabajos futuros.
- La generación de consultas SPARQL en la información obtenida, resuelven las interrogantes planteadas al inicio de la investigación. Dando la confianza de que la información que se almacena y reduciendo el tiempo de búsqueda.
- La información almacenada del entorno de trabajo del presente proyecto, al realizar las búsquedas por medio de consultas SPARQL es 100% confiable por el hecho de que se esta consultando de un tema en común y dichas consulta se basan en a exploración de nodos que contiene el Schema realizado para los Sitios OCW y sus respectivos OER. Si la misma búsqueda se realiza en buscadores por coincidencia los datos son variados pero únicamente 2 ó 4 de esos resultados nos brindan información de lo que buscamos.

DISCUSIÓN

En base a la problemática de extraer información de Recursos Educativos Abiertos almacenados en los diferentes sitios OCW, se aplicaron técnicas de extracción denominadas Web scraping y se planteo aplicar técnicas de Linked Data aportando de esta manera facilidad en el acceso, reutilización e interoperabilidad de dicha información.

Como primera instancia se procedió a conocer el ambiente de trabajo en el que se encuentran los OERs, para nuestro proyecto como son los diferentes sitios OCW, para ellos se buscaron los consorcios que se manejan bajo el concepto de OCW como son Open Consortium, Universia entre otros.

Analizar las estructuras implicaba inspeccionar el código fuente de cada una de las páginas en que se encuentra la información a extraer, obteniendo así los enlaces, detalle de cursos, menús y finalmente los OERs. Comparando al mismo tiempo que los enlaces que se almacenen sean correctos, puesto que en algunos casos luego de almacenada la información se podían verificar en la base de datos algunas URLs estaban acortadas, por lo que se procedía a través de consultas SQL a completar dicha información.

En base a los enlaces de los diferentes sitios OCW que se encuentran almacenados en una base de datos, se aplicó el proceso de extracción mediante el framework Scrapy, además aplicando una librería que analiza la validez de las diferentes URLs. Procediéndose como punto final de la extracción a la creación de scripts en base a los patrones obtenidos previo al análisis.

Tabla 16. Representación de datos obtenidos en extracción de información

Tipo de dato analizado	Cantidad
Patrones	
• Educommons	6
• Moodle	3
• HTML y otras	18
Universidades	85 enlaces
Cursos	7823
OERs	105929

El crear scripts que realicen el trabajo de una manera mecánica ahorra tiempo, recursos y dinero, puesto que la información amplia que nos tomaría meses en analizar cada uno de los lugares y recursos a obtener manualmente, es ahí donde realizar web scraping a los sitios OCW se vuelve nuestro punto clave para la obtención de estos en base a estructuras de similitud entre sitio y sitio. Puesto que el *Web scraping* se enfoca más en la transformación de datos sin estructura en datos estructurados que pueden ser almacenados y analizados en una base de datos general. Sin violar ninguna norma empleada por los diferentes sitios ya que prácticamente todos los OCW se manejan bajo licencias de reusó y disponibilidad de sus recursos.

La información obtenida por medio de este proyecto con el uso de Scrapy reduce el tiempo y el consumo de recursos humanos al automatizar los procesos para almacenar información, los datos expuestos son la muestra clara de la cantidad de información que se puede obtener en menos tiempo que si se lo realizará manualmente.

Luego de obtener la información de los distintos OERs se procedió a extraer los metadatos existentes en los recursos; por lo que genera la creación de un analizador y extractor de metadatos, con el cual inicialmente se descargan los recursos en una dirección física y luego se procede a analizar los diferentes tipos de metadatos por medio de librerías implementadas en dicho analizador y finalmente se almacenan en la base de datos general enriqueciendo así la información de los OERs.

Otro de los puntos a tratar es el modelado de la información obtenida, siendo importante afianzar el conocimiento que se obtuvo previamente al observar los sitios; creando así un modelo más robusto del entorno de trabajo del presente proyecto,

Conocer el ámbito en el que se maneja la información de dichos Recursos educativos abiertos nos lleva al momento de preguntarnos: ¿Para qué me va a servir esta información?, ¿A quien va dirigida? .Pues bien denotando la importancia de la información podemos concluir que servirá a toda persona ó institución que desee afianzar sus conocimientos en base a estos recursos, siendo ahí donde empezamos a conocer nuestro ambiente de trabajo; con esto nace una nueva incógnita ¿Cómo haremos compartir la información que tenemos almacenada?. Es en este punto en donde Linked Data nos ofrece la solución de publicar la

información en un formato accesible desde cualquier parte a través de la Web afianzando así al objetivo principal del proyecto en curso como es la accesibilidad e interoperabilidad de OERs.

Uno de los principios básicos de Linked Data es crear un vocabulario que explique semánticamente el entorno de trabajo, para ellos se analizaron varios vocabularios en los cuales no se encontraron propiedades que representen el entorno en su totalidad, por lo que se procede a crear algunos términos que aporten semánticamente información acerca los OERs además de reutilizar los existentes naciendo así el vocabulario OcwOer. En el cual se pueden observar las relaciones entre clases que van desde el espacio principal que son la Instituciones educativas, pasando por los sitios OCW sus cursos y finalmente los OERs; todo esto en base a los consorcios a los que pertenecen y el país en el que se encuentran.

Finalmente el uso del framework Jena empleado para java hace más fácil el volcar lo modelado a un entorno de programación, implementando vocabularios existentes y el nuestro, con este framework no se hizo necesario implementar nuevas herramientas para transformar datos a RDF siendo sencillo transformar simple información a data que será entendido por máquinas y por ende empleado por buscadores semánticos siendo así eminente el enlace de información con la nube de datos. Comprobándose además que al mantener la información en un mismo espacio y formato como es RDF, y basándonos en el punto final del proceso de publicación de datos de Linked Data como son las consultas SPARQL, podemos denotar que son de mayor precisión que si se realizarán en un buscador basado en coincidencias de texto.

Analizando las diferentes consultas y comparando las búsquedas entre un buscador sintáctico como lo es Google y un almacén de datos rdf como lo denominaremos al Virtuoso, podemos constatar que el mantener la información de los sitios OCW en un solo lugar, proporciona facilidad al usuario al evitar la perdida de tiempo en buscar enlace por enlace ó sitio por sitio como lo hace un buscador. Por ejemplo, se planteó una pregunta básica como es ¿Cuántas universidades contienen sitios OCW?, arrojando aproximadamente 62.900 enlaces en los cuales el usuario debería comprobar si son reales en el buscador Google; mientras que con consultas SPARQL en nuestro endpoint de Virtuoso nos brinda información es concisa y objetiva.

Podemos observar además al realizar consultas en Google que la información que maneja de los diferentes sitios OCW, no se encuentra relacionada directamente entre si; por ellos de cada una de las búsquedas realizadas el 1% de estas nos arroja información que podría tornarse como clara según el ambiente de este proyecto y el 99% restante es información suelta en la Web que por coincidencias pareciera que se tiene información del tema.

Al contrarrestar la información que se extrajo y bajo la cual se aplicó las técnicas de Linked Data, podemos denotar la importancia de aplicarlas a los diferente ambientes de trabajos, pues la información manejaría un mismo formato y podría acrecentar los almacenes de información en distintos aspectos como son la educación, el comercio, turismo, etc con datos que relacionados entre sí generen información de calidad.

CONCLUSIONES

- “Promoción de Recursos Educativos Abiertos” es una propuesta de trabajo que ha tomado gran fuerza, pues compartir la información de una manera libre y bajo el concepto de Open Data genera que los usuarios finales puedan tener recursos de calidad, además de que esta información pueda ser inferida por medio de máquinas a través de la implementación de la Web semántica.
- El aplicar técnicas de scrapeado de información en la Web (scraping Web), facilita a la extracción de información, implementando estructuras de búsqueda tomando en consideración la parte técnica del ambiente en que se encuentra la información; se han podido extraer aproximadamente 85 sitios OCW que manejan las Universidades a nivel mundial y una cantidad aproximada de 105,929 OERs. Esto es aproximadamente un 90% más de lo que nos brindaría una búsqueda en la Web; alcanzando así uno de los objetivos iniciales como es almacenar la mayor cantidad de información del entorno de trabajo para ser tratada bajo los principios de Linked Data.
- La estructuración e integración de los cursos OCW en una base de datos común, genera información clasificada y detallada sobre recursos educativos abiertos, permitiendo que dicha base de datos pueda ser explotada en diferentes escenarios de uso.
- El mantener un vocabulario consensuado que explique el ambiente de trabajo con el que se cuenta, es de gran importancia al momento de aportar nuevas clases ó conocimiento sobre el mismo tema, haciendo que este se vuelva inmensamente sólido, completo y reutilizable.
- El estado actual de las universidades que contienen sitios OCW están tomando conciencia de la importancia de generar conocimiento de alta calidad, aunque hay algunos sitios que están desactualizados la mayoría de las Universidades que tienen como pilar el concepto de compartir información están generando calidad de su servicio Web y a su vez están buscando plataformas similares para la generación de información.
- El mantener una colaboración activa en la generación de datos similares, ya sean académicos, artesanales, industriales, etc ayudarán a crear grandes

almacenes de información que pueden trabajarse bajo principios que aporten al crecimiento y desarrollo de la Web semántica.

- Los principios de Linked Data permiten crear verdaderos sistemas de información aportando de esta manera el cumplimiento del por qué fue desarrollada la Web. Tomando como punto de referencia cualquier ambiente de trabajo que necesite ser explotado semánticamente.

RECOMENDACIONES

Durante la elaboración del presente proyecto se adquirieron lecciones que pueden utilizarse como recomendaciones para proyectos futuros que manejen el mismo ambiente:

- Generar formatos de extracción de datos universales de acuerdo a las diferentes estructuras que contienen los Sitios Web. Pues esto evitara que muchos datos se queden fuera del scraping.
- Crear programas ó scripts para mantener una actualización activa, que exploren los cambios que han ocurrido en los diferentes sitios OCW, puesto que la información crece prácticamente a diario, influenciados por la competencia de mantener recursos educativos abiertos de calidad en los diferentes Organizaciones educativas.
- Utilizar una herramienta de modelado de conocimiento como lo es Cmap Tools COE, teniendo así un enfoque general del ambiente en el que se va a trabajar y la previa construcción del vocabulario.
- La reutilización de vocabularios es un punto importante, pues de esta manera evitamos la redundancia de información. Además que estos ya han sido revisados, se encuentran consensuados y muchos de estos son estándares; nos dan la seguridad de la calidad del término al reutilizarlo.
- Utilizar RDF Validator and Converter o W3C RDF Validation Service para validar la sintaxis del RDF. De esta manera conociendo si el modelado de la información está acorde con la generación de información al implementar la programación.
- Al implementar el uso de Jena se deberán primero incluir aquellos vocabularios que se fuesen a reutilizar, ya que al momento de generar la información necesaria algunos prefijos suelen exponerse con nombres por defecto, haciendo que las consultas SPARQL resulten prácticamente inútiles.
- En vista que la información que obtiene mediante el uso de scripts basados en el framework Scrapy y dicho proceso se realiza manualmente. Se recomienda como un trabajo futuro el uso de servicios Web que automaticen el proceso de extracción.

BIBLIOGRAFÍA

- Berners-Lee, T. (2005). *Uniform Resource Identifier (URI): Generic Syntax*. The Internet Society Sitio web: <http://tools.ietf.org/pdf/rfc3986.pdf>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). *The Semantic Web*. Scientific American. Sitio web: <http://www.disciplineoforganizing.org/wp-content/uploads/2013/01/SemanticWeb.pdf>.
- Berners-Lee, Tim (2006). *Linked Data*. <http://www.w3.org/DesignIssues/LinkedData.html>
- Borra, O. (2010), *Observatorio de Plataformas para OCW (2012)*: Madrid.
- CC Legal. Sitio Web: <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>
- Corcho, O. & Gómez, A... Universidad Politécnica de Madrid Florianópolis, *Herramientas de publicación de datos para portales de datos abiertos*, September 1st 2010 (3º OntoBras 2010). Sitio web: http://datos.gob.es/datos/sites/default/files/files/%2004.V3_Tecnologia_Portales_OpenData.pdf
- Creative Commons Organization (2012), *Creative Commons Attribution 3.0 License*. <http://creativecommons.org/licenses/by/3.0/>
- FUNDACION UNIVERSIA OCW. <http://OCW.universia.net>
- GNU: ¿Qué es el software libre? .2010. <http://www.gnu.org/philosophy/free-sw.html>
- Graham, K & Carroll, J. (2003). *Resource Description Framework (RDF): Concepts and Abstract Syntax*. W3C. Sitio web: <http://www.w3.org/TR/2003/PR-rdf-concepts-20031215/>
- Gruber, Thomas R. "Towards principles for the design of ontologies used for knowledge sharing". En: Guarino, N.; Poli, R. (eds.). *Formal ontology in conceptual analysis and knowledge representation*. Dordrecht: Kluwer Academic Publishers, 1993.
- Gutierréz, C. (2008). *Cómo funciona La Web*. Santiago de Chile: Gráfica LOM.
- Heath, T., Hepp, M., and Bizer, C. (eds.). *Special Issue on Linked Data*, International Journal on Semantic Web and Information Systems (IJSWIS). Sitio web: <http://linkeddata.org/docs/ijswis-special-issue>
- Heath, T. & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space* Synthesis Lectures on the Semantic Web: Theory and Technology. USA: Morgan & Claypool.

- Lmarca, M. Metadatos: *Hipertexto: El nuevo concepto de documento en la cultura de la imagen., Esquema RDF* , 2012 Disponible en <http://www.hipertexto.info/documentos/metadatos.htm>
- Manola, F., Miller, E. & McBride B. (2004). *RDF PRIMER*. <http://www.w3.org/TR/RDF-primer/>
- Mit, 2010. *Open Source Initiative OSI - The MIT License (MIT): Licensing*. Sitio Web: <http://www.opensource.org/licenses/mit-license.php>
- MITOCW statistics. http://OCW.mit.edu/about/site-statistics/monthly-reports/MITOCW_DB_2011_11.pdf
- Noy, & McGuinness Deborah L. *Ontology development 101: a guide to creating your first ontology*. Informe técnico. Universidad de Stanford, 2000. Sitio Web: <http://protege.stanford.edu/ontology101-noy-mcguinness.html>
- OPEN DATA FOUNDATION (2012). *OPEN DATA FOUNDATION*. <http://www.opendatafoundation.org/>
- OPEN EDUCATION CONSORTIUM. (2012). Sitio Web: <http://www.oeconsortium.org/>
- OpenCourseWare UTPL, (2012). *OpenCourseWare UTPL*. Sitio Web: <http://ocw.utpl.edu.ec/>
- Piedra, N. & Romero, A... (2012). *CALIDAD DE CONTENIDOS EN OCW*. Disponible en http://repositorial.cuaed.unam.mx:8080/jspui/bitstream/123456789/2706/1/romero_piedra_calidad_de_contenidos_OCW.pdf
- Piedra, N. ; Tovar, E.; Colomo-Palacios, R ., Lopez, J. , Chicaiza, A. , (2014) "*Consuming and producing linked open data: the case of OpenCourseWare*", Program: electronic library and information systems, Vol. 48 Iss: 1, pp.16 - 40
- Piedra, N.; Lopez, J.; Chicaiza J...: Universidad Técnica particular de Loja (2014). *Serendipity a platform to discover and visualize Open OER Data from OpenCourseWare repositories*. Disponible en: http://conference.oeconsortium.org/2014/wp-content/uploads/2014/02/Paper_82-serendipity.pdf
- Tim Berners-Lee, CERN March 1989, May 1990 : <http://www.w3c.es/Presentaciones/2009/1111-introduccionUM-MA/>
- Tovar, E., Piedra, N., Chicaiza, J., Lopez, J., Martínez, O.: Development and promotion of OERs. *Outcomes of an international research project under OpenCourseWare model*. Journal of Universal Computer Science (2012)

- Tovar, E., Piedra, N., Chicaiza, J., Lopez, J., Martínez, O.: 2012 World Open Educational Resources Congress 2012. Semantic Technologies: Linked Data and OER. Disponible en <http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/themes/piedra.pdf>
- Unesco. Foro sobre Impacto de los Cursos Abiertos para Educación Superior en los países en desarrollo, *RECURSOS EDUCATIVOS DE LIBRE ACCESO*. Sitio web: <http://www.unesco.org/new/es/unesco/themes/icts/open-educational-resources/>
- World Wide Web Consortium (2010). *Guía Breve de Web Semántica*. <http://www.w3c.es/divulgacion/quiasbreves/linkeddata>

ANEXOS

Anexo 1

En este espacio se podrán encontrar la descripción detallada de cada uno de los elementos empleados en el vocabulario que describe el ambiente de trabajo del presente proyecto como son los OCW.

Vocabulario de Clases

A continuación detallaremos cada una de las clases tomando en consideración URI, Metadato y Descripción.

Clase: EducationalOrganization

URI:	http://www.utpl.edu.ec/OERvocab#EducationalOrganization
Etiqueta:	Organización Educativa
Descripción:	Representa a toda institución educativa, pero principalmente a aquella q trabajan bajo el concepto OCW OpenData.

Clase: Country

URI:	http://www.utpl.edu.ec/OERvocab#Country
Etiqueta:	País
Descripción:	Representa el conjunto de países.

Clase: City

URI:	http://www.utpl.edu.ec/OERvocab#City
Etiqueta:	Ciudad
Descripción:	Representa el conjunto de ciudades

Clase: Continent

URI:	http://www.utpl.edu.ec/OERvocab#Continent
Etiqueta:	País
Descripción:	Engloba a todos los países que se encuentra en una determinada situación geográfica.

Clase: Consortium

URI:	http://www.utpl.edu.ec/OERvocab#Consortium
Etiqueta:	Consortio
Descripción:	Representa aquellas organizaciones en las que se han unificado diferentes Organizaciones educativas que comparten recursos educativos abiertos.

Clase: OCWSite

URI:	http://www.utpl.edu.ec/OERvocab# OCWSite
Etiqueta:	Sitios OCW
Descripción:	Representa aquellos sitios Webs que mantienen las universidades para compartir recursos educativos bajo el concepto de Open Data.

Clase: Course

URI:	http://www.utpl.edu.ec/OERvocab# Course
Etiqueta:	Cursos
Descripción:	Representa aquellos cursos que comparten las distintas Organizaciones Educativas en sus sitios OCW.

Clase: KnowledArea

URI:	http://www.utpl.edu.ec/OERvocab# KnowledArea
Etiqueta:	Área de Conocimiento
Descripción:	Representa las áreas de conocimiento al que pertenece cada curso según la temática que este trata. Ejm. Física, Química, etc.

Clase: Menu

URI:	http://www.utpl.edu.ec/OERvocab# Menu
Etiqueta:	Menu
Descripción:	Representa la estructura que mantiene un curso. Entre los diferentes menús se pueden citar Inicio, tareas, syllabus, Cronograma, Tareas, etc.

Clase: OER

URI:	http://www.utpl.edu.ec/OERvocab#OER
Etiqueta:	OER (Recurso Educativo Abierto)
Descripción:	Representa aquellos Recursos Educativos disponibles en varios formatos dentro de los cursos y estructura con la que trabajan los diferentes Sitios OCW.

Vocabulario de Propiedades.

En este punto se expondrán las propiedades creadas bajo nuestro ambiente de trabajo, tomando en consideración Dominios y Rangos.

Propiedad: Plataform

URI:	http://www.utpl.edu.ec/OERvocab#plataform
Etiqueta:	Plataforma
Descripción:	Es la estructura física y técnica en que esta desarrollado el sitio OCW.
Dominio:	OCW:OCWSite
Rango:	xsd:string
Tipo	owl:DatatypeProperty

Propiedad: license_link

URI:	http://www.utpl.edu.ec/OERvocab#license_URI
Etiqueta:	license_URI
Descripción:	Representa la dirección Web de la licencia bajo los cuales esta cada uno de lo sitios OCW.
Dominio:	OCW:OCWSite

Rango:	xsd:string
Tipo:	owl:DatatypeProperty
Similar a:	cc:legalcode

Propiedad: is_located_in

URI:	http://www.utpl.edu.ec/OERvocab#is_located_in
Etiqueta:	Is_located_in
Descripción:	Establece la relación de ubicación geográfica en la que se encuentran algún recurso, entidad, etc.
Dominio:	OCW:EducationalOrganization
Rango:	OCW:City
Tipo:	owl:ObjectProperty

Propiedad: corresponds_to

URI:	http://www.utpl.edu.ec/OERvocab#corresponds_to
Etiqueta:	corresponds_to
Descripción:	Indica la correspondencia de un ente particular a uno en general.
Dominio:	ocw:Country
Rango:	ocw:Continent
Tipo:	owl:ObjectProperty

Propiedad: belonging_to

URI: http://www.utpl.edu.ec/OERvocab#belonging_to

Etiqueta: belonging_to

Descripción: Establece relación de pertenencia de un recurso, entidad, etc a un todo.

Dominio: OCW:City

Rango: ocw:Country

Tipo: owl:ObjectProperty

Propiedad: is_member_of

URI: http://www.utpl.edu.ec/OERvocab#is_member_of

Etiqueta: is_member_of

Descripción: Hace referencia a la participación activa dentro de una entidad, consorcio, etc. Estableciendo la relación entre los Sitos OCW y los Consorcios.

Dominio: OCW:OCWSite

Rango: ocw:Consortium

Tipo: owl:ObjectProperty

Propiedad: contains

URI: <http://www.utpl.edu.ec/OERvocab#contains>

Etiqueta: is_member_of

Descripción: Indica que el recurso, entidad, etc esta siendo contenido por una clase principal.

Dominio: ocw:Course

Rango: ocw:Menu

Tipo: owl:ObjectProperty

Propiedad: offer**URI:** <http://www.utpl.edu.ec/OERvocab#offer>

Etiqueta:	Offer
Descripción:	Hace referencia la oferta que realiza una entidad a otra.
Dominio:	ocw:OCWSite
Rango:	ocw:Course
Tipo:	owl:ObjectProperty

Propiedad: organized_by**URI:** http://www.utpl.edu.ec/OERvocab#organized_by

Etiqueta:	organized_by
Descripción:	Indica la categorización que tiene una entidad, recurso, etc por otra.
Dominio:	Ocw:OCWSite
Rango:	Ocw:KnowledArea
Tipo:	owl: ObjectProperty

Propiedad: stored**URI:** <http://www.utpl.edu.ec/OERvocab#stored>

Etiqueta:	Stored
Descripción:	Indica el almacenamiento que tiene un curso ó la pertenencia a una entidad de rango mayor.
Dominio:	OCW:KnowledArea
Rango:	OCW:Course
Tipo:	owl: ObjectProperty

Propiedad: classified_by

URI: http://www.utpl.edu.ec/OERvocab#classified_by

Etiqueta: classified_by

Descripción: Indica que una clase esta siendo clasificada por ciertas características por otra clase.

Dominio: OCW: Course

Rango: OCW: KnowledArea

Tipo: owl: ObjectProperty

Propiedad: state

URI: <http://www.utpl.edu.ec/OERvocab#state>

Etiqueta: State

Descripción: Representa la ubicación física y geográfica de una ciudad en un área territorial política y social que maneja sus propias normas ó leyes dentro de una región.

Dominio: OCW: City

Rango: xsd:string

Tipo: owl:DataTypeProperty

Propiedad: provides

URI: <http://www.utpl.edu.ec/OERvocab#provides>

Etiqueta: Provides

Descripción: Indica que una clase contiene ó provee de alguna característica de otra clase.

Dominio: OCW: Course

Rango: OCW:OER

Tipo:	owl: ObjectProperty
--------------	---------------------

Propiedad: capital

URI:	http://www.utpl.edu.ec/OERvocab#capital
-------------	---

Etiqueta:	Capital
------------------	---------

Descripción:	Representa la situación geográfica de la ciudad o país dentro de un continente.
---------------------	---

Dominio:	OCW: Country
-----------------	--------------

Rango:	xsd:string
---------------	------------

Tipo:	owl:DataTypeProperty
--------------	----------------------

Propiedad: logo

URI:	http://www.utpl.edu.ec/OERvocab#logo
-------------	---

Etiqueta:	Logo
------------------	------

Descripción:	Representa el url de la imagen de las Organizaciones Educativas como de los Curso.
---------------------	--

Dominio:	OCW: Course and OCW: EducationalOrganization
-----------------	--

Rango:	xsd:string
---------------	------------

Tipo:	owl:DataTypeProperty
--------------	----------------------

Propiedad: accommodate

URI:	http://www.utpl.edu.ec/OERvocab#accommodate
-------------	---

Etiqueta:	Accommodate
------------------	-------------

Descripción:	Indica que una clase almacena las características de otra, clasificando por categorías.
---------------------	---

Dominio:	OCW: OCWSite
-----------------	--------------

Rango:	xsd:string
Tipo:	owl: ObjectProperty

Propiedad: metadata_type

URI:	http://www.utpl.edu.ec/OERvocab#metadata_type
Etiqueta:	metadata_type
Descripción:	Representa los tipos de metadatos encontrados en los OER. Ej creator
Dominio:	OCW:OER
Rango:	xsd:string
Tipo:	owl:DataTypeProperty

Propiedad: metadata_value

URI:	http://www.utpl.edu.ec/OERvocab#metadata_value
Etiqueta:	metadata_value
Descripción:	Es el valor de los metadatos ó la descripción de estos.
Dominio:	OCW:OER
Rango:	xsd:string
Tipo:	owl:DataTypeProperty

Propiedad: dbpedialink

URI:	http://www.utpl.edu.ec/OERvocab#dbpedialink
Etiqueta:	Dbpedialink
Descripción:	Representa el Url al que pertenece una entidad que se encuentra representada en la dbpedia.

Dominio:	OCW:EducationalOrganization
Rango:	xsd:string
Tipo:	owl:DataTypeProperty

Propiedad: dbpediaLanguage

URI:	http://www.utpl.edu.ec/OERvocab#dbpediaLanguage
Etiqueta:	DbpediaLanguage
Descripción:	Representa el Url que tiene un lenguaje representado en la dbpedia y que es empleado en los diferentes Sitios OCW.
Dominio:	OCW:OCWSite
Rango:	xsd:string
Tipo:	owl:DataTypeProperty

Propiedad: Language_alternative

URI:	http://www.utpl.edu.ec/OERvocab#language_alternative
Etiqueta:	language_alternative
Descripción:	Representa el nombre de los varios lenguajes que no son principales y que se utilizan en los Sitios OCW.
Dominio:	OCW:OCWSite
Rango:	xsd:string
Tipo:	owl:DataTypeProperty

Propiedad:dbpediaCapital

URI: <http://www.utpl.edu.ec/OERvocab#dbpediaCapital>

Etiqueta: DbpediaCapital

Descripción: Representa el URL de una capital que se encuentra representada en la dbpedia.

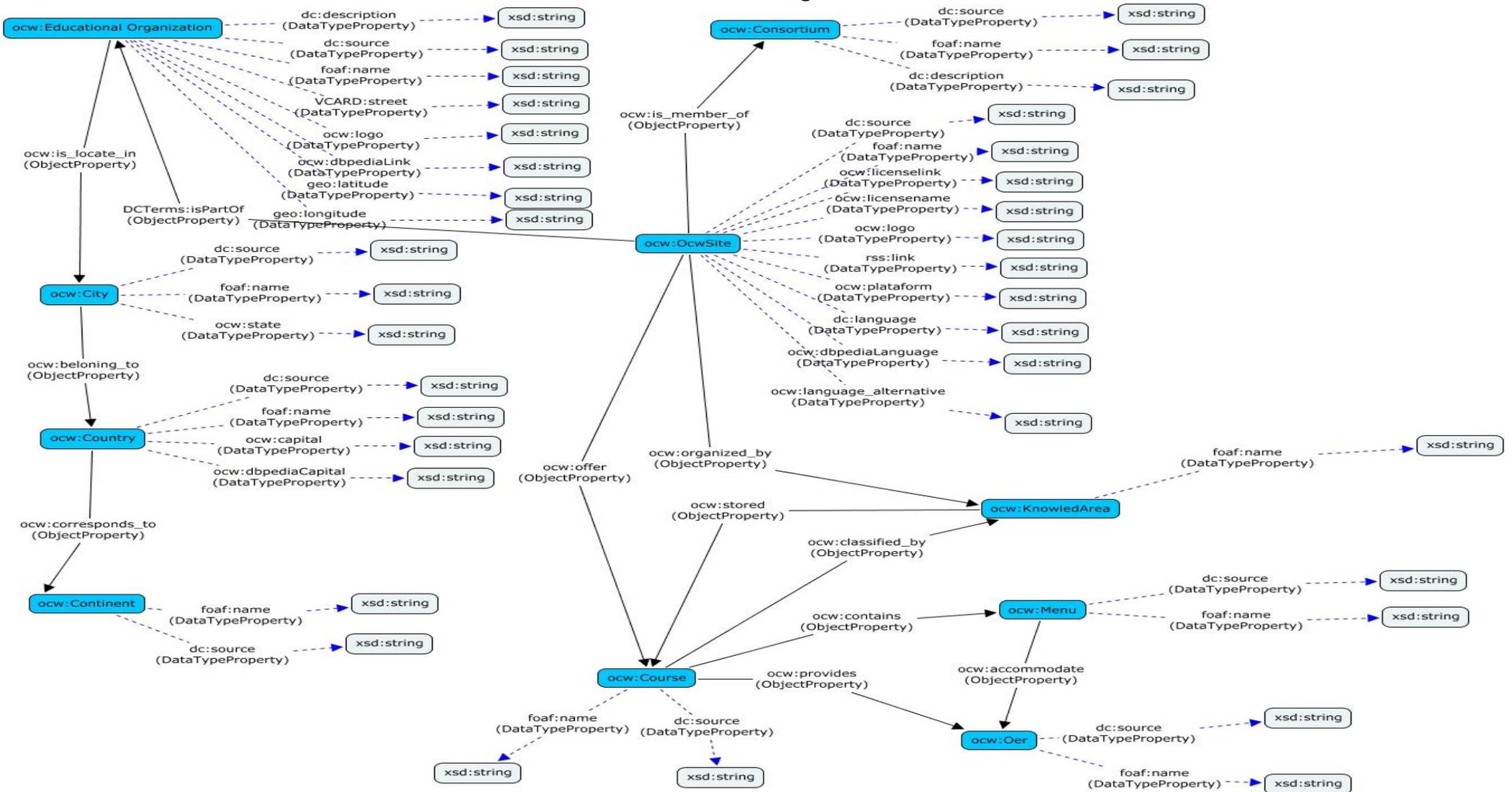
Dominio: OCW:Country

Rango: xsd:string

Tipo: owl:DataTypeProperty

Anexo 2

Creación de ontología OCW/OER



Anexo 3

Schema vocabulario OcwOer

```
<RDF:RDF
xmlns:RDF="http://www.w3.org/1999/02/22-RDF-syntax-ns#"
xmlns:j.0="http://xmlns.com/foaf/0.1/"
xmlns:vcard="http://www.w3.org/2001/vcard-RDF/3.0#"
xmlns:OCW="http://www.utpl.edu.ec/OERvocab#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rss="http://purl.org/rss/1.0/"
xmlns:RDFs="http://www.w3.org/2000/01/RDF-schema#"
xmlns:j.1="http://www.w3.org/2003/01/geo/wgs84_pos#" >
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#provides">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#OER"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#Course"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#is_member_of">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#Consortium"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#licenselink">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>
<RDFs:comment xml:lang="es">Representa la direcci3n Web de la licencia bajo
los cuales esta cada uno de lo sitios OCW</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#dbepediaLink">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain
RDF:resource="http://www.utpl.edu.ec/OERvocab#EducationalOrganization"/>
```

```

<RDFs:comment xml:lang="es">Representa el Url al que pertenece una entidad
que se encuentra representada en la dbpedia</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#offer">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#Course"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#corresponds_to">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#Continent"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#Country"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#OER">
<OCW:metadata_value
RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:metadata_type
RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<dc:source RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:comment xml:lang="es">Representa aquellos Recursos Educativos
disponibles en varios formatos dentro de los cursos y estructura con la que trabajan
los diferentes Sitios OCW.</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#metadata_type">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OER"/>
<RDFs:comment xml:lang="es">Representa los tipos de metadatos encontrados en
los OER. Ej creator</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#licensename">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>

```

```

<RDFs:comment xml:lang="es">Representa el nombre de aquellas licencias bajo
las cuales esta funcionando un sitio OCW</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#accommodate">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#OER"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#Menu"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#organized_by">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#KnowledArea"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#plataform">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>
<RDFs:comment xml:lang="es">Es la estructura física y técnica en que esta
desarrollado el sitio OCW</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description
RDF:about="http://www.utpl.edu.ec/OERvocab#dbpediaLanguage">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>
<RDFs:comment xml:lang="es">Representa el Url que tiene un lenguaje
representado en la dbpedia y que es empleado en los diferentes Sitios
OCW</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#Menu">
<dc:source RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:comment xml:lang="es">Representa la estructura que mantiene un curso.
Entre los diferentes menús se pueden citar Inicio, tareas, syllabus, Cronograma,
Tareas, etc.</RDFs:comment>

```

```

<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#City">
<OCW:state RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:comment xml:lang="es">Representa el conjunto de
ciudades</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#Course">
<dc:source RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:comment xml:lang="es">Representa aquellos cursos que comparten las
distintas Organizaciones Educativas en sus sitios OCW</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#classified_by">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#KnowledArea"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#Course"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description
RDF:about="http://www.utpl.edu.ec/OERvocab#metadata_value">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OER"/>
<RDFs:comment xml:lang="es">Es el valor de los metadatos Ã³ la descripciÃ³n de
estos</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#logo">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>
<RDFs:domain
RDF:resource="http://www.utpl.edu.ec/OERvocab#EducationalOrganization"/>
<RDFs:comment xml:lang="es">Representa el url de la imagen de las
Organizaciones Educativas como de los Curso.</RDFs:comment>

```

```

<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#Country">
<OCW:dbpediaCapital
RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:capital RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:comment xml:lang="es">Representa el conjunto de
paises</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#is_located">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#Country"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#Consortium"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#state">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#City"/>
<RDFs:comment xml:lang="es">Representa la ubicaci3n f3sica y geogr3fica de
una ciudad en un 3rea territorial pol3tica y social que maneja sus propias normas
3 leyes dentro de una regi3n.</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#Consortium">
<dc:description RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<dc:source RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:comment xml:lang="es">Representa aquellas organizaciones en las que se
han unificado diferentes Organizaciones educativas que comparten recursos
educativos abiertos</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#KnowledArea">
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>

```

```

<RDFs:comment xml:lang="es">Representa las Áreas de conocimiento al que pertenece cada curso según la temática que este trata. Ejm. Física, Química, etc.</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#capital">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#Country"/>
<RDFs:comment xml:lang="es">Representa la situación geográfica de la ciudad o país dentro de un continente</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#OCWSite">
<RDFs:comment xml:lang="es">Representa aquellos sitios Webs que mantienen las universidades para compartir recursos educativos bajo el concepto de Open Data</RDFs:comment>
<OCW:logo RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:plataform RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:licensename RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
<dc:source RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:licenselink RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:dbpediaLanguage
RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<dc:language RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:language_alternative
RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<rss:link RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#belonging_to">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#Country"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#City"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#stored">

```

```

<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#Course"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#KnowledArea"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#dbpediaCapital">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#Country"/>
<RDFs:comment xml:lang="es">Representa el URL de una capital que se
encuentra representada en la dbpedia</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#is_locate_in">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#City"/>
<RDFs:domain
RDF:resource="http://www.utpl.edu.ec/OERvocab#EducationalOrganization"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description
RDF:about="http://www.utpl.edu.ec/OERvocab#language_alternative">
<RDFs:range RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>
<RDFs:comment xml:lang="es">Representa el nombre de los varios lenguajes que
no son principales y que se utilizan en los Sitios OCW.</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</RDF:Description>
<RDF:Description
RDF:about="http://www.utpl.edu.ec/OERvocab#EducationalOrganization">
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
<j.1:longitude RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<dc:description RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:dbpediaLink
RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<dc:source RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<OCW:logo RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>

```

```

<RDFs:comment xml:lang="es">Representa a toda instituci3n educativa, pero
principalmente a aquella que trabajan bajo el concepto OCW
OpenData.</RDFs:comment>
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<j.1:latitude RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<vcard:Street RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#contains">
<RDFs:range RDF:resource="http://www.utpl.edu.ec/OERvocab#Menu"/>
<RDFs:domain RDF:resource="http://www.utpl.edu.ec/OERvocab#Course"/>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#Continent">
<j.0:name RDF:resource="http://www.w3.org/2001/XMLSchema#string"/>
<RDFs:comment xml:lang="es">Engloba a todos los pa3ses que se encuentra en
una determinada situaci3n geogr3fica</RDFs:comment>
<RDF:type RDF:resource="http://www.w3.org/2002/07/owl#Class"/>
</RDF:Description>
</RDF:RDF>

```

Anexo 4

Denominaciones de menús de cursos

Plataforma Educommons

- Home
- Programa
- Plan de curso
- Actividades y ejercicios
- Bibliografía
- Profesorado
- Descargar este curso
- Introducción
- Sobre Profesor
- Objetivos
- Metodología
- Material de clase
- Sistema de Evaluación
- Nombre de Materia/Asignatura
- Actividades
- Evaluación
- Recursos Complementarios
- Practicas
- Pruebas de Evaluación
- Otros Recursos
- Enlaces
- Presentación
- Bloques Temáticos
- Material de apoyo y trabajo
- Referencias Bibliográficas

- Redes Sociales(OJO)
- Bloques de Contenido
- Video de Presentación
- Equipo Docente
- Contenidos
- Exámenes
- Objetivos
- Metodología de Trabajo
- Enlace de interés
- Presentación(Bienvenida, créditos, Descripción, Estructura, Objetivos, Mapa, Evaluación)
- Material
- Programa y Descripción del Curso
- Guía de Aprendizaje
- Profesores
- Download This course
- Export into OpenOCW
- Ejercicios resueltos
- Ejercicios Propuestos
- Ejercicios de selectividad
- Webgrafía
- Enlaces de Interés
- Acerca del Profesor
- Material de Soporte
- Investigación
- Evaluaciones
- Laboratorios
- Guía docente
- Lecturas recomendadas
- Materiales de Estudio

- Practicas y ejercicios
- Procedimientos de autoevaluación
- Fundamentación y competencias necesarias
- Competencias genéricas y específicas
- Plan de trabajo
- Autoevaluaciones
- Sobre el autor
- Tareas
- Syllabus
- Actividades y Materiales
- Exámenes
- Problemas
- Presentación/Programa
- Descarga todos los materiales
- Materiales
- Lectura obligatoria
- Lecturas Complementarias
- Ejercicios, proyectos y casos
- Software
- Autores del material
- Grabaciones de clase
- Conclusiones
- Colaboradores
- Plan de estudios
- Módulos
- Glosario
- Calendario
- Conferencias
- Contenido popular

- Agradecimientos

Plataforma Moodle:

- Información General
- Programa
- Guías de Aprendizaje
- Material de Clase
- Actividades, prácticas, proyectos y/o casos.
- Materia de estudio
- Evaluación
- Descargar curso
- Descargar el curso completo
- Presentación
- Prácticas y ejercicios
- Exámenes
- Otros recursos
- Bibliografía

Parece moodle o propio:

- Curso
 - Temario
 - Profesores
 - Organización Docente
- Materiales
 - Documentos
 - Autores
- Enlaces externos
 - Aula virtual

Se desconoce la plataforma:

- Contenido guía docente
 - Descripción general de la asignatura
 - Competencias

- Conocimientos recomendados
- Selección y estructuración de las Unidades Didácticas
- Distribución
- Evaluación
- Bibliografía
- Información
 - Profesorado
 - Materiales docentes

Plataforma Drupal/HTML:

- Ejercicios/problemas
- Apuntes
- Imágenes
- Otros
- Trabajos de asignatura
- Guía de estudio
 - Objetivos de la asignatura
 - Objetivos de aprendizaje
 - Metodología docente
 - Temario
 - Planificación de actividades
 - Sistema de calificación
 - Normas de las actividades
- Materiales del curso
 - Apuntes
 - Audiovisual
 - Trabajos de asignatura
 - Practicas
 - Otros
- Otros recursos
 - Bibliografía

Anexo 5

Manual de Instalación de Scrapy en Windows

1. Instalamos [Python](#) descargando la versión que sea compatible con Scrapy como son 2.5, 2.6, 2.7.
2. Se deberá cambiar las variables de entorno del sistema mediante: Iniciar -> Panel de Control -> Sistema -> Acceso remoto -> Variables de Entorno. En la que se configurara la ruta de instalación de Python C:\Python26 dependiendo de la versión a instalar.

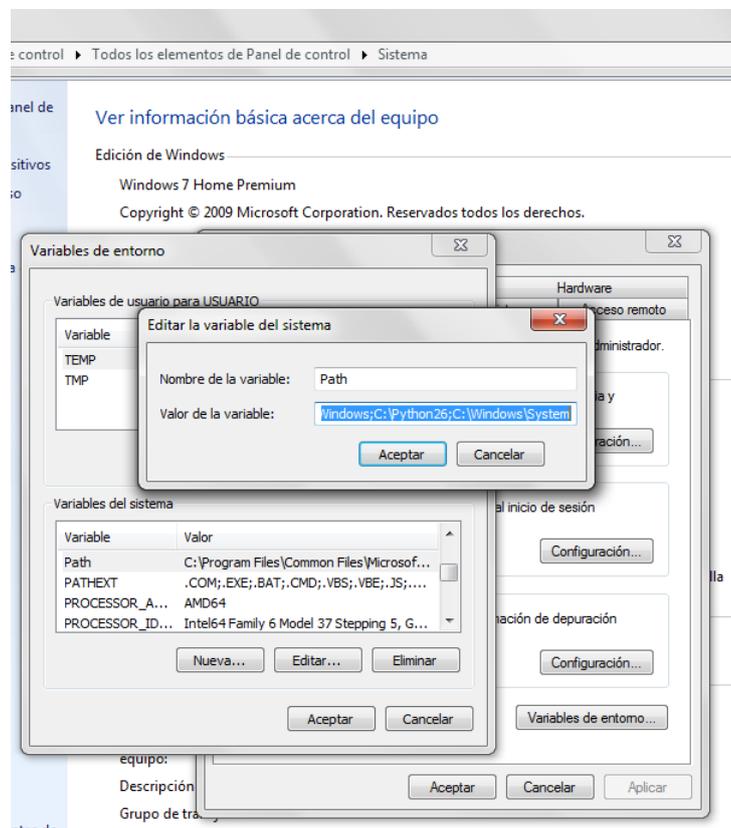


Fig.1. Variables de Entorno

3. El primer de los requerimientos a Instalar es [pywin32](#) pues es un requerimiento de [Twisted](#), evitando así errores.

Nota: Se deberá elegir la versión que vaya de acuerdo a la que se instalo de Python, en mi caso instale la versión 2.6.

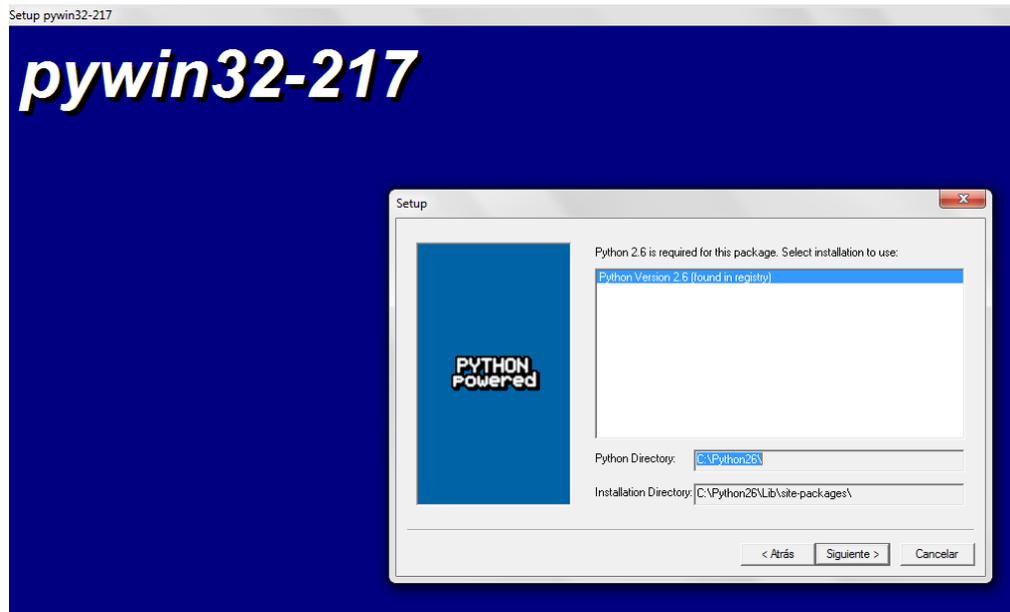


Fig2.Instalación de Pywin32

4. Instalaremos de la misma manera [Twisted](#).

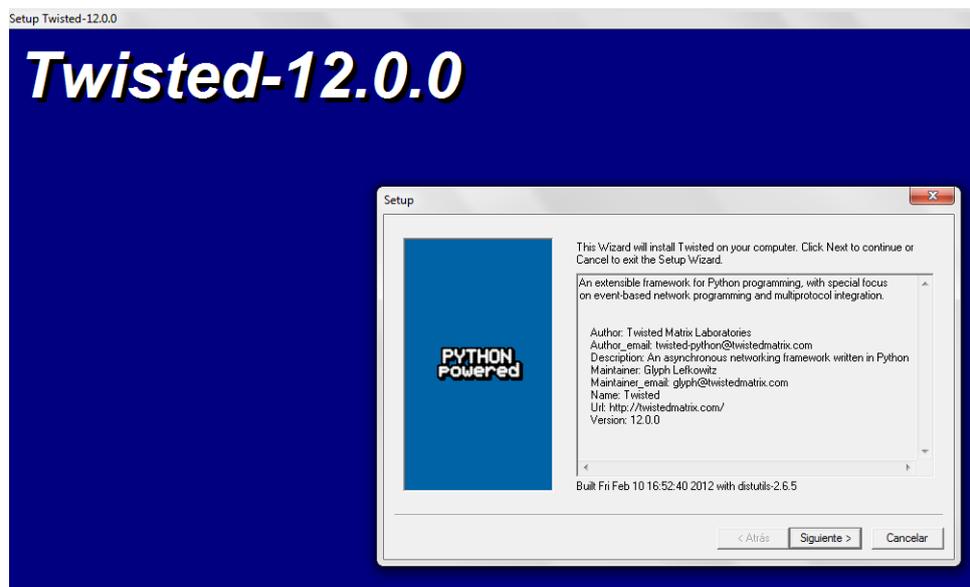
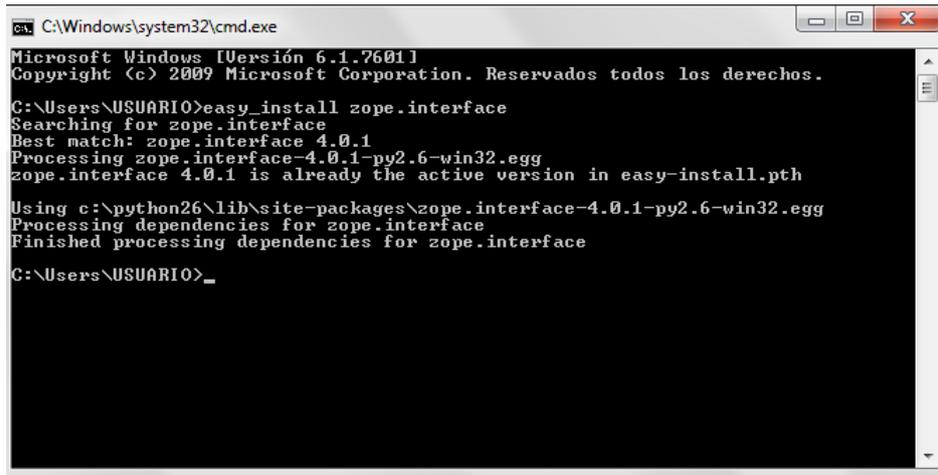


Fig3. Instalación Twisted

5. Se procederá a instalar [Zope.Interface](#) que es requerida por Twisted. En el caso de no encontrar instalador.exe se procederá la de la manera mas sencilla mediante la sentencia `easy_install zope.interface`. Esta instalación se realizara en el c:/



```
C:\Windows\system32\cmd.exe
Microsoft Windows [Versión 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. Reservados todos los derechos.

C:\Users\USUARIO>easy_install zope.interface
Searching for zope.interface
Best match: zope.interface 4.0.1
Processing zope.interface-4.0.1-py2.6-win32.egg
zope.interface 4.0.1 is already the active version in easy-install.pth

Using c:\python26\lib\site-packages\zope.interface-4.0.1-py2.6-win32.egg
Processing dependencies for zope.interface
Finished processing dependencies for zope.interface

C:\Users\USUARIO>
```

Fig4. Instalación Zope.interface

Nota: Puede que al momento de ingresar el comando obtenga un mensaje como "easy_install" no se reconoce como un comando interno o externo, programa o archivo por lotes ejecutable.

Para solucionar dicho problema deberá ingresar nuevamente y editar la variable Path agregando: C:\Windows\System; C:\Windows; C:\Windows\System32, C:\Python26\Scripts; ya que por lo general esta variable tomara por defecto solo alguna de estas direcciones y reiniciar la máquina.

En caso de no poder solucionar el problema puede dirigirse a: <http://blogtaco.blogspot.com/2011/10/usar-easyinstall-de-python-en-windows.html>

6. Instalar [libxml2](#) se puede realizar de 2 maneras:

- a. Por medio del comando `easy_install` de la línea de comandos.
- b. Si no se encontró un instalador [.exe](#) se deberá descargar dos librerías, la `iconv.h` de la que depende `libxml2` y además `zlib`.

Debemos descomprimir las carpetas en C:\libxml2, C:\iconv, C:\zlib; internamente cada una de estas contienen en el bin un `.dll` los cuales deberán ser copiados en la carpeta C:\WINDOWSSystem32.

La compilación de este proceso se realizara por medio de:

```
gcc -o main main.c -IC:libxml2include -IC:libxml2lib -IC:iconvinclude -  
LC:libxml2include -LC:libxml2bin -LC:libxml2lib -LC:iconvinclude -  
lxml2 -lm
```

7. Se procederá descargar [PyOpenSSL](#) para Windows o también desde consola con el comando `easy_install` se podrá conseguir el mismo resultado.

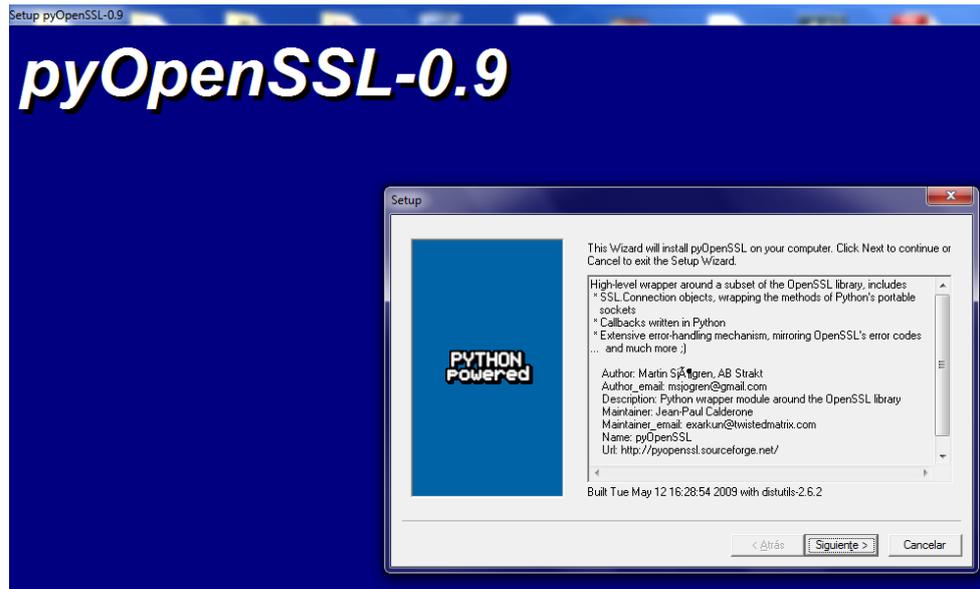


Fig5. Instalación PyOpenssl

8. Finalmente procedemos a descargar [Scrapy.exe](#) o también por medio del comando `easy_install` este ultimo lo recomiendo pues podría haber alguna dependencia que no se tomo en cuenta y por medio de este comando la instalación es automática y no tendremos líos de compatibilidad.

```
ca. C:\Windows\system32\cmd.exe  
Copyright (c) 2009 Microsoft Corporation. Reservados todos los derechos.  
C:\Users\USUARIO>easy_install zope.interface  
Searching for zope.interface  
Best match: zope.interface 4.0.1  
Processing zope.interface-4.0.1-py2.6-win32.egg  
zope.interface 4.0.1 is already the active version in easy-install.pth  
Using c:\python26\lib\site-packages\zope.interface-4.0.1-py2.6-win32.egg  
Processing dependencies for zope.interface  
Finished processing dependencies for zope.interface  
C:\Users\USUARIO>easy_install scrapy  
Searching for scrapy  
Best match: scrapy 0.14.4  
Processing scrapy-0.14.4-py2.6.egg  
scrapy 0.14.4 is already the active version in easy-install.pth  
Installing scrapy script to C:\Python26\Scripts  
Installing scrapy.bat script to C:\Python26\Scripts  
Using c:\python26\lib\site-packages\scrapy-0.14.4-py2.6.egg  
Processing dependencies for scrapy  
Finished processing dependencies for scrapy  
C:\Users\USUARIO>
```

Fig 6. Instalación de Scrapy

9. Para mayor comodidad se procedió a instalar Ipython el cual proporciona facilidad al momento de generar sentencia sin complicarnos la tabulación de la misma.
 - a. Debemos instalar [PyReadline](#) para tener funcionalidades como el autocompletado con la tecla Tab.

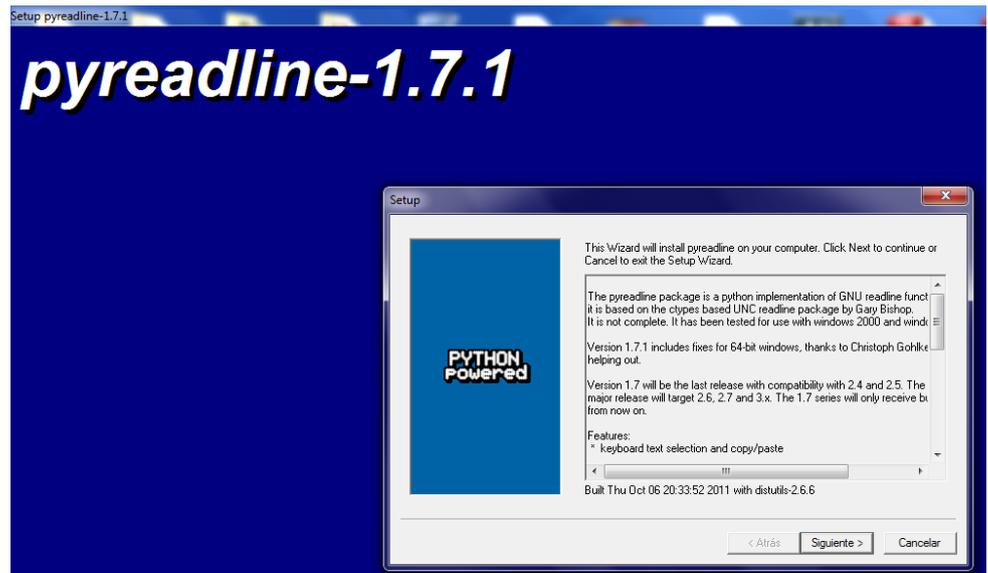


Fig 7. Instalación pyreadline

- b. Finalmente procederemos a Instalar [Ipython](#)

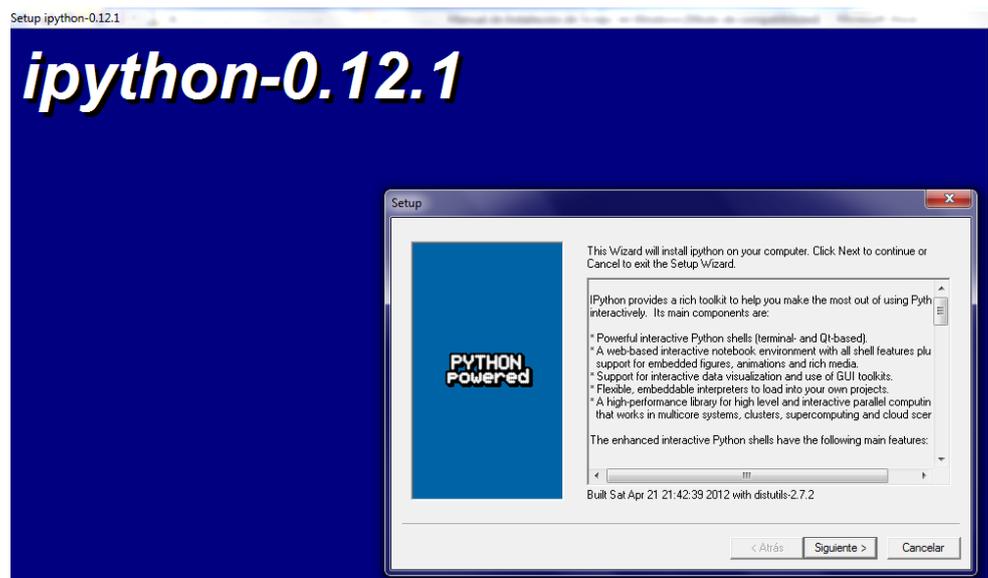
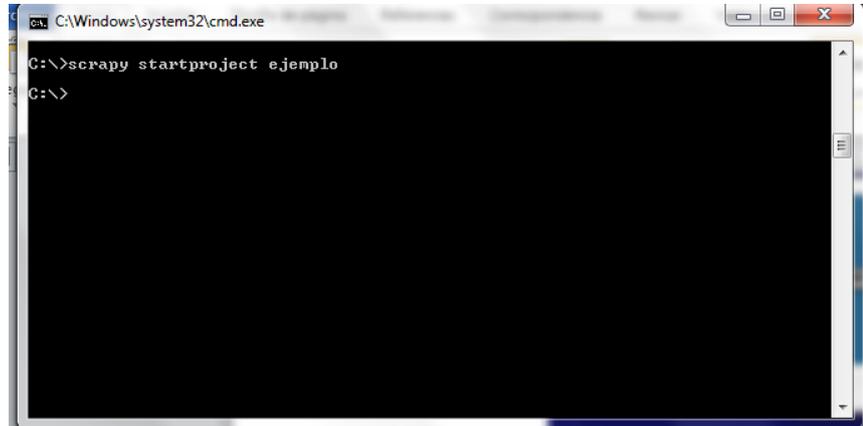


Fig 8. Instalación ipython

10. Procederemos a trabajar en el Shell de scrapy:

- a. Creamos un proyecto por medio del comando **scrapy startproject ejemplo**

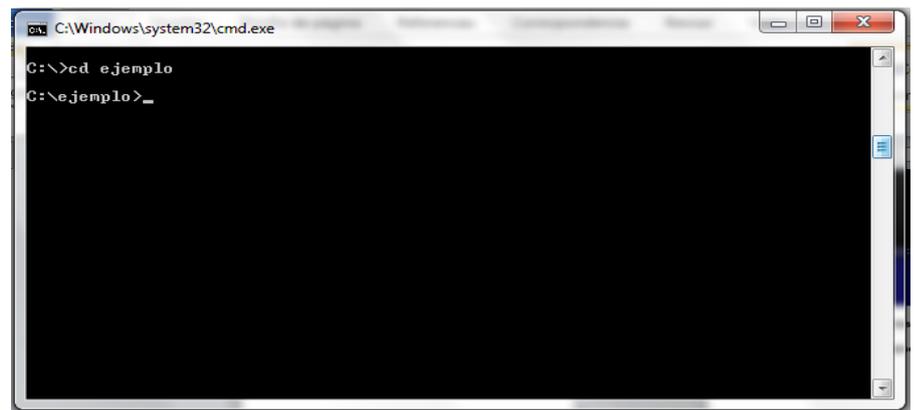


```
C:\Windows\system32\cmd.exe
C:\>scrapy startproject ejemplo
C:\>
```

- b. Una vez creado el proyecto procederemos a verificar le funcionamiento del scrapy por medio del Shell.

Ejemplo de scrapy

1. Ingresaremos al proyecto pues desde este nivel podremos correr el Shell.



```
C:\Windows\system32\cmd.exe
C:\>cd ejemplo
C:\ejemplo>_
```

2. Por medio del comando **scrapy shell** procederemos a analizar el sitio OCW y podrá empezar a realizar sus ejemplos.

```

C:\Windows\system32\cmd.exe - scrapy shell http://ocw.utpl.edu.ec
2012-06-05 10:17:46-0500 [scrapy] INFO: Scrapy 0.14.4 started (bot: tutorial)
2012-06-05 10:17:46-0500 [scrapy] DEBUG: Enabled extensions: TelnetConsole, Clo
eSpider, Webservice, CoreStats, SpiderState
2012-06-05 10:17:46-0500 [scrapy] DEBUG: Enabled downloader middlewares: HttpAut
hMiddleware, DownloadTimeoutMiddleware, UserAgentMiddleware, RetryMiddleware, De
faultHeadersMiddleware, RedirectMiddleware, CookiesMiddleware, HttpCompressionMi
ddleware, ChunkedTransferMiddleware, DownloaderStats
2012-06-05 10:17:46-0500 [scrapy] DEBUG: Enabled spider middlewares: HttpErrorMi
ddleware, OffsiteMiddleware, RefererMiddleware, UrlLengthMiddleware, DepthMiddle
ware
2012-06-05 10:17:46-0500 [scrapy] DEBUG: Enabled item pipelines:
2012-06-05 10:17:46-0500 [scrapy] DEBUG: Telnet console listening on 0.0.0.0:602
3
2012-06-05 10:17:46-0500 [scrapy] DEBUG: Web service listening on 0.0.0.0:6080
2012-06-05 10:17:46-0500 [default] INFO: Spider opened
2012-06-05 10:17:49-0500 [default] DEBUG: Retrying <GET http://ocw.utpl.edu.ec>
<failed 1 times>: 500 Internal Server Error
2012-06-05 10:17:50-0500 [default] DEBUG: Crawled <200> <GET http://ocw.utpl.edu
.ec> (referer: None)
[!] Available Scrapy objects:
[!] hxs      <HtmlXPathSelector xpath=None data=u'<html xmlns="http://www.w3
.org/1999/xhtml">
[!] item     {}
[!] request  <GET http://ocw.utpl.edu.ec>
[!] response <200 http://ocw.utpl.edu.ec>
[!] settings <CrawlerSettings module=<module 'tutorial.settings' from 'C:\tu
torial\tutorial\settings.pyc'>>
[!] spider   <BaseSpider 'default' at 0x3961b30>
[!] Useful shortcuts:
[!] shelp()   Shell help (print this help)
[!] fetch(req_or_url) Fetch request (or URL) and update local objects
[!] view(response) View response in a browser
Python 2.6 (r26:66721, Oct 2 2008, 11:35:03) [MSC v.1500 32 bit (Intel)]
Type "copyright", "credits" or "license" for more information.

IPython 0.12.1 -- An enhanced Interactive Python.
?          -> Introduction and overview of IPython's features.
%quickref  -> Quick reference.
help       -> Python's own help system.
object?    -> Details about 'object', use 'object??' for extra details.

In [1]:

```

c. Este es un ejemplo básico de la obtención de resultados

```

C:\Windows\system32\cmd.exe - scrapy shell http://ocw.utpl.edu.ec
[!] shelp()   Shell help (print this help)
[!] fetch(req_or_url) Fetch request (or URL) and update local objects
[!] view(response) View response in a browser
Python 2.6 (r26:66721, Oct 2 2008, 11:35:03) [MSC v.1500 32 bit (Intel)]
Type "copyright", "credits" or "license" for more information.

IPython 0.12.1 -- An enhanced Interactive Python.
?          -> Introduction and overview of IPython's features.
%quickref  -> Quick reference.
help       -> Python's own help system.
object?    -> Details about 'object', use 'object??' for extra details.

In [1]: datos=hxs.select('//div[@class="visualPadding"]/div/dl[@id="portlet-simple-nav"]/dd/a')

In [2]: for d in datos:
...:     print d.select("text()").extract()[0]
...:
Economía
Educación Continua
Ingeniería Civil
Instituto de Pedagogía
Sistemas Informáticos y Computación

In [3]:

```

Anexo 6

RDF OCW/OER con información.

```
<RDF:RDF
  xmlns:RDF="http://www.w3.org/1999/02/22-RDF-syntax-ns#"
  xmlns:j.0="http://xmlns.com/foaf/0.1/"
  xmlns:vcard="http://www.w3.org/2001/vcard-RDF/3.0#"
  xmlns:OCW="http://www.utpl.edu.ec/OERvocab#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rss="http://purl.org/rss/1.0/"
  xmlns:j.1="http://www.w3.org/2003/01/geo/wgs84_pos#" >
<RDF:Description RDF:about="http://OCW.mit.edu/index.htm">
  <OCW:logo>http://OCW.mit.edu/images/OCW_mast.png</OCW:logo>
  <OCW:is_member_of RDF:resource="http://www.openedconsortium.org"/>
  <OCW:organized_by RDF:resource="http://www.utpl.edu.ec/OERvocab#Aeronautics
and Astronautics"/>
  <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#OCWSite"/>

<OCW:dbpediaLanguage>http://dbpedia.org/resource/English_language</OCW:dbpedi
aLanguage>
  <OCW:licensename>Creative Commons License</OCW:licensename>
  <rss:link>http://OCW.mit.edu/rss/all/</rss:link>
  <OCW:plataform>Desarrollo Propio</OCW:plataform>
  <j.0:name>MIT OpenCourseWare | Free Online Course Materials</j.0:name>
  <OCW:licenselink>http://creativecommons.org/licenses/by-nc-
sa/3.0/us/deed.en_US</OCW:licenselink>
  <OCW:is_member_of RDF:resource="http://OCW.mit.edu"/>
  <OCW:language_alternative>No Information</OCW:language_alternative>
  <dc:source>http://OCW.mit.edu/index.htm</dc:source>
  <dc:language>English</dc:language>
</RDF:Description>
```

```

<RDF:Description RDF:about="http://www.openedconsortium.org">
  <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#Consortium"/>
  <j.0:name>Open Education Consortium</j.0:name>
  <dc:description>The OpenCourseWare Consortium is a worldwide community of
hundreds of higher education institutions and associated organizations committed to
advancing open education and its impact on global education</dc:description>
  <dc:source>http://www.openedconsortium.org</dc:source>
</RDF:Description>
<RDF:Description RDF:about="http://dbpedia.org/page/North_America">
  <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#Continent"/>
  <j.0:name>North America</j.0:name>
  <dc:source>http://dbpedia.org/page/North_America</dc:source>
</RDF:Description>
<RDF:Description RDF:about="http://OCW.mit.edu">
  <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#Consortium"/>
  <dc:description>MIT OpenCourseWare (OCW) is a Web-based publication of virtually
all MIT course content. OCW is open and available to the world and is a permanent MIT
activity.</dc:description>
  <dc:source>http://OCW.mit.edu</dc:source>
  <j.0:name>MIT OpenCourseWare</j.0:name>
</RDF:Description>
<RDF:Description      RDF:about="http://OCW.mit.edu/courses/aeronautics-and-
aeronautics/16-00-introduction-to-aerospace-engineering-and-design-spring-2003/">
  <OCW:classified_by  RDF:resource="http://www.utpl.edu.ec/OERvocab#Aeronautics
and Astronautics"/>
  <OCW:contains      RDF:resource="http://OCW.mit.edu/courses/aeronautics-and-
aeronautics/16-00-introduction-to-aerospace-engineering-and-design-spring-
2003/related-resources"/>
  <OCW:provides      RDF:resource="http://www.grc.nasa.gov/WWW/K-
12/FoilSim/index.html"/>
  <OCW:provides RDF:resource="http://onlineethics.org/CMS/glossary/12893.aspx"/>

```

```

    <OCW:contains      RDF:resource="http://OCW.mit.edu/courses/aeronautics-and-
astronautics/16-00-introduction-to-aerospace-engineering-and-design-spring-
2003/calendar"/>
    <j.0:name>No Information</j.0:name>
    <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#Course"/>
    <dc:source>http://OCW.mit.edu/courses/aeronautics-and-astronautics/16-00-
introduction-to-aerospace-engineering-and-design-spring-2003/</dc:source>
    </RDF:Description>
    <RDF:Description      RDF:about="http://OCW.mit.edu/courses/aeronautics-and-
astronautics/16-00-introduction-to-aerospace-engineering-and-design-spring-
2003/related-resources/">
    <OCW:accommodate      RDF:resource="http://www.grc.nasa.gov/WWW/K-
12/FoilSim/index.html"/>
    <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#Menu"/>
    <j.0:name>Related Resources</j.0:name>
    </RDF:Description>
    <RDF:Description RDF:about="http://dbpedia.org/page/Cambridge,_Massachusetts">
    <OCW:belonging_to RDF:resource="http://dbpedia.org/page/United_States"/>
    <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#City"/>
    <OCW:state>Massachusetts</OCW:state>
    <j.0:name>Cambridge</j.0:name>
    <dc:source>http://dbpedia.org/page/Cambridge,_Massachusetts</dc:source>
    </RDF:Description>
    <RDF:Description      RDF:about="http://OCW.mit.edu/courses/aeronautics-and-
astronautics/16-00-introduction-to-aerospace-engineering-and-design-spring-
2003/calendar/">
    <OCW:accommodate      RDF:resource="http://www.grc.nasa.gov/WWW/K-
12/FoilSim/index.html"/>
    <j.0:name>Calendar      &#xD;</j.0:name>
    <OCW:accommodate
RDF:resource="http://onlineethics.org/CMS/glossary/12893.aspx"/>
    <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#Menu"/>
    <j.0:name>Calendar</j.0:name>

```

```

</RDF:Description>
<RDF:Description RDF:about="http://www.grc.nasa.gov/WWW/K-12/FoilSim/index.html">
  <OCW:metadata_value>text/html;charset=iso88591</OCW:metadata_value>
  <OCW:metadata_type>Content-Type</OCW:metadata_type>
  <OCW:metadata_value>FoilSim Download</OCW:metadata_value>
  <OCW:metadata_type>titulo</OCW:metadata_type>
  <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#OER"/>
  <j.0:name>FoilSim</j.0:name>
</RDF:Description>
<RDF:Description RDF:about="http://dbpedia.org/page/United_States">
  <OCW:corresponds_to RDF:resource="http://dbpedia.org/page/North_America"/>
  <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#Country"/>

<OCW:dbpediaCapital>http://dbpedia.org/page/Washington,_D.C.</OCW:dbpediaCapital>
  <OCW:capital>Washington D. C.</OCW:capital>
  <j.0:name>United States</j.0:name>
</RDF:Description>
<RDF:Description RDF:about="http://www.utpl.edu.ec/OERvocab#Aeronautics and Astronautics">
  <OCW:stored RDF:resource="http://OCW.mit.edu/courses/aeronautics-and-astronautics/16-00-introduction-to-aerospace-engineering-and-design-spring-2003"/>
  <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#KnowledArea"/>
  <j.0:name>Aeronautics and Astronautics</j.0:name>
</RDF:Description>
<RDF:Description RDF:about="http://onlineethics.org/CMS/glossary/12893.aspx">
  <OCW:metadata_value>No Information</OCW:metadata_value>
  <OCW:metadata_type>keywords</OCW:metadata_type>
  <OCW:metadata_value>OEC CHALLENGE STUDY</OCW:metadata_value>
  <OCW:metadata_type>titulo</OCW:metadata_type>
  <RDF:type RDF:resource="http://www.utpl.edu.ec/OERvocab#OER"/>

```

```

    <j.0:name>Challenger Case Study</j.0:name>
  </RDF:Description>
  <RDF:Description RDF:about="http://Web.mit.edu/">
    <vcard:Street>77 Massachusetts Ave, Cambridge, MA 02139, Estados
    Unidos</vcard:Street>

  <OCW:dbpediaLink>http://dbpedia.org/page/Massachusetts_Institute_of_Technology</
  OCW:dbpediaLink>
    <RDF:type
  RDF:resource="http://www.utpl.edu.ec/OERvocab#EducationalOrganization"/>
    <j.0:name>MIT - Massachusetts Institute of Technology</j.0:name>
    <OCW:is_locate_in
  RDF:resource="http://dbpedia.org/page/Cambridge,_Massachusetts"/>
    <dc:source>http://Web.mit.edu/</dc:source>
    <OCW:logo>http://Web.mit.edu/img/MIT_logo.gif</OCW:logo>
    <j.1:latitude>42.35982</j.1:latitude>
    <j.1:longitude>-71.09211</j.1:longitude>
    <dc:description>The mission of MIT is to advance knowledge and educate students in
    science, technology and other areas of scholarship that will best serve the nation and the
    world in the 21st century — whether the focus is cancer, energy, economics or
    literature</dc:description>
  </RDF:Description>
</RDF:RDF>

```

Anexo 6

Datos Obtenidos de sitios OCW

Universidad	Cursos	OER
http://Web.mit.edu/	2281	59016
http://www.open.ac.uk/	509	0
http://www.uci.edu/	365	0
http://www.uct.ac.za/	283	634
http://www.uoc.edu/portal/ca/index.html	275	274
http://www.us.es/	243	7
http://www.nottingham.ac.uk/	234	71
http://uc3m.es/	232	12746
http://www.avu.org/	201	434
http://www.um.es/	180	5110
http://www.ua.es/es/index.html	162	0
http://www.kyoto-u.ac.jp/ja	153	3
http://www.unican.es/index.html	140	761
http://www.upm.es/institucional	133	3503
http://www.uv.es/es	130	1387
http://www.camins.upc.edu/	121	0
http://www.upv.es/	118	0
http://www.hokudai.ac.jp/	114	402
http://umich.edu/	114	3528
http://www.metu.edu.tr/	111	2241
http://www.jhsph.edu/	107	2841
http://www.korea.edu/	99	1697
http://www.uca.es/es/	99	585
http://www.tsukuba.ac.jp/	92	131
http://www.unc.edu.ar/	87	56
http://www.agh.edu.pl/	86	66
http://yct.acad.ncku.edu.tw/bin/home.php	65	754
http://www.unizar.es/	63	1421
http://hanyangOCW.hanyang.ac.kr/class	61	0
http://www.nd.edu/	54	1145
w4.utmspace.utm.my/	48	0

http://www.umb.edu/	46	1245
http://www.udl.cat/	41	493
http://www.udem.edu.mx/Esp/Paginas/default.aspx	39	1702
http://portal.fgv.br/	37	0
http://unu.edu/	34	195
http://www.anahuac.mx/Pages/NvoHome.aspx	32	0
http://www.unia.es/	31	107
http://www.usal.es/Webusal/	30	348
http://www.uni.edu.pe/	26	16
http://gunadarma.ac.id/	26	1
http://www-e.openu.ac.il/	25	0
http://www.univalle.edu.co/	23	195
http://www.ugr.es/	23	12
http://www.capilanou.ca/	23	24
http://www.unav.edu/	22	925
http://portal.uned.es/	21	18
http://www.pucv.cl/	21	42
http://www.uhu.es/index.php	21	16
http://www.kyoto-seika.ac.jp/	20	0
http://www.njit.edu/	20	335
http://www.tudelft.nl/	15	0
http://www.nctu.edu.tw/	15	0
http://www.ucv.ve/	15	0
http://www.tmu.edu.tw/v3/main.php	15	0
http://www.uib.es/es/	14	185
http://www.utpl.edu.ec/	13	328
http://www.uab.es/	13	106
http://www.flacso.edu.mx/	13	18
http://www.uis.edu.co/WebUIS/es/index.jsp	12	17
http://www.uned.ac.cr/	12	32
http://www.marhi.ru/	12	24
http://www.iimb.ernet.in/	11	1
http://www.unsa.edu.pe/	11	0
http://www.itesm.mx/	11	0
http://www.titech.ac.jp/english/index.html	10	0
http://www.kyushu-u.ac.jp/	10	69

http://www.uc.cl/	10	0
http://www.uchile.cl/	10	0
http://www.ie.edu/	10	0
http://en.sbu.ac.ir/pages/home.aspx	10	48
http://www.uwc.ac.za/Pages/default.aspx	9	9
http://OCWWeberstateuniversity.blogspot.com/	7	24
http://www.icesi.edu.co/	7	354
https://www.uam.es/ss/Satellite/es/home.htm	7	112
https://www.kaplanuniversity.edu/home.aspx	5	1
http://ujiapps.uji.es/	5	56
http://www.wheelock.edu/	4	32
http://www.ceu.es/	4	2
http://www.ehu.es/	3	0
http://www.paristech.fr/	3	0
http://usu.ac.id/en/	2	24
http://www.upct.es/	2	0
http://www.uniovi.es/	1	0
http://uci.edu/	1	0
Total	7823	105929

Anexo 7

Plataformas empleadas en sitios OCW.

1. EduCommons⁵⁴

Esta plataforma es un sistema de gestión de contenido diseñado específicamente para apoyar proyectos de OpenCourseWare.

Este software fue creado por la Universidad del Estado de Utah

Características:

- Una de sus principales características es su configuración sencilla y flexible.
- Estructura en el flujo de trabajo.
- Incluye mecanismos de gestión de propiedad intelectual y algunos derechos de utilización como Creative Commons.
- Es compatible con estándares e-learning.
- Soporte de edición de HTML.
- Contenido RDF, basado en estándares Dublín Core.
- Basado en Zope (servidor de aplicaciones orientado a objetos, OpenSource) y Plone (sistema de administración de contenido, CMS).
- Cabeceras RDF basados en el estándar Dublin Core.
- Sindicación RSS, versión expandida de RSS en EduCommons
- Enlaces a marcadores sociales, posibilidad de adaptación.
- Exportación/Importación de paquetes IMS.
- Exportación/Importación a WordPress.

2. Moodle⁵⁵

⁵⁴ <http://edcommons.com/>

Moodle es un sistema de gestión de cursos (CMS), también conocido como un Learning Management System (LMS) o un Entorno Virtual de Aprendizaje (EVA). Se trata de una aplicación Web gratuita que los educadores pueden utilizar para crear eficaces sitios de aprendizaje en línea.

Es un paquete de software para producir cursos basados en internet; considerada como un entorno de aprendizaje virtual. Distribuida gratuitamente con Open source y con una licencia GNU.

Su denominación viene dada por Modular-Object-Oriented-Dynamic Learning by Education. Esta plataforma además cuenta con derechos de autor Copyright.

Esta plataforma es una de las más usadas a nivel de colaboración de información de tipo educativo y de aprendizaje en línea. Por lo que algunos sitios OCW han sido implementados en esta plataforma acoplándola de tal manera que la compartición de los recursos sea abierta.

La dificultad que muestra esta plataforma es que no se apega mucho al concepto de libertad en el sentido que cuenta con acceso restringido por medio de un login. En algunos sitios OCW se ha logrado trabajar con módulos que ayuden a mejorar estos aspectos pero la mayoría de los que usan esta plataforma aun trabajan de una manera básica y cuentan con un usuario denominado Invitado para el ingreso al área de recursos.

Dificultades

A continuación se mencionan ciertas dificultades que se tenían en versiones antiguas, pero poco a poco se han ido mejorando conforme se acoplaban a los sitios OCW:

⁵⁵ <http://moodle.org/>

- La mayor dificultad es hacer público o privado un curso.
- Añadir metadatos.
- Flujo de trabajo.
- Soporte RSS.

3. Drupal ⁵⁶

Drupal es un sistema de gestión de contenido modular multipropósito y muy configurable de código abierto que permite publicar artículos, imágenes, u otros archivos y servicios añadidos como foros, encuestas, blogs y administración de usuarios y permisos.

Es un software Open Source con licencia GNU/GPL multiplataforma, multilinguaje, modular, Entendible, Flexible y de contenido; creado por Dries Buytaert en el 2001 siendo así utilizado por algunos sitios OCW.

Características

- Gran gestor de contenidos: como son control de versiones, objetos, plantillas y formatos RDF/RSS.
- Amplio marco de trabajo.
- Categorización y organización de contenido.
- Contenido en varios formatos.
- Privacidad y control de acceso.
- Apoyo al campo educativo E-learning.
- Incluye soporte a búsqueda Semántica.

Su poco uso en la creación de estos sitios se debe a que su aprendizaje es difícil, su interfaz es poco amigable y cuenta con compatibilidad inversa.

4. WordPress ⁵⁷

⁵⁶ <http://drupal.org/>

Es una plataforma libre y gratuita, diseñada como un sistema de gestión de contenido semántico.

Características:

- Queda a su criterio que los datos sean privados o públicos en opciones de privacidad.
- Contiene estadísticas integradas.
- Fácil Edición.
- Control absoluto sobre Spam.
- Seguridad Integrada.
- Variedad de Idiomas (50)
- Permite crear paquetes de Web estática.
- Contiene un área de contenido actual en organización por categorías.
- Es compatible con archivos Feeds (RDF, RSS,ATOM).
- Cuenta con licencia GPL.
- Soporte PHP y MySQL
- Núcleo flexible.
- Fácil Instalación y actualización.
- Generación dinámica de páginas.
- Extensiones y plugins necesarios.

⁵⁷ <http://es.wordpress.org/>