



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

ÁREA TÉCNICA

TITULACIÓN DE INGENIERO EN INFORMÁTICA

Extracción de información de uso de cursos OpenCourseWare

TRABAJO DE FIN DE TITULACIÓN

AUTORAS: Molina Sandoval, Diana Maricela
Soto Fuertes, Paula Elizabeth

DIRECTOR: Chicaiza Espinosa, Janneth Alexandra Ing.

CENTRO UNIVERSITARIO LOJA

2014

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN

Ingeniera.

Janneth Alexandra Chicaiza Espinosa.

DOCENTE DE LA TITULACIÓN

De nuestras consideraciones:

El presente trabajo de fin de titulación: denominado “Extracción de información de uso de cursos OpenCourseWare” realizado por las profesionales en formación: Molina Sandoval Diana Maricela y Soto Fuertes Paula Elizabeth; ha sido orientado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, Abril de 2014

f.-----

Ing. Janneth Chicaiza E.
DIRECTORA DE TESIS

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

"Nosotras Molina Sandoval Diana Maricela y Soto Fuertes Paula Elizabeth declaramos ser autoras del presente trabajo de fin de titulación: denominado "Extracción de información de uso de cursos OpenCourseWare" de la Titulación Ingeniería en Informática siendo la Ing. Janneth Alexandra Chicaiza Espinosa directora del presente trabajo; y eximamos expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales.

Además certificamos que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de nuestra exclusiva responsabilidad.

Adicionalmente declaramos 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.-----

Molina Sandoval Diana Maricela
1719591206

f.-----

Soto Fuertes Paula Elizabeth
1103827851

DEDICATORIA

A Dios por darme la oportunidad de vivir y por estar conmigo en cada paso que doy, por derramar todas sus bendiciones, protegerme, fortalecer mi corazón e iluminar mi mente y por haber puesto en mi camino aquellas personas que han sido mi soporte y compañía durante todo el periodo de mis estudios.

A mí amado esposo Ángel y a mis padres Ángel y Gloria por haberme apoyado incondicional, moral y económicamente; a mis niñas Karen y Vivian y a mis hermanos Patricia, Cristian y Lucy por ser el motor de mi vida y haber creído en mí, y por todas las palabras de aliento que me impulsaron y motivaron a seguir adelante en los momentos de flaqueza.

A cada uno de los docentes de la extensión Cariamanga por todos los momentos compartidos y por haberme impartido sus conocimientos, experiencias, pero sobre todo por haberme brindado su amistad sincera.

Con todo cariño para ustedes.

Diana

DEDICATORIA

El invaluable amor y unión de una familia se ve reflejado en el éxito de vida de cada uno de sus miembros; con todo mi amor para mi papi Ángel, mi mami Elsa, Paúl, Israel, Andrea e Israelito, que dedican todo en la vida para que yo logre mis sueños; por motivarme y darme la mano cuando sentía que el camino se terminaba, a ustedes por siempre mi corazón y fortaleza.

Paula

AGRADECIMIENTO

Agradecemos de manera especial y sincera a nuestra directora de tesis, Ing. Janneth Chicaiza Espinosa por su apoyo, confianza y capacidad para guiar nuestras ideas, facilitado siempre los medios suficientes para llevar a cabo todas las actividades propuestas y no cabe duda que su participación ha sido un aporte invaluable, enriqueciendo nuestro conocimiento para el progreso y finalización de este proyecto de Tesis.

A la Lcda. Lidia Villacís por su sólida amistad, apoyo, orientación y eficiencia total con la colectividad de estudiantes universitarios de la modalidad abierta y a distancia de la UTPL; nuestra gratitud por siempre.

Diana Molina Sandoval, Paula Soto Fuentes

ÍNDICE DE CONTENIDOS

CARATULA.....	i
APROBACIÓN DEL DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA	iv
DEDICATORIA	v
AGRADECIMIENTO	vi
ÍNDICE DE CONTENIDOS	vii
RESUMEN.....	- 9 -
ABSTRACT	- 10 -
CAPITULO I: MARCO CONCEPTUAL Y DESCRIPCIÓN DE LA PROPUESTA DE TRABAJO OCW.....	- 11 -
1. Conceptualización y propuesta de trabajo OCW.....	- 12 -
1.1. Presentación de tecnología de la Web Semántica.....	- 12 -
1.1.1. <i>Evolución de la web.</i>	- 12 -
1.1.2. <i>Web Semántica.</i>	- 13 -
1.1.3. <i>Arquitectura de la Web Semántica.</i>	- 15 -
1.1.4. <i>Vocabularios.</i>	- 17 -
1.1.4.1. <i>Vocabularios Consensuados.</i>	- 17 -
1.1.4.2. <i>Iniciativas de vocabularios desarrollados por la UTPL.</i>	- 19 -
1.1.5. <i>Linked Data.</i>	- 20 -
1.1.6. <i>Herramientas para extracción y segmentación de datos desde páginas HTML y publicación de datos.</i>	- 23 -
1.1.6.1. <i>Python.</i>	- 23 -
1.1.6.2. <i>LODSPeaKr.</i>	- 26 -
1.2. OCW	- 27 -
1.2.1. <i>Syllabus.</i>	- 29 -
1.3. Propuesta de trabajo OCW	- 32 -
CAPÍTULO II: DESCRIPCIÓN DE FUENTE DE DATOS.....	- 34 -
2. Recolección de datos	- 35 -
2.1. Universo de datos.....	- 35 -

2.2. Selección de datos	- 36 -
CAPÍTULO III: PREPROCESAMIENTO DE DATOS	- 38 -
3.1. Extracción de datos de cursos OCW	- 39 -
3.2. Segmentación de la data.....	- 44 -
3.3. Categorización de data procesada.....	- 47 -
CAPÍTULO IV: IMPLEMENTACIÓN DE VOCABULARIO DOCW	- 51 -
4.1. Adaptación y modelado de vocabulario para describir planes de OCWs.....	- 52 -
4.2. Generación de datos RDF.....	- 56 -
4.2.1. <i>Creación de aplicativo para generación de fichero RDF.</i>	- 57 -
4.2.2. Validación del código RDF.....	- 60 -
CAPÍTULO V: ALMACENAMIENTO, ACCESO Y VISUALIZACIÓN DE DATOS	- 64 -
5. Almacenamiento, acceso y visualización de datos	- 65 -
5.1. Almacenamiento de tripletas en TripleStore	- 65 -
5.1.1. <i>Explotación de datos mediante lenguaje de consulta SPARQL.</i>	- 66 -
5.2. Visualización de datos RDF	- 75 -
5.2.1. <i>Creación de servicios usando LODSPeakr con la librería D3</i>	- 76 -
5.2.1.1. <i>Creación de Servicios.</i>	- 77 -
5.2.1.2. <i>Aplicaciones creadas con LODSPeakr.</i>	- 78 -
CONCLUSIONES	- 86 -
RECOMENDACIONES	- 87 -
BIBLIOGRAFÍA.....	- 88 -
ANEXOS.....	- 91 -
Anexo 1: MANUAL DE INSTALACIÓN DE PYTHON EN WINDOWS.....	- 91 -
Anexo 2: MANUAL DE INSTALACIÓN DE SCRAPY EN WINDOWS.....	- 95 -
Anexo 3: MANUAL DE INSTALACIÓN DE VIRTUALENV	- 99 -
Anexo 4: MANUAL DE INSTALACIÓN DE WAMPSERVER 2.1	- 101 -
Anexo 5: MANUAL DE INSTALACIÓN DE PROTEGÉ.....	- 105 -
Anexo 6: INSTALACION DE VIRTUOSO EN LINUX	- 106 -
Anexo 7: CÓDIGO FUENTE DEL SERVICIO TOPICCOURSE	- 109 -
Anexo 8: CÓDIGO FUENTE DEL SERVICIO DETAILCOURSE	- 114 -
Anexo 9: CÓDIGO FUENTE DEL SERVICIO DETAILSyllabus	- 117 -
Anexo 10: CONSULTAS SPARQL	- 120 -

RESUMEN

El presente proyecto pretende ampliar el programa general de extracción Linked OCW Data en el que ha venido trabajando el departamento de Tecnologías Avanzadas de la Web y Sistemas Basados en el Conocimiento, centrándose este trabajo en obtener calidad de información pedagógica de los cursos OCW; extrayendo datos de relevancia como objetivos, contenidos, criterios de evaluación, descripción del curso, bibliografía, requisitos, actividades, calendario académico, recursos, políticas y conducta.

Para el efecto se trabajó con técnicas de extracción y segmentación de datos como Scrapy y Leaf respectivamente, métodos basados en el lenguaje de programación Python; categorizada la data se procedió a crear un vocabulario denominado DOCW (Descripción de cursos OCW), donde se reutilizó vocabularios consensuados como BIBO, DC, AIISO, etc. y vocabularios de autoría de la UTPL como LOCWD e IPD.

Se generó un archivo RDF y para la publicación de estos datos se trabajó en el marco de código abierto LODSPeaKr, herramienta que permite la explotación y visualización de datos enlazados.

PALABRAS CLAVES: OCW, Syllabus, Scrapy, Leaf, Python, Vocabularios, RDF, LODSPeaKr, Linked Data.

ABSTRACT

This project aims to extend the general program OCW Linked Data extraction in which the department of Tecnologías Avanzadas de la Web y Sistemas basados en el Conocimiento has been working, this work is focused on getting quality educational information about OCW courses; extracting relevant data such as objectives, contents, evaluation criteria, course description, bibliography, requirements, activities academic calendar, resources, policies and conduct.

For the purpose , it worked with extraction techniques and data segmentation as Scrapy and Leaf respectively, based on Python methods; categorized the data ,the next process was to create a vocabulary called DOCW (description OCW courses), where the main factor was reusing vocabularies as BIBO, DC AIISO etc. and UTPL authoring vocabularies as LOCWD and IPD.

A software application for generating an RDF file was created, for the publication of such data it worked under LODSPeaKr, an open source tool that allows the exploitation and visualization of linked data.

KEY WORDS: OCW, Syllabus, Scrapy, Leaf, Python, Ontology, RDF, LODSPeaKr, Linked Data

**CAPITULO I: MARCO CONCEPTUAL Y DESCRIPCIÓN DE LA PROPUESTA DE
TRABAJO OCW**

1. Conceptualización y propuesta de trabajo OCW

En este apartado se realiza una breve descripción sobre el estado del arte de los principales conceptos que permiten el desarrollo de esta tesis, enfatizando en elementos relacionados a la dimensión tecnológica del proyecto: Linked Data, Web semántica, Extracción de datos desde la web de documentos.

En la sección 1.3 se presenta la propuesta y los procesos desarrollados a lo largo de este proyecto.

1.1. Presentación de tecnología de la Web Semántica.

La web semántica es considerada una extensión de la web actual, en la cual tanto los datos como su definición semántica pueden ser procesados por las personas y por aplicaciones en una computadora.

La web semántica identifica un conjunto de tecnologías, herramientas y estándares que dan soporte a la visión de web asociada con el significado.

1.1.1. *Evolución de la web.*

- **Web 1.0:** Se habla de una web estática sólo de lectura, los usuarios no pueden interactuar con el contenido existente en las mismas; el predominio del texto sobre la imagen; la escasa o nula actualización de los contenidos eran algunas de sus características.
- **Web 2.0:** Nacen las comunidades virtuales, se fomenta la colaboración y el intercambio de información entre los consumidores dando origen a las redes sociales, wikis y blogs; la principal característica de esta web es la interacción y colaboración entre los miembros como creadores de contenido.
- **Web 3.0:** también conocida como web semántica; fundamenta su razón de ser en adaptar la información existente en Internet con contenido propio para que pueda ser procesada por los ordenadores a través de búsquedas más precisas. Referenciando a (Cantera Fonseca, Hierro Sureda, & Romo Zabala, 2007), la web semántica mantiene los principios que han hecho que la web actual sea un éxito, como son los principios

de descentralización, compartición, compatibilidad, máxima facilidad de acceso y contribución, o la apertura al crecimiento y uso no previstos de antemano.

En la Figura 1.1 se puede apreciar la constante evolución de la web por décadas y su posible futuro.

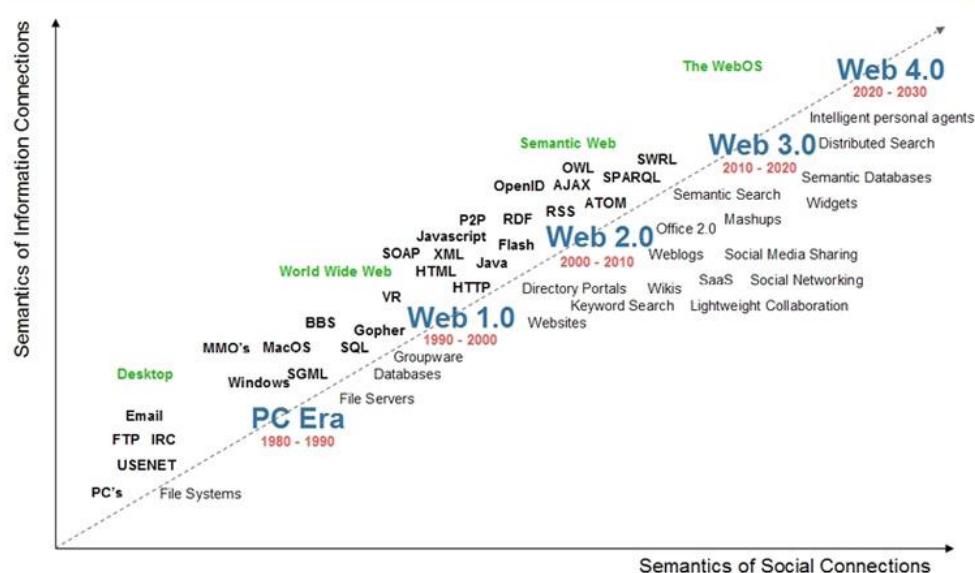


Figura 1.1 Evolución de la web

Fuente: (Radar Networks & Nova Spivack, 2007

http://novaspivack.typepad.com/nova_spivacks_weblog/2007/02/steps_towards_a.html

1.1.2. Web Semántica.

En la actualidad, la World Wide Web¹ está basada principalmente en documentos escritos en HTML²; este lenguaje es válido para aportar el aspecto visual y multimedia de un documento, pero ofrece escasas posibilidades de trabajo inteligente al momento de comprender el significado de toda esa información; esto se debe a que los documentos que se encuentran en la web no comparten un formato común y los ordenadores no pueden procesarlos de manera eficiente.

De acuerdo a la visión de (Berners-Lee, Hendler, & Lassila, The Semantic Web, 2001), la web semántica es una extensión de la web actual en la que la información se proporciona con un significado bien definido para que las computadoras y las personas trabajen mejor en conjunto.

¹ <http://www.w3c.es/>

² <http://www.w3.org/html/>

“El objetivo más general de la Web Semántica es que la información en la Web esté definida y ligada de tal modo que pueda ser utilizada por máquinas no sólo para propósitos de despliegue si no para la automatización, integración y reutilización de los datos entre diferentes aplicaciones”, (López Guzmán & García Peñalvo, 2008).

La Figura 1.2 esquematiza la diferencia entre la web que conecta datos HTML (web de documentos) y la web que conecta significado de datos (web de datos).

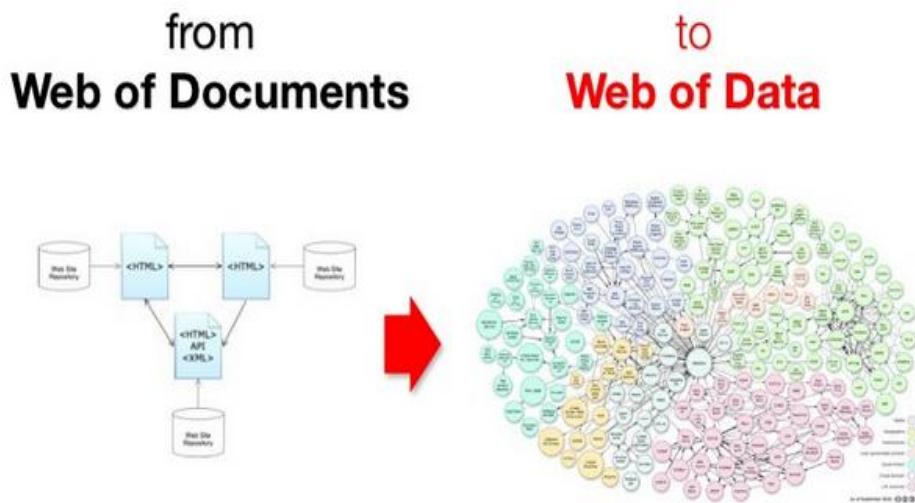


Figura 1.2 De la web de documentos a la web de datos

Fuente: <http://es.slideshare.net/emadridnet/2012-01-20-upm-emadrid-etovar-upm-npiedra-utpl-linked-data-repositorios-ocw>

En el campo educativo la web semántica añade ventajas orientadas principalmente a la clasificación y búsqueda del conocimiento en ambientes de aprendizaje, existe la necesidad de reutilización de la información, para hacer más eficiente el autoaprendizaje.

Las tendencias de desarrollo de la web semántica se centran en tres áreas aplicadas a la educación: la informática, el diseño instruccional y los sistemas de bibliotecas. (Santacruz Valencia , Aedo, & Delgado Kloos, 2004).

Como lo cita (Leguizamo León, 2007), la Web semántica ayuda a las personas a desempeñar sus roles en espacios de aprendizaje grandes, distribuidos, basados en problemas, multi-actores, multi-recursos, crear y modificar ambientes de aprendizaje centrados en el aprendiz, no lineales y autodirigidos.

1.1.3. Arquitectura de la Web Semántica.

La propuesta de desarrollo de la Web Semántica del consorcio W3C³ se sugiere una arquitectura básica en capas (ver Figura 1.3), comenzando por la capa del nivel inferior XML, que permite estructurar sintácticamente los datos, siguiendo por una capa RDF que define la semántica de dichos datos, luego por la capa ontológica que define consensuadamente conceptos y relaciones para distintos dominios y por último, la capa lógica, que define las reglas lógicas y mecanismo para hacer inferencias, (García Ruiz, 2008).

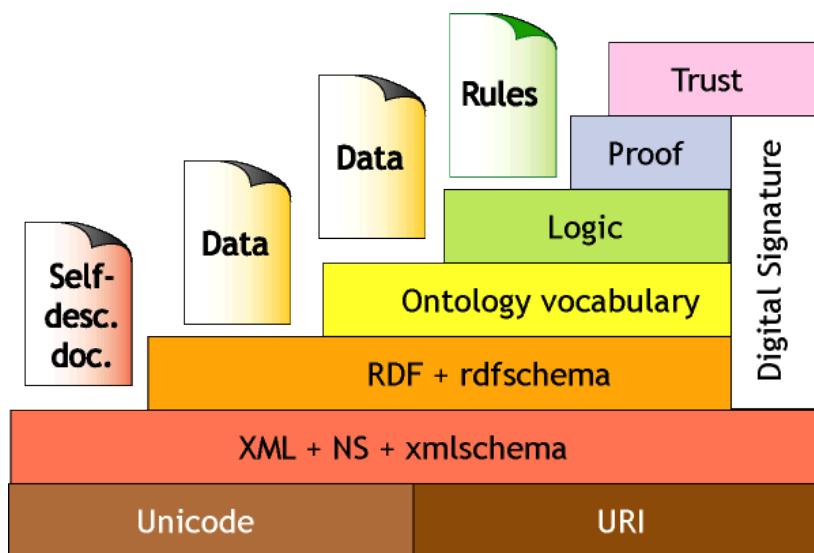


Figura 1.3 Tim Berners-Lee. Semantic Web -XML2000. Architecture
Fuente: <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slides10-0.html>

A continuación se presentan las características de las capas de la estructura:

- **XML**

XML⁴ (Extended Markup Language) se plantea como lenguaje para intercambio de información estructurada entre las diferentes plataformas, incluye etiquetas específicas para diferentes componentes en un documento.

- **XML Schema**

XML Schema⁵ es un lenguaje para definir la estructura de los documentos XML.

³ <http://www.w3.org/>

⁴ <http://www.w3.org/XML/>

⁵ <http://www.w3.org/XML/Schema.html>

- **RDF**

RDF⁶ (Resource Description Framework), en español marco de descripción de recursos; es un modelo estándar para intercambiar datos en la web, este lenguaje describe recursos e identifica únicamente entidades (personas, organizaciones, lugares, cursos...etc) y relaciones sobre las mismas.

Una descripción RDF es un conjunto de proposiciones simples, también conocidas como tripletas⁷. Las tripletas son “piezas” de conocimiento que tienen la siguiente estructura:



Figura 1.4: Tripleta RDF

Fuente: <http://glosariowebsem1a.wikispaces.com/RDF>

- **Sujeto:** indica la entidad sobre la que se describe el hecho.
- **Predicado:** indica el tipo de hecho a describir.
- **Objeto:** indica un valor adicional que ayuda a completar el hecho.

RDF permite la interoperabilidad entre aplicaciones mediante el intercambio de información entendible para el ordenador mediante la web.

RDF tiene varias áreas de aplicación; como la recuperación de recursos educativos, inventario en bibliotecas digitales, vocabularios, agentes inteligentes entre otros.

- **RDF Schema**

RDF SCHEMA⁸ (RDFS), es un lenguaje que complementa el modelo RDF ayudando en la interoperabilidad semántica en el contenido web.

RDFS proporciona información sobre la interpretación de una sentencia dada en un modelo de datos RDF, además facilita la combinación de sentencias RDF por medio de URIs.

- **OWL**

OWL⁹ (Web Ontology Language) está diseñado para ser usado por aplicaciones que necesiten procesar el contenido de la información. OWL permite la interoperabilidad entre

⁶ <http://blog.classora.com/2012/10/10/describiendo-el-conocimiento-en-un-formato-estandar-para-la-web-semantica-rdf/>

⁷ http://es.wikipedia.org/wiki/Resource_Description_Framework

⁸ <http://www.w3.org/TR/rdf-schema/>

ontologías¹⁰ proporcionando de vocabulario adicional junto con una semántica formal, proporciona clases, propiedades, personas y los valores de datos y se almacenan como documentos de la Web Semántica.

1.1.4. *Vocabularios.*

En Web Semántica, los vocabularios¹¹ definen los conceptos y relaciones que describen y representan un área de conocimiento. Los vocabularios son utilizados para clasificar los términos que se pueden utilizar en una aplicación particular, caracterizar posibles relaciones, y definir posibles restricciones sobre el uso de esos términos.

Los conceptos y entidades del mundo académico como universidades, material educativo, áreas de conocimiento, cursos on-line, Syllabus, etc. deberían estar descritos mediante vocabularios consensuados, y los datos deberían estar representados mediante lenguajes semánticos; de esta manera, un agente inteligente podría procesar y entender el contenido.

1.1.4.1. *Vocabularios Consensuados.*

Como lo cita (Chicaiza, Piedra, & López, 2011), la definición, construcción y uso de vocabularios consensuados para describir conceptos del mundo real, permite entre otras, las siguientes ventajas:

- Habilita el intercambio, la reutilización y la compartición de recursos a diferentes niveles: institucional, gubernamental, entre aplicaciones de software, etc.
- Facilita el acceso y localización de recursos en la Web, pues agentes de búsqueda podrán comprender e interpretar el significado del contenido, y así emparejar de mejor manera con las necesidades de información de los usuarios.

A continuación se describen los vocabularios utilizados a lo largo de este proyecto.

⁹ <http://www.w3.org/TR/owl-features/>

¹⁰ Ontología: Es la especificación explícita de una conceptualización.

[<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/1861/1/42897.pdf>]

¹¹ <http://www.w3.org/standards/semanticweb/ontology>

- **Dublin Core (dc, terms)**

Dublin Core¹² es un conjunto de elementos de metadatos que incluye todos los términos DCMI¹³ (es decir, los refinamientos, esquemas de codificación, y los términos de vocabulario controlado) destinadas a facilitar el descubrimiento de recursos.

(Lamarca, 2011), señala que los metadatos Dublin Core tratan de ubicar, dentro de Internet, los datos necesarios para describir, identificar, procesar, encontrar y recuperar un documento introducido en la red.

- **Friend – Of – A – Friend (FOAF)**

FOAF¹⁴ tiene como fin crear una web legible por máquina, para describir gente y enlazar a dichas personas con las cosas que han creado y hecho. FOAF es una aplicación de XML y RDF. Básicamente se trata de un archivo RDF escrito en XML con el cual se describen personas, documentos, imágenes, etc.

- **Bibliographic Ontology (BIBO)**

BIBO¹⁵ permite describir sujetos bibliográficos mediante el lenguaje semántico RDF. Proporciona principales conceptos y propiedades para describir las citas y referencias bibliográficas (es decir, citas, libros, artículos, etc.) en la Web Semántica.

- **DBpedia Property (PROPERTY)**

DBpedia¹⁶ es una ontología de dominios cruzados poco profundas que se ha creado manualmente basado en los InfoBoxes más comúnmente utilizados dentro de Wikipedia¹⁷. La ontología cubre actualmente 529 clases que forman una jerarquía de subsunción y se describen por 2.333 propiedades diferentes.

¹² <http://dublincore.org/documents/usageguide/glossary.shtml>

¹³ DCMI: Dublin Core Metadata Initiative [<http://dublincore.org/>]

¹⁴ <http://xmlns.com/foaf/spec/#sec-glance>

¹⁵ <http://bibliontology.com/specification>

¹⁶ <http://dbpedia.org/Ontology>

¹⁷ Wikipedia: es una enciclopedia libre, políglota y editada colaborativamente. [<http://es.wikipedia.org/wiki/Wikipedia>]

- ***Academic Institution Internal Structure Ontology (AIISO)***

Academic Institution Internal Structure Ontology, AIISO¹⁸ ofrece clases y propiedades para describir la estructura interna de una organización académica.

- ***Teaching Core Vocabulary (TEACH)***

TEACH¹⁹ es un vocabulario liviano que ofrece términos para que los profesores relacionen cosas correspondientes a sus cursos; TEACH está basado en las necesidades prácticas establecidas, proporcionando seminarios y descripciones de cursos como datos enlazados.

1.1.4.2. Iniciativas de vocabularios desarrollados por la UTPL.

La sección departamental de tecnologías Avanzadas de la Web y Sistema Basados en el Conocimiento ha venido proponiendo y desarrollando proyectos relacionados con recursos educativos abiertos en especial de tipo OCW basados en la tecnología Linked Data; marco del trabajo del cual han creado un vocabulario denominado LOCWD y bajo la misma líneas actualmente se encuentra en proceso de desarrollo el vocabulario IPD; trabajos detallados a continuación.

- ***Linked OpenCourseWare Data (LOCWD)***

LOCWD²⁰ es una ontología que describe y conecta repositorios OCW.

El vocabulario LOCWD ofrece términos de los sitios OCW como: títulos, áreas de conocimientos, instituciones educativas de alto nivel, materiales educativos abiertos y otros. Como lo citan (Piedra, Tovar, Chicaiza, & López, 2013) el objetivo de este vocabulario es describir recursos OCW de organizaciones educativas.

- ***Planes Docentes En Educación Superior (IPD)***

IPD²¹ es un vocabulario en proceso de desarrollo, el cual describe información acerca de planes docentes. El proyecto denominado IPD extiende el programa general de extracción Linked OCW Data de la UTPL.

¹⁸ <http://vocab.org/aiiso/schema>

¹⁹ <http://linkedscience.org/teach/ns/>

²⁰ <http://es.slideshare.net/OCWConsortium/ocwc-global-2013-locwd-a-vocabulary-for-ocw-based-on-linked-open-data-technologies>

1.1.5. *Linked Data*.

Datos enlazados o Linked Data²² es un término usado para describir una práctica recomendada para exponer, compartir y conectar piezas de datos, información y conocimientos sobre la web semántica utilizando URIs y RDF.

(Berners-Lee & Handler, The Semantic Web, 2001), define que para lograr tener los datos interconectados, como si la Web fuese una gran base de datos, se deben respetar cuatro reglas básicas.

1. Usar URIs para identificar las cosas
2. Usar URIs HTTP
3. Ofrecer información sobre los recursos usando RDF
4. Incluir enlaces a otros URIs

Esta interconexión permite reutilizar la información, lo que ofrece un valor añadido a la Web.

La Figura 1.5 muestra los datasets²³ que han sido publicados en formato Linked Data por colaboradores de la comunidad del Linked Open Data y por individuos de otras organizaciones.



As of September 2011

Figura 1.5 Nube de datos Enlazados (2011)
Fuente: (<http://lod-cloud.net/>)

²¹ IPD: Trabajo de titulación de fin de carrera desarrollado por estudiantes de la escuela de Informática de la Modalidad Abierta y a Distancia de la UTPL

²² Linked Data: Datos enlazados [<http://linkeddata.org/home>]

²³ Datasets: Conjunto de datos [http://es.wikipedia.org/wiki/Conjunto_de_datos]

En el campo educativo se han realizado varios proyectos basados en la tecnología Linked Data, a continuación se detalla brevemente las algunas propuestas.

- ***Serendipity***

Serendipity²⁴ es un proyecto auspiciado por el grupo GICAC²⁵ de la Universidad Politécnica de Madrid y la Universidad Técnica Particular de Loja en colaboración con instituciones OCW, el propósito principal del servicio desarrollado es facilitar a los estudiantes, maestros y autodidactas un motor de búsqueda de facetas, que permite a los usuarios visualizar repositorios OCW de un conjunto de datos basado en las tecnologías de Linked Data relacionadas con el OpenCourseWare Consortium y OCW-Universia.

- ***LOIUS***

El proyecto Linking Italian University Statistics - LOIUS²⁶ tiene como objetivo representar las estadísticas universitarias publicadas por el Ministerio de Educación Italiano a través del uso de lenguas semánticas y exponer estos datos de acuerdo con las normas establecidas por Linked Data; la ontología fue creada mediante la ampliación del vocabulario SCOOVO²⁷.

- ***University Of Southampton***

La Escuela de Electrónica e Informática (ECS) de la Universidad de Southampton²⁸ ha hecho público sus datos en formato de datos enlazados abiertos para su reutilización, esto incluye datos sobre trabajos de investigación, personal de la escuela, grupos de investigación, módulos didácticos, seminarios, eventos y edificios.

Todos los datos están disponibles en formato RDF y pueden ser reusados para cualquier propósito legal, incluyendo trabajos relacionados de uso comercial.

²⁴ <http://serendipity.utpl.edu.ec/map/>

²⁵ GICAC: Grupo de Innovación en la Gestión de Calidad de Centros Universitarios

²⁶ <http://sw.unime.it/loius/info.html>

²⁷ SCOOVO: Statistical Core Vocabulary [<http://vocab.deri.ie/scovo>]

²⁸ <http://www.ecs.soton.ac.uk/news/3313>

- ***The LUCERO Project***

LUCERO²⁹ es un proyecto perteneciente a la Open University³⁰ (OU), el objetivo es establecer una infraestructura completa para la exposición de los contenidos educativos y de investigación; tanto a nivel técnico como organizativo, es decir crear un Web OU³¹ de datos, con la conexión de todos los repositorios entre sí y con conjuntos de datos externos.

- ***LODUM***

LODUM³² es el proyecto Linked Open Data de la Universidad de Münster³³, construyó una base de datos científica denominada CRIS³⁴, que contiene información interconectada sobre las 5 P (personas, publicaciones, proyectos, patentes y premios) asociados a la Universidad. LODUM lleva su aproximación al siguiente nivel al implementar una estrategia que persigue mejorar la transparencia y la visibilidad de la universidad.

- ***OAIster***

OAIster³⁵ es un proyecto de la Universidad de Michigan³⁶ orientado al e-learning³⁷ y al ámbito educativo; es un catálogo de millones de registros que representan recursos digitales de acceso abierto. Entre los materiales disponibles se puede encontrar libros digitalizados, archivos de audio, archivos de video, imágenes fotográficas entre otras.

Hoy en día, OAIster incluye más de 25 millones de registros que representan recursos digitales de más de 1.100 colaboradores.

²⁹ <http://lucero-project.info/lb/>

³⁰ <http://www.open.ac.uk/>

³¹ http://www.slideshare.net/slideshow/embed_code/7836982#btnNext

³² <http://lodum.de/>

³³ <http://www.uni-muenster.de/>

³⁴ <http://lodum.de/about/>

³⁵ <http://www.oclc.org/oaister/about.en.html>

³⁶ <http://www.ns.umich.edu/new/releases/4848-Web%20search%20service%20reveals%20hidden%20scholarly%20resources>

³⁷ e-learning: Aprendizaje electrónico a la Educación a Distancia [<http://es.wikipedia.org/wiki/E-learning>]

1.1.6. Herramientas para extracción y segmentación de datos desde páginas HTML y publicación de datos.

Como se indicó al inicio de este proyecto, el presente trabajo forma parte de un propósito integral del equipo de investigación de Tecnologías Avanzadas de la Web y Sistemas Basados en el Conocimiento.

Los investigadores de este departamento han establecido un conjunto de herramientas y tecnología probadas con éxito para proyectos de obtención de datos desde la Web, las cuales han sido sugeridas y reutilizadas para el fin de este proyecto.

A continuación se conceptualizan las herramientas utilizadas para los procesos de extracción (scraping³⁸), segmentación³⁹ y publicación de datos.

1.1.6.1. Python.

Python⁴⁰ es un potente lenguaje de programación interpretado, de código abierto, multiplataforma, dinámico, de alto nivel, orientado a objetos.

El intérprete de Python se extiende fácilmente con nuevas funciones y tipos de datos implementados en C⁴¹ o C++⁴². Python también es adecuado como un lenguaje de extensión para aplicaciones personalizables.

- **Scrapy**

Scrapy⁴³ es un marco de trabajo usado para el rastreo de sitios web y la extracción de datos estructurados que pueden ser utilizados para una amplia gama de aplicaciones útiles, como la minería, procesamiento de la información de datos o archivo histórico.

Scrapy fue diseñado pensando en la simplicidad, proporcionando las características de rapidez, extensibilidad y portabilidad; es de código abierto, escrito en Python y se ejecuta en Linux⁴⁴, Windows⁴⁵, Mac OS X⁴⁶ y BSD⁴⁷.

³⁸ Scraping: es un término que, traducido al español, literalmente quiere decir “rascado”, en este contexto, se refiere a la extracción, limpieza y filtro de los datos. [<http://es.schoolofdata.org/introduccion-a-la-extraccion-de-datos-de-sitios-web-scraping>]

³⁹ Segmentación: en este contexto se lo utiliza como sinónimo de división de datos.

⁴⁰ <https://docs.python.org/3/tutorial/index.html>

⁴¹ <http://code-reference.com/c>

⁴² <http://www.cplusplus.com/info/>

⁴³ <http://scrapy.org/>

La Figura 1.6 muestra de manera general la arquitectura de Scrapy con cada uno de sus componentes y como se desarrolla el flujo de datos en el sistema.

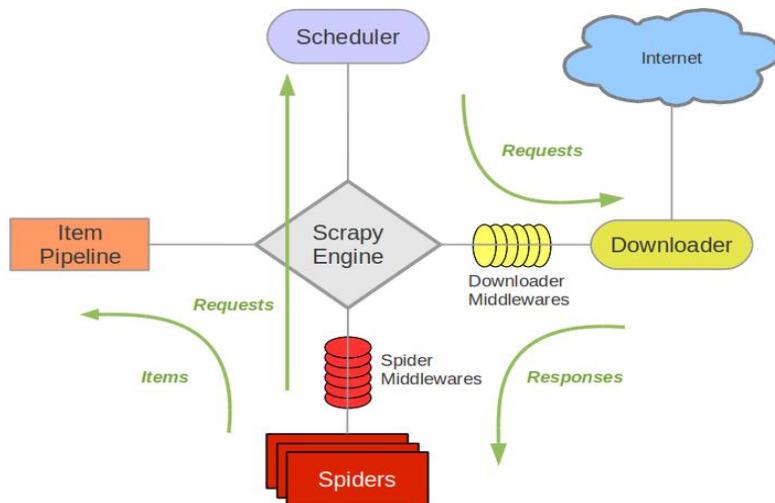


Figura 1.6 Arquitectura de Scrapy (scrapy.org, 2013)
Fuente: <http://doc.scrapy.org/en/0.16/topics/architecture.html>

La información sobre el flujo de información, descrita a continuación se la ha tomado de la página oficial de Scrapy⁴⁸.

El flujo de datos en Scrapy es controlado por el motor de ejecución:

1. El motor abre un dominio, localiza el spider que se encarga de ese dominio, y le pide a la spider la primera URL⁴⁹ que se rastreará.
2. El motor obtiene la primera URL que se rastreará de la spider
3. El motor pide al componente Scheduler para la siguiente URL a rastrear.
4. El componente Scheduler devuelve el siguiente URL para llevarlo hasta el motor y el motor los envía al componente Downloader, pasando por el Downloader Middleware (dirección petición).

⁴⁴ <http://www.ubuntu.com/download/desktop>

⁴⁵ <http://windows.microsoft.com/es-es/windows-8/meet>

⁴⁶ <https://www.apple.com/es/osx/what-is/>

⁴⁷ <https://en.wikipedia.org/wiki/FreeBSD>

⁴⁸ <http://doc.scrapy.org/en/latest/topics/architecture.html>

⁴⁹ URL: Un localizador de recursos uniforme [http://es.wikipedia.org/wiki/Localizador_uniforme_de_recursos]

5. Una vez que la página termine la descarga desde Downloader genera una respuesta (con esa página) y lo envía al motor, pasando por el Downloader Middleware (dirección de respuesta).
6. El motor recibe la respuesta desde el Downloader y lo envía al Spider⁵⁰ para el procesamiento, que pasa a través de Spider Middleware (dirección de entrada).
7. La araña procesa la respuesta y vuelve raspado artículos y nuevas peticiones (a seguir) al motor.
8. El motor envía ítems raspados (devuelto por la Araña) a Item Pipeline y solicitudes (devuelto por la araña) para el Scheduler
9. El proceso se repite desde el paso (2) hasta que no hay más peticiones desde el Scheduler, y el motor cierra el dominio.

- ***Leaf***

Leaf⁵¹ es una biblioteca para segmentar data en Python, se ejecuta en dos capas⁵² diferentes como se puede observar en la Figura 1.7 la capa Leaf y la capa User.

La capa User incluye código Python creado por el desarrollador y con las interacciones del Shell de Python.

El nivel Leaf incluye las bibliotecas por defecto para que acepte peticiones de los usuarios.

El usuario escribe su código de Python personalizado y define un protocolo de alto nivel para la descripción de las dependencias entre sus funciones, la capa Leaf ejecuta el código apropiado de acuerdo con el protocolo y almacena todos los recursos derivados para su uso futuro.

⁵⁰Spider: son las clases que definen cómo se raspó un sitio determinado
[<http://doc.scrapy.org/en/0.16/topics/spiders.html>]

⁵¹<https://pypi.python.org/pypi/leaf>

⁵²<http://pythonhosted.org/pyleaf/>

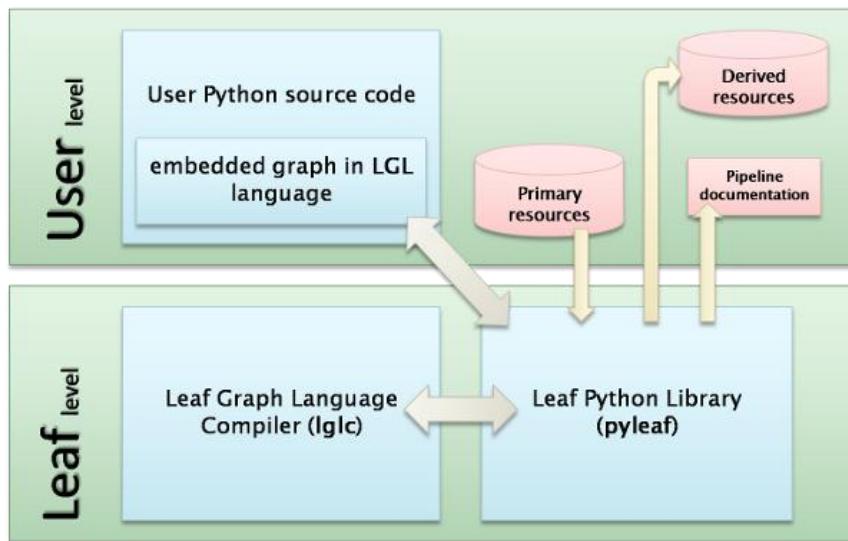


Figura 1.7 Flujo de trabajo de Leaf
Fuente: <http://pythonhosted.org/pyleaf/>

1.1.6.2. *LODSPeaKr.*

LODSPeaKr⁵³ es un marco de trabajo de código abierto para crear aplicaciones Linked Data y publicar dato RDF de una manera rápida y amigable, de fácil instalación, provee la posibilidad de crear nuestro propio flujo de ejecución de consultas SPARQL⁵⁴ en múltiples EndPoints⁵⁵ SPARQL.

Asimismo, LODSPeaKr proporciona la negociación de contenido, lo que permite obtener los mismos datos en múltiples formatos (RDF/XML, N-Triples⁵⁶, RDFJSON⁵⁷, etc.).

El marco de trabajo LODSPeaKr se ha utilizado para diferentes proyectos personalizados⁵⁸ como por ejemplo un demo utilizando los datos del portal abierto del gobierno Chileno, en el cual se describen los medicamentos bioequivalentes para diferentes tratamientos, el cual está disponible en español.

⁵³ <http://lodsppeakr.org/>

⁵⁴ <http://www.w3.org/TR/rdf-sparql-query/>

⁵⁵EndPoint: Nombre asignado a la dirección web del TripleStore donde se acceden los datos.

⁵⁶ <http://www.w3.org/TR/n-triples/>

⁵⁷ <https://dvcs.w3.org/hg/rdf/raw-file/default/rdf-json/index.html>

⁵⁸ <http://lodsppeakr.org/applications.html>

1.2. OCW

Actualmente los recursos para la enseñanza, el aprendizaje y la investigación que ofrecen entidades educativas se encuentran dispersos en diferentes repositorios en formato de cursos OCW.

Como lo cita (Oriol Borrás, 2010), "OpenCourseWare (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 Massachusetts⁵⁹ (MIT) en colaboración con la Fundación William and Flora Hewlett y la Fundación Andrew W. Mellon".

UNESCO⁶⁰ en su web oficial señala que en una conferencia de prensa el MIT anunció que publicaría en la web los materiales educativos, lecciones, ejercicios, exámenes de 2.000 cursos para que puedan ser utilizados por cualquiera de forma gratuita; la creación del OCW del MIT fue una inspiración para otras entidades universitarias que también emprendieron en la publicación de sus Recursos Educativos Abiertos⁶¹.

En el año 2005 se creó el OpenCourseWare Consortium⁶², este consorcio está formado por organizaciones universitarias que tienen el objetivo de trabajar juntas para promover la autonomía educativa de las personas en todo el mundo.

A continuación presentamos datos estadísticas que muestran el continuo crecimiento que ha tenido el movimiento OCW, los datos están actualizados a fecha de febrero del 2014⁶³ y han sido tomados de los reportes mensuales del MIT OpenCourseWare.

Se puede contrastar como punto positivo que el número total de visitas globales al OCW Consortium de este año se ha incrementado en un 14% en relación al año anterior (2013), como punto negativo el número de visita a la página del mit.edu ha disminuido en un 18%, por citar algunos datos de la tabla 1.1.

⁵⁹ <http://ocw.mit.edu/index.htm>

⁶⁰ <http://www.unesco.com.org/news/es/education/>

⁶¹ REA: en inglés, Open Educational Resources (OER), fue acuñado en 2002 por la Unesco. "Son recursos para la enseñanza, el aprendizaje y la investigación, que residen en el dominio público bajo una licencia de propiedad intelectual que permite su libre uso para otras personas."

⁶² <http://www.ocwconsortium.org/>

⁶³ <http://ocw.mit.edu/about/site-statistics/monthly-reports/>

Tabla 1.1 Número de visitas realizadas al sitio OCW

Site Traffic Measure	Total	2014	
		February	
Total Visits including affiliates¹	196,760,701	2,647,595	
Last year/% change		2,369,912	12%
Total Visits	150,145,335	2,312,184	
Last year/% change		2,034,501	14%
Total Unique Visitors	92,913,621	1,468,713	
Last year/% change		1,230,567	19%
MIT.EDU Visits	2,158,510	24,854	
Last year/% change		30,301	-18%
Highlights for High School Visits	3,639,914	48,924	
Last year/% change		53,293	-8%
OCW Scholar Visits	9,212,970	329,311	
Last year/% change		375,457	-12%
Page Views	929,874,904	10,791,079	
Last year/% change		10,336,033	4%
.ZIP Downloads	19,585,973	222,657	
Donation Page Visits	684,480	6,289	
iTunesU Downloads ²	45,897,418	284,616	
YouTube Views	57,039,499	1,213,624	
OCW enewsletter subscribers	218,746	3,561	

¹ Some data estimated. ² Data provided in weekly increment

Source: Akamai, SiteWise, and MailChimp. Akamai data for November zip downloads is currently not available. This number reflects October data.

Fuente: http://ocw.mit.edu/about/site-statistics/monthly-reports/MITOCW_DB_2014_02.pdf

Elaborado por: OCW Consortium

La tabla 1.2 muestra el valor de cursos publicados, cursos actualizados, cursos traducidos, entre otros datos importantes.

Tabla1.2 Publicación de datos de cursos OCW

Publication Metric	2014
	February
Total Tenure Track Faculty Participating	601 (66%*)
Courses Published	2,204
Updated Courses Published	761
Archived to D-Space	857
Full Video Courses	71
Mirror Sites	342
Translated Courses	1,018

* Based on 913 tenure track teaching faculty as of 2013.

Fuente: http://ocw.mit.edu/about/site-statistics/monthly-reports/MITOCW_DB_2014_02.pdf

Elaborado por: OCW Consortium

En la Figura 1.8 se puede apreciar el número de afiliaciones mensuales con respecto a los OCW. El consorcio Universia ha disminuido en el último año su número de afiliaciones, mientras que el Consorcio OCW sigue liderando la popularidad internacional entre educadores y autodidactas.

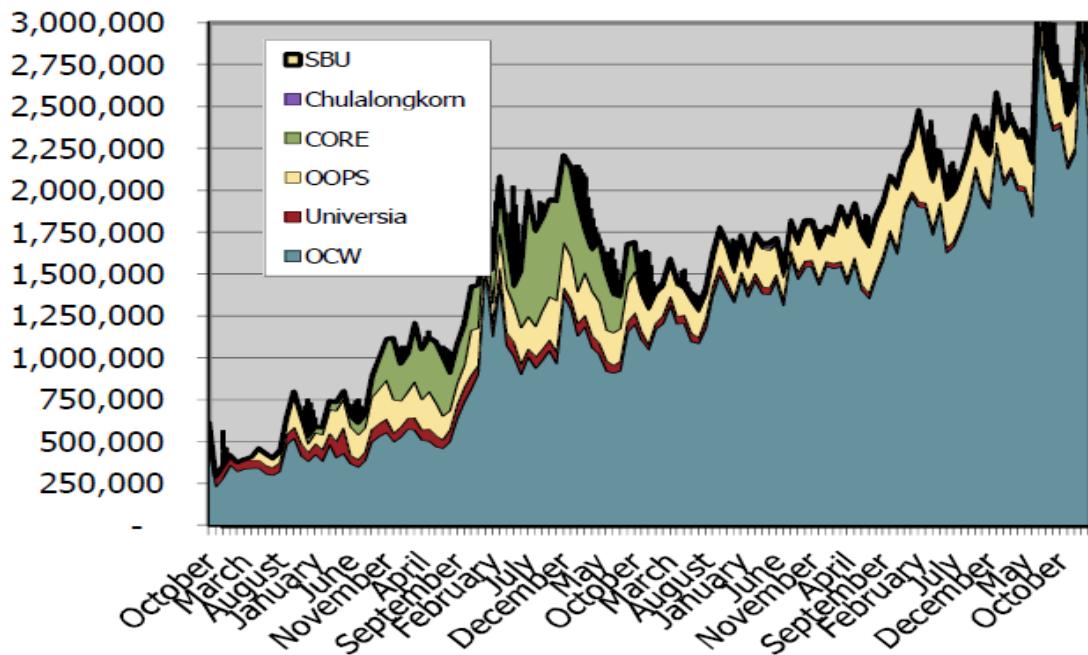


Figura 1.8 OCW y visitas mensuales de afiliados

Fuente: http://ocw.mit.edu/about/site-statistics/monthly-reports/MITOCW_DB_2014_02.pdf

Elaborado por: OCW Consortium

La información de un OCW se organiza dentro de cursos, la estructura de estos cursos se modelan de acuerdo a la información académica que la Universidad desea transmitir a los estudiantes a través del curso, aquí es donde se hace imperativo realizar una breve descripción de la relación entre la palabra Syllabus y los cursos OCW.

1.2.1. *Syllabus*

Término empleado desde hace algún tiempo por los expertos norteamericanos en educación, el vocablo Syllabus es objeto de análisis, particularmente, en la expresión “planes académicos”, entendiendo por tal un concepto equivalente al arte de diseñar y elaborar itinerarios de enseñanza-aprendizaje.

(Núñez Paris, 2008), en *Teoría del currículum y didáctica de las lenguas extranjeras*, toma de M. P. Breen el significado de “silabo⁶⁴ como un plan acerca de lo que se pretende alcanzar a través del proceso de enseñanza-aprendizaje. Éste forma parte del currículum global de lenguas o de un curso que está compuesto por cuatro elementos: objetivos, contenidos, metodología y evaluación”.

Como lo cita (Ferreiro R. , 2011) en su blog oficial, tanto uno como el otro son en esencia lo mismo, aunque hay que reconocer que el concepto de “Syllabus” en la práctica es más abarcador ya que contempla un mayor número de aspectos del curriculum escolar que el concepto de “Programa de la Asignatura” o Planes Docentes.

El syllabus debe contener toda la información que el profesor deba transmitir a sus estudiantes con respecto al curso y que es lo que los estudiantes deberán aprender de dicho curso.

Como lo describe (VIRTUAL UNAB)⁶⁵ en un artículo educativo, un Syllabus debe contener como mínimo la siguiente información:

- Nombre del curso
- Descripción del curso
- Prerrequisitos del curso
- Objetivos de aprendizaje
- Textos requeridos
- Lista de todas las actividades o tareas
- Explicación de los criterios de evaluación
- Mecanismos para la entrega de tareas
- Política de trabajos atrasados
- Programación semanal

Analizando las fuentes de información de cursos OCW previo al proceso de extracción de la data, se pudo evidenciar varias particularidades las cuales citamos a continuación:

- No existe una normativa estandarizada para publicar cursos OCW.
- El Syllabus tiene varias denominaciones tales como Planes Académicos, Planes Docentes, Programa de la Asignatura, etc.

⁶⁴ Silabo: Es el español de Syllabus

⁶⁵ UNAB: Universidad Andrés Bello [http://www.unab.cl/sitio-html/rr_ii/welcome.asp]

En la Figura 1.9 se puede visualizar las diferentes denominaciones y esquematización del syllabus de algunos cursos OCW.

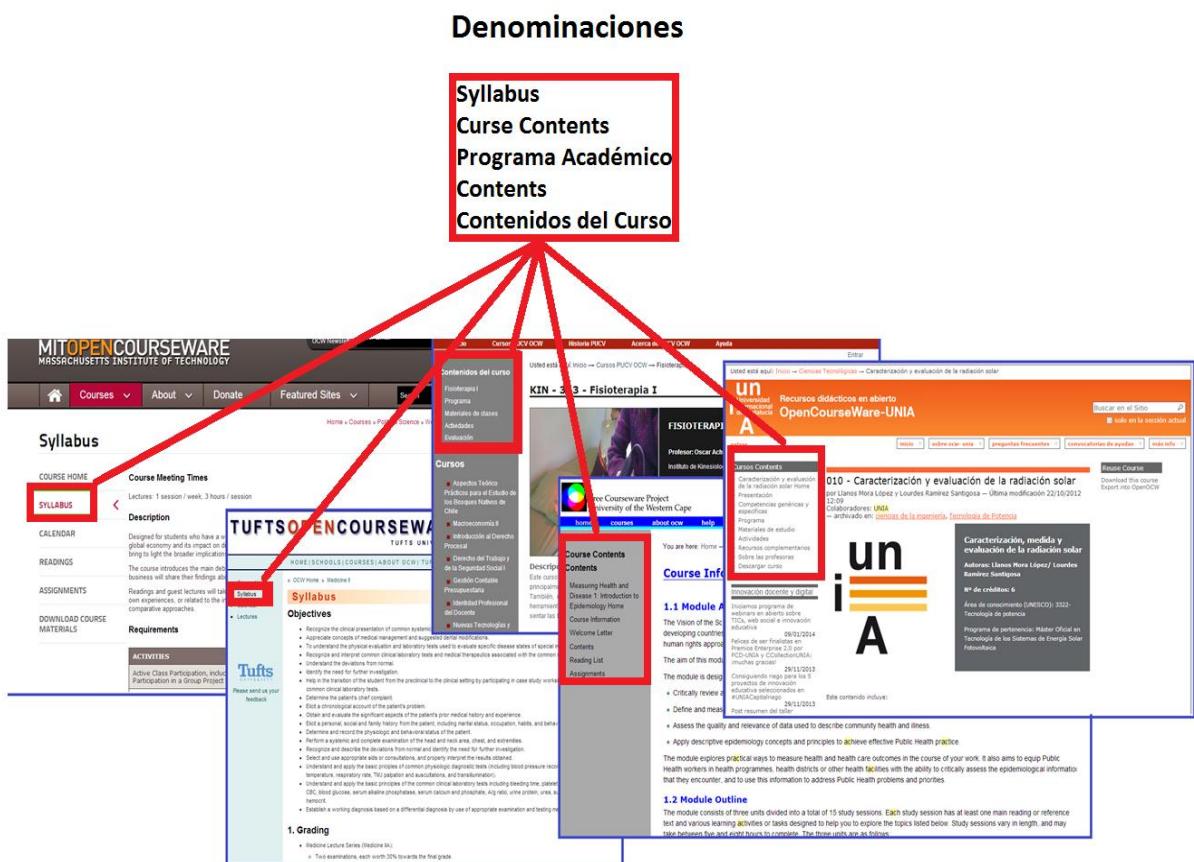


Figura 1.9 Diferentes denominaciones de Syllabus de cursos OCW.
Elaborado: Diana Molina, Paula Soto F.

En ocasiones los datos no están dispuestos para el usuario como nos gustaría, o en el caso de los cursos OCW la mayoría de los datos no están siquiera con una estructura uniforme, los OCW ofrecen su información académica en diferentes formatos, PDFs, multimedia, documentos Word, presentaciones Power Point, y otros; particularizamos que el proceso de extracción de información pedagógica se enfocará solo en datos en formato HTML excluyendo cualquier formato descargable.

Al finalizar el capítulo I, luego de conceptualizar la representación del contenido académico de un curso OCW, la técnica que mejor se adapta para conectar e interoperar los datos del Syllabus es la tecnología Web Semántica debido a que permite:

- Describir y organizar los recursos académicos.

- Presentar la semántica de los metadatos en término de ontologías.
- Crear un archivo RDF.
- Publicar estos datos para que puedan ser utilizados por usuarios.

1.3. Propuesta de trabajo OCW

El aporte de este proyecto para ayudar en la problemática de la búsqueda de recursos educativos en la Web, es convertir la data HTML de los cursos OCW a tecnología RDF, tomando como punto de partida la información recolectada por el proyecto Consuming and producing linked open data: The case of OpenCourseWare.

En la Figura 1.10 se observa de forma global los diferentes procesos creados a lo largo de este proyecto, los dos primeros procesos describen la extracción y segmentación de datos utilizando las herramientas Scrapy y Leaf de Python respectivamente; el siguiente proceso hace referencia a la categorización de datos, una vez clasificado los datos se procede a crear un vocabulario denominado DOCW (Descripción de cursos OCW), donde se reutiliza vocabularios consensuados como BIBO, AIISO, DC, etc. y vocabularios autoría de la UTPL como LOCWD e IPD; a continuación se genera un archivo RDF el cual se encuentra de manera permanente en el triple Store Virtuoso. Como proceso final se tiene la publicación de los datos utilizando LODSPeaKr que permite la explotación y visualización de datos enlazados.

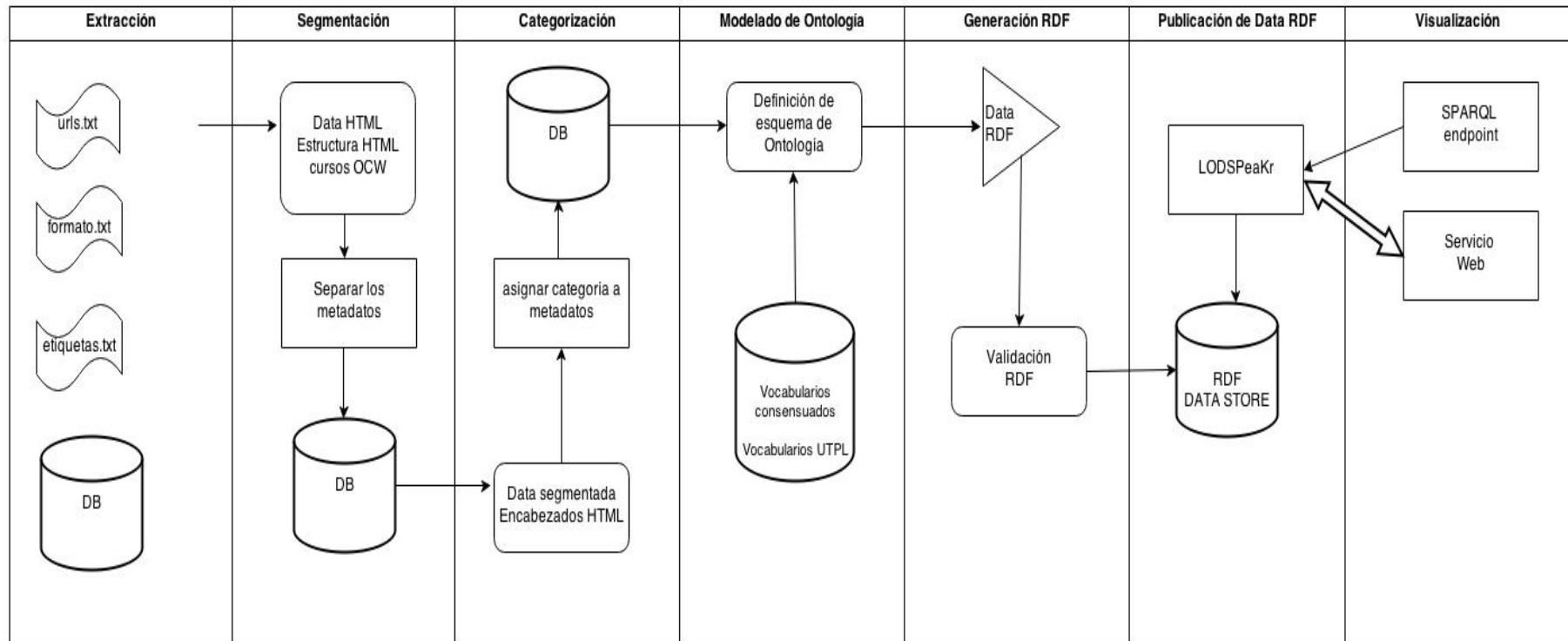


Figura 1.10 Diagrama de proceso de la propuesta de trabajo OCW.
Elaborado: Diana Molina, Paula Soto F.

CAPÍTULO II: DESCRIPCIÓN DE FUENTE DE DATOS

2. Recolección de datos

Para el presente capítulo se hace referencia al global de los datos obtenidos de la investigación denominada “Consuming and producing linked open data: The case of OpenCourseWare”⁶⁶, trabajada por la UTPL; proyecto que presenta la iniciativa de aplicar los principios de Linked Data para mejorar el descubrimiento y búsqueda de contenidos OCW, datos que serán descritos en el apartado 2.2., mientras que la sección 2.3. se puntualiza los criterios de selección de datos con los que se trabaja a lo largo de este proyecto.

2.1. Universo de datos

El ambiente de los OCWs está caracterizado por grandes cantidades de datos, en la tabla 2.1 se aprecia la colección de datos de los repositorios OCW tomados del proyecto “Consuming and producing linked open data: The case of OpenCourseWare.

Tabla 2.1 Colección de datos OCW del proyecto “Consuming and producing linked open data: The case of OpenCourseWare

Publicación de datos	
Repositorios Heterogéneos OCW (universidades)	80
Total Cursos OCW	7239
OERs	17053 aprox.
Países	28
Palabras claves	27588
Áreas de Conocimiento	657
Consorcios	2

Fuente: <http://www.emeraldinsight.com/journals.htm?articleid=17104601>

Elaborado por: Piedra Nelson, Chicaiza Janneth, Tovar Edmundo, López Jorge, Colomo Ricardo.

⁶⁶ <http://www.emeraldinsight.com/journals.htm?articleid=17104601>

2.2. Selección de datos

Al analizar el global de las 80 universidades, se establece como criterio de selección el idioma, ya que la mayoría de los cursos OCW se encuentran publicados en inglés y español, a continuación se detalla varios datos numéricos que permiten determinar el conjunto de universidades con las que se trabaja.

- 17 universidades están en otros idiomas (Polaco, Portugués, Coreano, Japonés, Russo, Chino, Francés, Hebreo, Chamorro, Catalán e Indonesio).
- 63 universidades están en idioma inglés y español, donde 21 de ellas poseen algún tipo de problema (ver tabla 2.2) lo cual nos da un total de 42 universidades.

Tabla 2.2 Problemas con varios cursos OCW

Inconvenientes encontrados en varias Universidades
Servidor caído
URL no existe
Pide ingresar con password de la universidad.
Menú en inglés pero la información en otros idiomas.
Cursos OCWs se encuentran en formato PDFs, presentaciones Power Point, videos, audio y archivos descargables
URL estático no se modifica.
Algunos cursos solo se rotulan sus nombres pero carecen de información.

Elaborado por: Diana Molina, Paula Soto F.

En la tabla 2.3 se detallan las universidades que son objeto de la investigación y desarrollo de este proyecto.

Tabla 2.3 Listado de Universidades con su respectivo idioma y número de cursos OCW

Universidad	Idioma	Nº Cursos
African Virtual University	Inglés	35
Capilano University	Inglés	23
Fundación Universitaria San Pablo CEU	Español	4
Gunadarma University	Inglés	17
Indian Institute of Management Bangalore	Inglés	10
Johns Hopkins Bloomberg School of Public Health	Inglés	86
Kyoto University	Inglés	117

Massachusetts Institute of Technology	Inglés	2092
New Jersey Institute of Technology	Inglés	20
Pontificia Universidad Católica de Valparaíso	Español	21
Tecnológico de Monterrey	Español	11
Tokyo Institute of Technology	Inglés	36
Tufts University	Inglés	50
United Nations University	Inglés	16
Universidad Carlos III de Madrid	Español	165
Universidad de Alicante	Español	316
Universidad de Cantabria	Español	139
Universidad de Huelva	Español	30
Universidad de Monterrey	Español	24
Universidad de Murcia	Español	162
Universidad de Navarra	Español	20
Universidad de Salamanca	Español	89
Universidad de Sevilla	Español	236
Universidad de Zaragoza	Español	31
Universidad del Valle	Español	11
Universidad Estatal a Distancia	Español	12
Universidad Icesi	Español	9
Universidad Internacional de Andalucía	Español	43
Universidad Nacional de Córdoba	Español	52
Universidad Nacional de Educación a Distancia	Español	17
Universidad Nacional de Ingeniería	Español	21
Universidad Politécnica de Cartagena	Español	43
Universidad Politécnica de Madrid	Español	118
Universidad Técnica Particular de Loja	Español	11
Universitat de les Illes Balears	Español	13
Universitat de València	Español	62
University of Cape Town	Inglés	183
University of Massachusetts Boston	Inglés	41
University of Michigan	Inglés	73
University of Notre Dame	Inglés	47
University of the Western Cape	Inglés	6
Wheelock College	Inglés	1
Total Cursos		4513

Elaborado por: Diana Molina, Paula Soto F.

Luego del análisis de datos expuesto en este capítulo, se puntuiza que se trabajará con 4513 cursos OCW repartidos entre 42 universidades.

CAPÍTULO III: PREPROCESAMIENTO DE DATOS

Una vez determinados los orígenes de la información y definida la selección de las universidades, el siguiente punto es implementar los procesos de extracción, segmentación y categorización a los contenidos de cursos OCW, procesos detallados en los siguientes apartados.

3.1. Extracción de datos de cursos OCW

Conocer la estructura HTML de un curso OCW es el siguiente paso para extraer los datos, para ello se debe identificar los patrones⁶⁷ basados en tags⁶⁸ donde se encuentra la información.

Se seleccionan aquellos tags que engloben todo el contenido académico de un curso OCW; encontrado diez patrones (ver tabla 3.1) a nivel de las 42 universidades.

El trabajo de búsqueda de patrones se lo realizó manualmente debido a la particularidad de las diferentes estructuras HTML en las que están construidas los cursos, de esa manera se puede tener un control más preciso al momento de extraer la data.

Tabla 3.1 Patrones identificados para extraer información de contenidos académicos de cursos OCW

Patrones
<div id="parent-fieldname-text">
<div id="course_inner_section">
<div id="about">
<div class="plain">
<div id="region-content">
<div id="description">
<dl class="outlines">
<table class="documentTable">
<table class="documentTable plain">
<table class="grid listing">

Elaborado por: Diana Molina, Paula Soto F.

⁶⁷ Patrón: tipo de tema de sucesos u objetos recurrentes. [<http://es.wikipedia.org/wiki/Patr%C3%B3n>]

⁶⁸ <http://www.htmlquick.com/es/reference/tags/div.html>

En la Figura 3.1 se visualiza el tag <div id="course_inner_section"> donde se encuentra la información de la categoría Syllabus; para poder observar la estructura HTML se utilizó el complemento Firebug⁶⁹.

The screenshot shows a browser window for MIT OpenCourseWare. At the top, there's a navigation bar with links for OCW Newsletter, Help, Contact Us, Home, Courses, About, Donate, Featured Sites, Search, and Advanced Search. The main content area has a header 'Syllabus'. On the left, there's a sidebar with links for Course Home, Syllabus (which is highlighted in red), Calendar, and Readings. The main content area contains sections for 'Course Meeting Times' (with text about 2 sessions/week, 1.5 hours/session) and 'A Cognitive Science Approach To Engineering' (with a detailed description of the course's focus on objects, prototyping, and evaluation). A red box highlights the 'Course Meeting Times' section. Below the main content, the Firebug developer tool is visible, showing the DOM structure. The 'HTML' tab is selected, and the code for the 'course_inner_section' div is shown:

```

<div id="course_inner_section">
    <h2 class="subhead">Course Meeting Times</h2>
    <p>Lectures: 2 sessions / week, 1.5 hours / session</p>
    <h2 class="subhead">A Cognitive Science Approach To Engineering</h2>
    <p>The look, feel and use of objects communicate their value to us. This course applies cognitive science and technology to the industrial design process. The course will introduce prototyping techniques and approaches for objective evaluation as part of the design process. Students will practice evaluating products with mechanical and electronic aspects. The evaluation process will then be applied to creating functioning product prototypes. This is a project-oriented course that will draw on engineering, aesthetic, and creative skills. The course is geared towards students interested in creating physical products which encompass electronics and computers in order to include them in scenarios. Students will present</p>

```

The right side of the Firebug interface shows the CSS styles for the 'course_inner_section' class.

Figura 3.1 Tag de contenido del Syllabus de un OCW del MIT⁷⁰

Fuente: <http://ocw.mit.edu/courses/media-arts-and-sciences/mas-742-industrial-design-intelligence-a-cognitive-approach-to-engineering-fall-2003/syllabus/>

Elaborado: Diana Molina, Paula Soto F.

Los datos que se obtendrán serán del contenido del menú principal de las páginas de los curso OCW como Página principal, Syllabus, Calendario, Textos de Lectura, Tareas, etc., (ver Figura 3.2); se hace hincapié en que se extraerá la mayor cantidad de datos posibles, ya que la falta de homogeneidad de la estructura HTML entre cursos OCW dificulta el proceso.

⁶⁹ <https://addons.mozilla.org/es/firefox/addon/firebug/>

⁷⁰ MIT: Massachusetts Institute of Technology [<http://ocw.mit.edu/index.htm>]

Información académica de un curso OCW

The diagram illustrates the structure of an OCW course page. A large red arrow points from the left sidebar to the main content area of each of the five screenshots. The screenshots show:

- Syllabus:** Course Meeting Time, Syllabus, Calendar, Readings, Lecture Notes, Projects.
- Calendar:** COURSE HOME, Instructor(s), Syllabus, Calendar, Readings, Lecture Notes, Projects.
- Readings:** COURSE HOME, Instructor(s), Syllabus, Calendar, Readings, Lecture Notes, Projects.
- Lecture Notes:** COURSE HOME, Instructor(s), Syllabus, Calendar, Readings, Lecture Notes, Projects.

Figura 3.2 Información académica de un curso OCW pertenecientes al MIT
Elaborado: Diana Molina, Paula Soto F.

Una vez analizada la estructura HTML donde se encuentra la información académica de los cursos OCW se procede a automatizar el proceso de extracción de datos; con este fin se creó un proyecto en el lenguaje de programación Python que se ha denominado genérico (Ver Figura 3.3), el cual utiliza el framework de desarrollo Scrapy; con este marco de trabajo se crea un spider llamado generico_spider.py; básicamente este programa lee las

URLs, rastrea la información de acuerdo al tag HTML, extrae la data y la almacena en la base de datos.

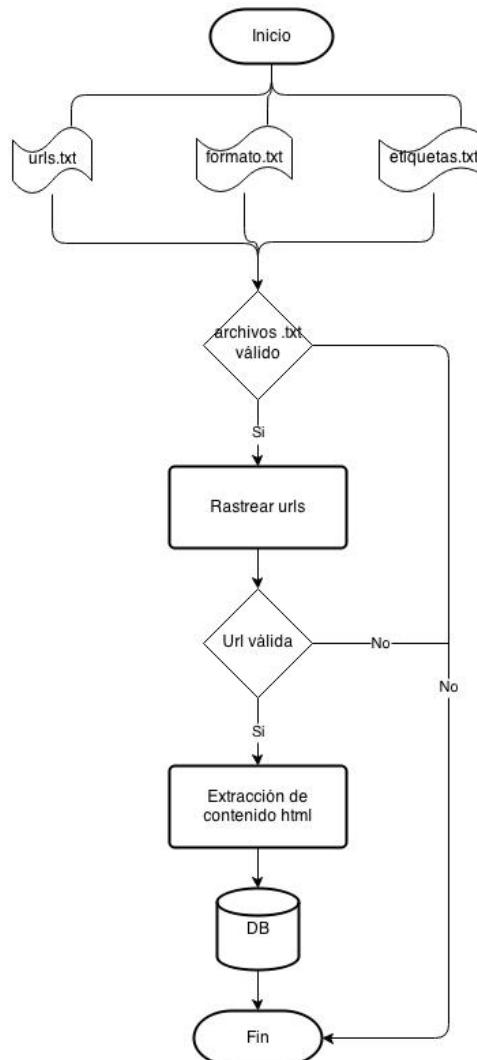


Figura 3.3 Diagrama de flujo del proceso de extracción de datos de curso OCW
Elaborado: Diana Molina, Paula Soto F.

Para el funcionamiento de genérico_spider.py se necesita los siguientes elementos que hacen posible la extracción:

- ***run.py***

Módulo que inicia la ejecución del Spider (generico_spider.py).

- **urls.txt**

En este archivo se lista las URLs de los cursos OCW de las cuales se va a extraer los datos.

- **formatos.txt**

Archivo en el cual se especifica el tag HTML mencionado anteriormente en la tabla 3.1.

- **tag_genericos.txt**

En este archivo se ubica la etiqueta del tag (`<div>`) que se utiliza en el archivo formatos.txt.

- **generico_spider.py**

Módulo encargado de combinar los archivos urls.txt, formatos.txt, tags_genericos.txt y bd.py mediante llamadas de métodos y funciones para extraer la data de los contenidos académicos.

- **bd.py**

Módulo sirve para conectar, guardar, insertar y cerrar la base de datos.

En la Figura 3.4 se visualiza el resultado de la ejecución del programa genérico_spider.py aplicado a un curso OCW del MIT. La información extraída esta almacenada en una BD donde se observa que los datos conservan su estructura HTML.

```
<div class="" id="parent-fieldname-text">
<h2 class="subhead">Course Meeting Times</h2>
<p>Lectures: 2 sessions / week, 1.5 hours / session</p>
<h2 class="subhead">A Cognitive Science Approach To Engineering</h2>
<p>The look, feel and use of objects communicate their value to us. This course applies cognitive science and te
the design process. Students will practice evaluating products with mechanical and electronic aspects. The evalua
aesthetic, and creative skills. The course is geared towards students interested in creating physical products which
product prototype, and complete a publication style paper. We will mill, cut, mold solder, program, and draw our
<p>Model making and prototyping skills will be taught.</p>
<h2 class="subhead">Requirements</h2>
<div class="maintabletemplate">
<table class="tablewidth50" summary="Course Table Listing"><thead><tr class="tableheader"><th id="col1" sc
<th id="col2" scope="col">PERCENTAGES</th>
</tr></thead><tr class="row"><td headers="col1"><em>Discussion</em><br><br>
Artifacts, Images, Papers, Books, Ideas</td>
<td headers="col2">10%</td>
</tr><tr class="alt-row"><td headers="col1"><em>Design We Do in Class</em><br><br>
2002: Dirty Coke Can, Doorstop Bikes<br><br>
2003: Brainstorm Every Two Weeks - Silos of Knowledge and Activity</td>
<td headers="col2">10%</td>
</tr><tr class="row"><td headers="col1"><em>Projects</em><br><br>
2 Liter Soda Bottle Door Stop<br><br>
Vacuum Form Container<br><br>
Water Jet Door Stop with MIT or ML Embedded<br><br>
3D Printed Thing You Can Hold in Your Hand<br><br>
Electronics and Sensors<br><br>
Final Project</td>
<td headers="col2">60%</td>
</tr><tr class="alt-row"><td headers="col1"><em>Presentations</em><br><br><br>
2 Formal Presentations of Book or Paper<br><br>
1 In-class Project Taken Farther<br><br>
```

Figura 3.4 Datos extraídos de un curso OCW almacenados en una BD.
Elaborado: Diana Molina, Paula Soto F.

Al analizar este proceso de extracción de datos se puede argumentar que debido a la falta de homogeneidad⁷¹ de las estructuras HTML entre cursos OCW se tuvo que realizar de forma manual la determinación de patrones de contenido de información, para tener un control de la información extraída.

3.2. Segmentación de la data

En este apartado se describe como se ha procedido a segmentar la data extraída de los cursos OCW alojada en la base de datos.

Para poder aplicar el proceso de segmentación el primer paso es analizar en que tags se encuentran ubicados los diferentes elementos de los contenidos OCW (Syllabus, calendario, objetivos de aprendizaje, contenidos, bibliografía, lecturas, etc.).

Al analizar un conjunto considerable de información extraída se pudo evidenciar que no existe una estructura HTML común de diseño de contenido académico, esto significa que mientras los objetivos de un curso se encuentran en un tag <H2>, en otro curso los objetivos están un tag o <p>; demostrando que no existe un protocolo⁷² de diseño de Syllabus de cursos OCW.

Una vez diferenciados estos tags se procede a crear un programa denominado separa_data.py; el cual utiliza el framework Leaf de Python; básicamente este marco de trabajo sirve para segmentar estructuras HTML.

Separar_data.py recorre cada registro de la base de datos, y por cada tag padre asignado (por ejemplo <h2> “Objetivos”) selecciona el texto, luego lee los hijos que le pertenecen a ese tag los extrae y los separa. La información segmentada es almacenada en tripletas (sujeto, predicado, objeto) en una BD⁷³.

En la Figura 3.5 se visualiza el diagrama de flujo del proceso de segmentación de datos.

⁷¹ Homogeneidad: Que está formado por elementos con una serie de características comunes referidas a su clase o naturaleza que permiten establecer entre ellos una relación de semejanza [http://es.thefreedictionary.com/homog%C3%A9neas]

⁷² Protocolo: Uno o un conjunto de procedimientos destinados a estandarizar un comportamiento humano u sistemático artificial frente a una situación específica [http://es.wikipedia.org/wiki/Protocolo]

⁷³ BD: Base de Datos [http://es.wikipedia.org/wiki/Base_de_datos]

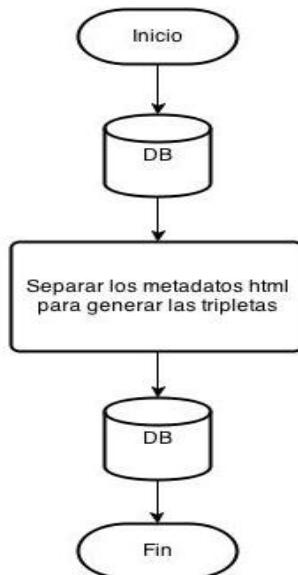


Figura 3.5 4 Diagrama de flujo para segmentar la data
Elaborado: Diana Molina, Paula Soto F.

Para el funcionamiento de `separa_data.py` se necesita los siguientes elementos que hacen posible el proceso de segmentación de los datos:

- ***separa_data.py***

Módulo que recorre cada uno de los registros de la BD, este contiene métodos que sirve para segmentar la información que tenga una tag asignado (h1, h2, h3,...,h6). Esta información se almacena en la BD en forma de tripletas:

sujeto: ur
 predicado: texto h1
 objeto: contenido de ese tag h1

- ***bd.py***

Módulo sirve para conectar, guardar, insertar y cerrar la base de datos.

En la tabla 3.2 se visualiza el resultado de la ejecución del programa `separa_data.py` aplicado a la información de cursos OCW almacenada en la BD.

Tabla 3.2 Ejemplo de datos de cursos OCW segmentados y almacenados en la BD.

http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/PrinciplesIndustrialHygiene/coursePage/syllabus/	Course Descriptions	<p>Principles of Industrial Hygiene provides an introduction to the field of industrial hygiene and to occupational health in general. The instructor focuses on introducing concepts, terminology, and methodology in the practice of industrial hygiene and identifies resource materials. The class would benefit those wishing to pursue a Master's degree in industrial hygiene, those wishing to complete a certificate in occupational health, or for students in allied health fields needing a basic understanding of industrial hygiene.</p>
http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/PrinciplesIndustrialHygiene/coursePage/syllabus/	Course Objectives	<p>Describe the legal, professional, and ethical framework for the practice of industrial hygiene</p>
http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/PrinciplesIndustrialHygiene/coursePage/syllabus/	Course Objectives	<p>Define basic terms and technical concepts integral to the practice of industrial hygiene</p>
http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/PrinciplesIndustrialHygiene/coursePage/syllabus/	Readings	<p>You can buy the textbook, Fundamentals of Industrial Hygiene 5th edition, National Safety Council Chicago, IL, from any online reseller or through the Matthews Johns Hopkins Medical Book Center</p>

Elaborado: Diana Molina, Paula Soto F.

Al analizar el proceso de segmentación de datos de cursos OCW podemos señalar que debido a: la extensa composición en número (4513 cursos), información académica (distribución del Syllabus) y falta de homogeneidad de estructura HTML entre cursos OCW hacen que el proceso resulte largo aunque esté debidamente automatizado; puesto que se necesita tomar una muestra al azar de OCWs por universidades para poder seleccionar en que tags se encuentra ubicada la información, estableciendo de esta manera patrones que ayudan a dividir el contenido académico apropiadamente.

3.3. Categorización de data procesada

Una vez concluida la segmentación de datos, la siguiente tarea con el objetivo de ir consolidando la información es la categorización de la data, esta clasificación es una experiencia de las formas de ordenar y de disponer la información de etiquetas que representen nuestras categorías con fines específicos.

El proceso de categorización juega un papel importante en cuanto permite establecer estructuras implícitas en el material almacenado, las cuales a futuro ayudan a establecer un lenguaje común para que se puedan diferenciar los componentes académicos de los OCW y así las máquinas y personas puedan procesar de mejor manera esa información facilitando actividades.

“Categorizar es poner juntas las cosas que van juntas. Es agrupar datos que comportan significados similares. Es clasificar la información por categorías de acuerdo a criterios temáticos referidos a la búsqueda de significados. Es conceptualizar y codificar, con un término o expresión que sea claro e inequívoco, el contenido de cada unidad temática con el fin de clasificar, contrastar, interpretar, analizar, teorizar”, (Galeano Marín, 2003)

En el dominio de trabajo OCW, cada institución tiene diferente terminología para organizar la información del Syllabus, el caso más evidente de la importancia de categorizar los datos es la del vocablo⁷⁴ Objetivos; se puede evidenciar que esta palabra indistinta del idioma del curso tiene varios equivalentes tales como Objective, Goal, Skills, Competencias, Propósitos y Learning Outcomes.

Basándonos en el concepto y características de construcción de categorías mencionados en el apartado superior, se ha procedido a clasificar o categorizar la data en 10 grupos, los cuales abarcan toda la información segmentada.

⁷⁴ Vocablo: Palabra, como expresión de una idea [http://es.thefreedictionary.com/vocablo]

En la tabla 3.3 se visualiza algunas de las palabras que cumplen con la correspondencia y exclusión en significado a la categoría Learning outcomes.

Tabla 3.3 Palabras correspondientes y excluyentes para la categoría de Learning outcomes.

Learning Outcomes	
Palabra correspondientes a la categoría	Palabra excluyente a la categoría
Course Objetive	Curse Information
Objetivos del curso	Assigment
Goal	Seminar
Skill	Feedback
Competencias	Resource
Propósito	Activity

Elaborado: Diana Molina, Paula Soto F.

Para este fin se creó un programa llamado categorizacion.py; el cual compara y analiza cada registro de la BD bajo parámetros de caracterización de agrupación de palabras definidas por el usuario, enviando esta información a un nuevo registro con la categoría asignada. En el siguiente diagrama de flujo (Figura 3.6) se observa el proceso de categorización de datos.

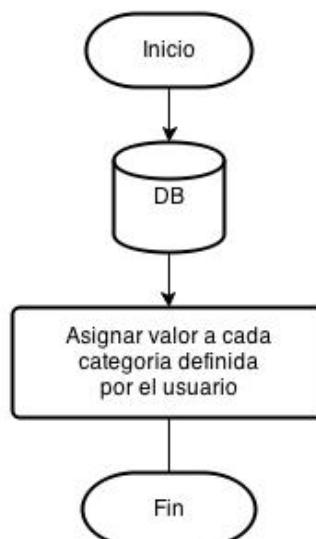


Figura 3.6 Diagrama de flujo del proceso de categorizacion.py
Elaborado: Diana Molina, Paula Soto F

Para el funcionamiento del programa categorizacion.py se necesita los siguientes elementos que hacen posible el proceso de categorización de los datos:

- ***Tabla _categorizador***

Esta tabla está alojada en la BD y en esta se ubica las reglas o restricciones de acuerdo a la categoría creada (ver tabla 3.4)

Tabla 3.4 Palabras de correspondencia y exclusión de las categorías Course Information y Learning Outcomes (tabla_categorizador)

Palabras de correspondencia	Palabras excluyentes	Predicado	Objeto
%Description%,%Introduction%, %Summary%,%Preparation%, %Overall%,%About%Course%,	%Assignment%,%Participation %,%Lecture%Preparation%,%Lecture%Introduction%,.....	tipo	Course Information
%Goal%, %Objective%, %Purpose%, %Mission%, %Aims%, %Outcome%, %Skills%,%Objetivos%,%ACTITUDES,.....	%Submission%, %Assignments%,%Feedback,%requisito%,%Defensa%competicencia%,%competencia%desleal%,%competencia%empresarial%,.....	tipo	Learning outcomes

Elaborado: Diana Molina, Paula Soto F

- ***categorizacion.py***

Módulo que contiene métodos para categorizar la data segmentada de acuerdo a las reglas definidas anteriormente en tabla_categorizador y guarda los resultados en la BD en forma de tripletas.

El resultado del proceso de categorización aplicado a los datos almacenados en la BD se aprecia en la Figura 3.7.

sujeto	predicado	objeto
Class Attendance (for students enrolled in thi...	tipo	Policy and conduct
Class Attendance and Participation	tipo	Policy and conduct
Class Blog	tipo	Resources
Class Components	tipo	Course Information
Class Debates and Presentations	tipo	Activities
Class Description	tipo	Course Information
Class Description and Learning Objectives	tipo	Learning outcomes
Class Diary Assignment	tipo	Activities
Class Discussion	tipo	Activities
Class Discussions	tipo	Activities
Class Exercises	tipo	Activities
Class Expectations	tipo	Learning outcomes

Figura 3.7 Ejemplo de datos OCW categorizados.
Elaborado: Diana Molina, Paula Soto F

El producto de los datos procesados anteriormente está distribuido en 10 tablas alojadas en una base relacional denominada Ontología, ver Figura 3.8.

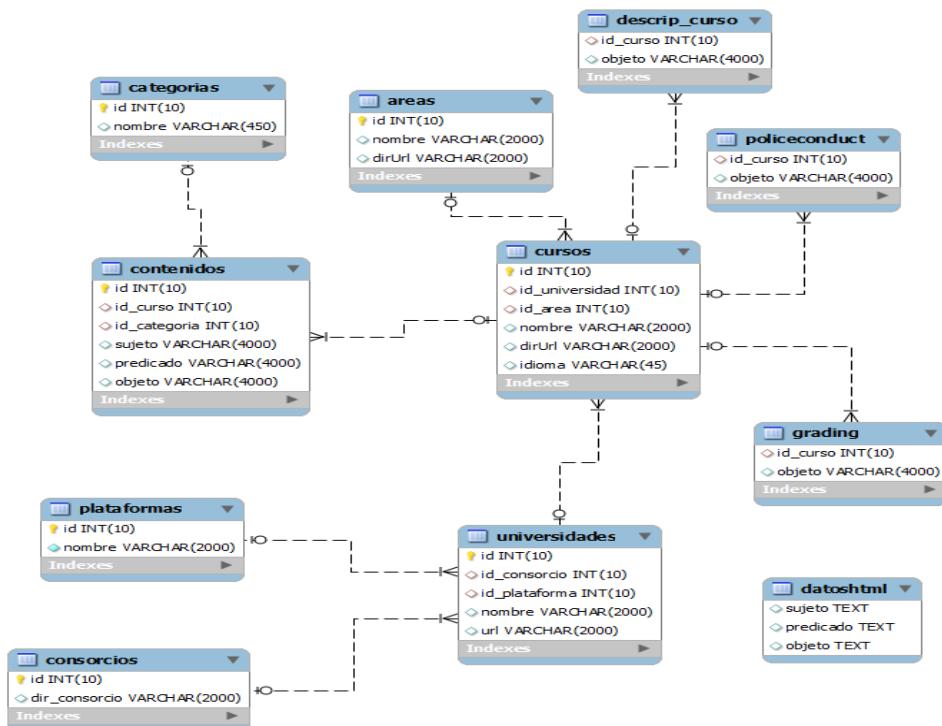


Figura 3.7 Modelo relacional de la base de datos “Ontología”
Elaborado: Diana Molina, Paula Soto

Al finalizar el proceso de categorización podemos acotar que aunque cada institución OCW tiene diferente terminología para nombrar la información académica de los OCWs no existe impedimento para agrupar los datos, ya que se realiza un trabajo de análisis de palabras para poder establecer y reunir la información de acuerdo a sus similares en concepto.

CAPÍTULO IV: IMPLEMENTACIÓN DE VOCABULARIO DOCW

4.1. Adaptación y modelado de vocabulario para describir planes de OCWs

“El modelamiento y descripción de los datos, entidades y relaciones de los recursos educativos y de los tópicos específicos que traten, mediante RDF; la posterior publicación de estos datos bajo principios de Linked Data; y la creación de enlaces con otros datos y entidades de la Web (como profesores, universidades, localizaciones, áreas de conocimiento, etc.), permitirán a un buscador encontrar y recuperar resultados mucho más relevantes para un usuario final.” (Educación Superior Virtual Inclusiva – América Latina, 2013)

Basándonos en el objetivo principal de este proyecto, el cual es la extracción de datos de cursos ofrecidos bajo la modalidad OCW, nuestro vocabulario ha sido denominado DOCW (Descripción de cursos OpenCourseWare), este vocabulario nombre clases y propiedades utilizando la tecnología W3Cs RDF, representando de una manera genérica recursos académicos como objetivos, descripción de curso, políticas, calendario, syllabus, etc.

Se empezó con la identificación de ontologías que puedan ser reutilizadas para poder representar los datos de los cursos OCW. Bajo este concepto se trabaja con vocabularios propiedad de la UTPL como LOCWD, IPD y vocabularios consensuados como TEACH, FOAF, BIBO, AIISO, DC.

Analizando detalladamente la información relacional y los vocabularios existentes se ha reutilizado las siguientes especificaciones (ver tablas 4.1 - 4.8) que serán representadas en el nuevo vocabulario.

Tabla 4.1 Clases y propiedades reutilizadas del Vocabulario TEACH

Nombre	Tipo	Descripción
courseDescription	Propiedad	Describe la información del curso.
courseTitle	Propiedad	Presenta el nombre del curso.
Grading	Propiedad	Representa el sistema de calificación en porcentaje
Teacher	Clases	Profesor o tutor del curso
Course	Clase	Curso didáctico por una organización educativa en la modalidad OCW.

Elaborado: Diana Molina, Paula Soto F.

Tabla 4.2 Clases y propiedades reutilizadas del Vocabulario FOAF

Nombre	Tipo	Descripción
Name	Propiedad	El nombre del docente
		El nombre de la Universidad

Elaborado: Diana Molina, Paula Soto F.

Tabla 4.3 Clases y propiedades reutilizadas del Vocabulario LOCWD

Nombre	Tipo	Descripción
EducationalOrganization	Clase	Representación de un centro educativo
oerPlataform	Propiedad	Sistema de repositorio de cursos OCW.
orgURL	Propiedad	Url del Consorcio
url	Propiedad	Url Universidad, Curso
Language	Propiedad	Idioma del curso

Elaborado: Diana Molina, Paula Soto F.

Tabla 4.4 Clases y propiedades reutilizadas del Vocabulario IPD

Nombre	Tipo	Descripción
KnowledgeArea	Propiedad	Área de conocimiento
Requisite	Clase	Conocimiento requerido para la iniciación del curso.
isOfferedBy	Propiedad	Es ofrecido por Universidad
Topic	Propiedad	Temas de estudio de un Curso
Content	Propiedad	Contenidos de estudio de un curso
Activity	Propiedad	Actividades complementarias del syllabus
Syllabus	Clase	Contenido del curso que se imparte
AcademicCalendar	Clase	Distribución del tiempo del curso
LearningOutcome	Clase	Objetivos de aprendizaje.

Elaborado: Diana Molina, Paula Soto F.

Tabla 4.5 Clases y propiedades reutilizadas del Vocabulario DC

Nombre	Tipo	Descripción
isPartOf	Propiedad	Parte del Syllabus
Description	Propiedad	Recursos que tiene un curso
		Requisitos de un curso.
		Calendario Académico de un curso.
		Objetivos de aprendizaje de un curso
		Criterios de evaluación

Elaborado: Diana Molina, Paula Soto F.

Tabla 4.6 Clases y propiedades reutilizadas del Vocabulario AIISO

Nombre	Tipo	Descripción
Subject	Clase	Elementos de evaluación utilizados dentro del curso

Elaborado: Diana Molina, Paula Soto F.

Tabla 4.7 Clases y propiedades reutilizadas del Vocabulario BIBO

Nombre	Tipo	Descripción
Resource	Clase	Recursos asociados al aprendizaje del curso.
Document	Clase	Recursos de tipo Documentos.
Activity	Clase	Actividades que complementan el curso.

Elaborado: Diana Molina, Paula Soto F.

Tabla 4.8 Clases y propiedades reutilizadas del Vocabulario DOCW

Nombre	Tipo	Descripción
taughtBy	Propiedad	Enseñado por
policeAndConduct	Propiedad	Políticas de un curso
hasResource	Propiedad	Recursos de un curso
Bibliography	Clase	Bibliografía de un curso
isRequired	Propiedad	Es requerido por el curso
urlArea	Propiedad	Url Area de Conocimiento
Offer	Propiedad	Ofrece cursos
consistingOf	Propiedad	Consiste de un syllabus
Timeline	Propiedad	Tiempo del calendario académico
Outcomes	Propiedad	Objetivos de aprendizaje de un syllabus.
evaluationCriteria	Propiedad	Criterios de evaluación del Syllabus.

Elaborado: Diana Molina, Paula Soto F.

Las clases creadas para esta ontología se describen a continuación.

- **Educational Organization**

Representación de un centro educativo, en nuestro caso una universidad

- **Course**

Curso dictado por una organización educativa en la modalidad OCW

- **Syllabus**

Contenido del curso que se imparte

- **Subject**

Determina los elementos de evaluación utilizados dentro del curso

- **LearningOutcome**

- Objetivos de aprendizaje que posee el curso

- **AcademicCalendar**

Información relacionada con la distribución en el tiempo del curso

- **Requisite**

Conocimientos requeridos para la iniciación de un curso

- **Resource**

SubClase de Bibliography para indicar recursos asociados al aprendizaje del curso

- **Document**

SubClase de Bibliography que especifica los recursos del tipo documentos

- **Activity**

SubClase de Bibliography que representa actividades que complementan el curso

- **Teacher**

Profesor, tutor o persona relacionada académicamente con el curso

La unión y estructura del vocabulario se la realiza mediante la herramienta Protégé⁷⁵.

Se han definido un conjunto de propiedades tanto de datos como de objetos que permitirán representar tanto la información como la relación que existe entre la misma dentro del vocabulario.

Un vocabulario posee dos tipos de propiedades que se pueden asociar a una clase, que pueden ser Propiedades Objeto y Propiedades de Datos; en la Figura 4.1 se visualiza estas propiedades.

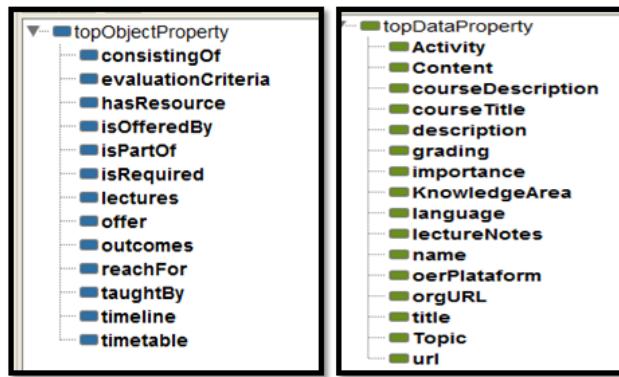


Figura 4.1 Object Property y Data Property de la ontología DOCW
Elaborado: Diana Molina, Paula Soto F.

Terminada la identificación, listado, y descripción de cada uno de los términos del Syllabus, se procedió a elaborar el vocabulario DOCW, cuya representación se puede visualizar en la Figura 4.2. Dentro de cada clase se distinguen los prefijos de los vocabularios que se han reutilizado (entre otros: FOAF, DC, AIISO, LOCWD)

⁷⁵ <http://protege.stanford.edu/>

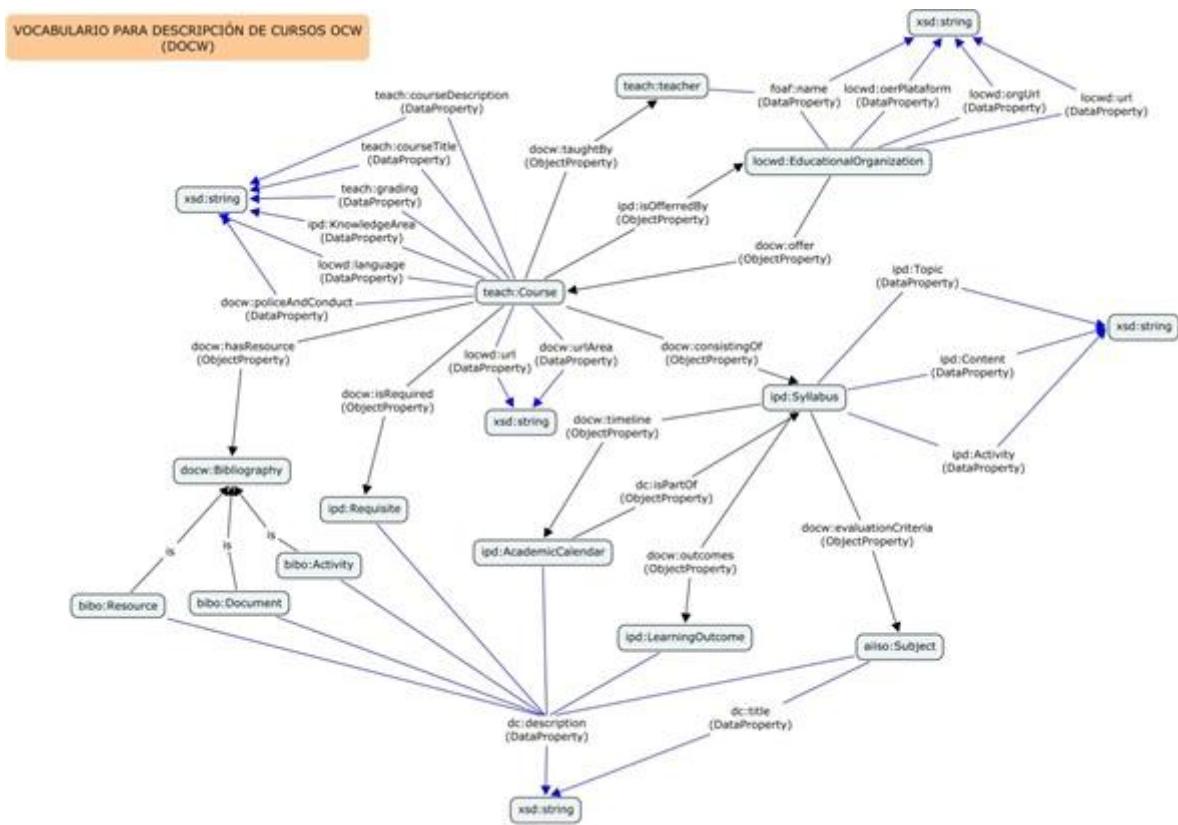


Figura 4.2 Vocabulario para la descripción de cursos OCW (DOCW)
Elaborado: Diana Molina, Paula Soto F.

“Antes de decidir sobre las herramientas a utilizar para la publicación de los datos, se debe conocer el dominio de datos a publicar, su nivel de cambios, su grado de confidencialidad y por supuesto su origen. Para datos existentes en bases de datos relacionales, planillas Excel o páginas web existen varias herramientas accesibles para generar su transformación a la representación RDF deseada” (Guzmán, Motz, & Rodrigue da Siva, 2011).

Al finalizar el proceso de modelado del vocabulario DOCW determinamos que a excepción de los vocabularios LOCWD e IPD trabajados por la UTPL, no existe ningún otro vocabulario que haga referencia a información sobre recursos OCW, lo cual demuestra el compromiso de la UTPL en el campo de la investigación y desarrollo de proyectos relacionados con Web Semántica y Linked Data liderando el movimiento a nivel de América Latina.

4.2. Generación de datos RDF

Una vez elegida la representación de los datos y la información categorizada de los cursos OCW, se debe elegir la herramienta para generar la información en formato RDF.

Aunque en el mercado encontramos herramientas para generación de archivos RDF como Google Refine⁷⁶, existen limitantes en cuanto a la relación entre la cantidad y estructura de datos; por ello se decide crear manualmente el archivo .RDF ya que en la programación se controla la jerarquización de datos y relaciones entre los mismos.

Para la generación de la información en formato RDF se la realiza volcando el contenido de la base de datos a un fichero RDF.

- Se parte desde la base de datos relacional (ver Figura 3.7).
- Se realiza un proceso de relación directa entre las columnas de la base de datos, las clases y propiedades del vocabulario.
- Se definen abreviaturas para cada clase de individuos y se maneja el siguiente estándar:

abrev00000

Donde “abrev” corresponde a la abreviatura de la clase (ver Figura 4.3) y “00000” hace referencia a un identificador único para el individuo

```
//Abreviaturas para las clases de los objetos.  
abrev_teacher      = "teach";  
abrev_EducOrganiz = "eduorg";  
abrev_Course       = "course";  
abrev_Resource     = "resor";  
abrev_Document     = "docum";  
abrev_Activity     = "activ";  
abrev_Requisite    = "requis";  
abrev_Syllabus     = "syllb";  
abrev_AcadCal      = "acadc";  
abrev_LearOut      = "lerot";  
abrev_Subject      = "subj";
```

Figura 4.3 Abreviaturas para las clases de los objetos
Elaborado: Diana Molina, Paula Soto F.

4.2.1. Creación de aplicativo para generación de fichero RDF.

Para la construcción del aplicativo de generación de RDF se ha utilizado el lenguaje de programación JAVA⁷⁷ realizando la codificación bajo el IDE NetBeans⁷⁸.

El objetivo del aplicativo es obtener un archivo de salida con los individuos (instancias) del vocabulario a partir de la Base Relacional.

⁷⁶ <http://openrefine.org/>

⁷⁷ https://www.java.com/es/download/faq/whatis_java.xml

⁷⁸ <https://netbeans.org/>

Para la comprensión general del proceso (ver Figura 4.4) presentamos el esquema de trabajo para la obtención del fichero RDF.

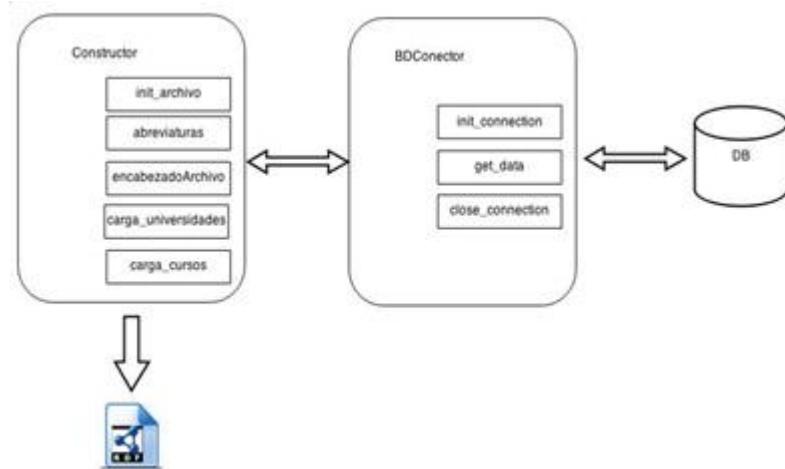


Figura 4.4 Esquema para obtención de archivo RDF
Elaborado: Diana Molina, Paula Soto F.

El aplicativo se encuentra contenido bajo el nombre de “**Constructor**” y dentro del paquete “**source.code**”. Conformado por dos clases:

- **BDConector**: Encargada de la comunicación con la Base de Datos relacional con los siguientes métodos (ver Figura 4.5)
 - **init_connection**: Encargado de establecer la conexión inicial con la Base de Datos
 - **get_data**: Ejecuta la sentencia SQL⁷⁹ enviada como parámetro y devuelve los resultados en una estructura del tipo ResultSet.
 - **close_connection**: Cierra la conexión y libera los recursos de memoria.

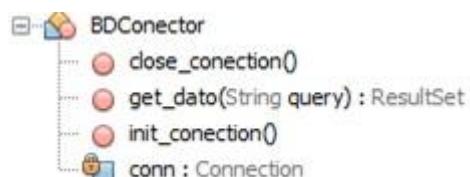


Figura 4.5 Métodos de BDConector
Elaborado: Diana Molina, Paula Soto F.

- **Constructor**: La clase principal (ver Figura 4.6) y encargada de construir el archivo RDF, para lo cual utiliza los siguiente métodos.

⁷⁹ es.wikipedia.org/wiki/SQL

- **ini_archivo:** Crea un archivo vacío en el directorio especificado y lo enlaza a una variable para que el aplicativo pueda escribirlo.
- **abreviaturas:** Carga en memoria todas las abreviaturas de los diferentes individuos que serán utilizados en la generación del RDF.
- **encabezadoArchivo:** Escribe en el archivo el encabezado correspondiente a la definición de los vocabularios y sus abreviaturas.
- **carga_universidades:** Crea los individuos de la clase EducationalOrganization por considerarse una clase menos dependiente.
- **carga_cursos:** Se trasladan todos los objetos restantes y dependientes iniciando por los cursos y sus propiedades, considerando que la generación de objetos dependientes es almacenada en memoria hasta que todos los cursos hayan sido escritos al archivo.



Figura 4.6: Clase constructor
Elaborado: Diana Molina, Paula Soto F.

El proceso de generación dentro del aplicativo sigue un orden jerárquico de mayor a menor; donde se procederá a generar primero la clase universidades y se terminará con las propiedades de cada uno de los cursos.

Este proceso ordenado permite que la estructura del archivo pueda tener una relación lógica con cada una de sus partes.

En el aplicativo utilizamos consultas SQL (ver Figura 4.7) que nos permitirán extraer la información necesaria para posteriormente procesarla.

```
String consulta = "SELECT *  
+ " c.id, c.id_universidad, c.id_area, *  
+ " c.nombre nomb_curso, c.dirUrl dir_curso, "  
+ " c.idioma, a.nombre nomb_area, "  
+ " a.dirUrl url_area"  
+ "FROM cursos c, areas a"  
+ "where a.id = c.id_area";
```

Figura 4.7 Consulta SQL para extraer Cursos
Elaborado: Diana Molina, Paula Soto F.

La salida del aplicativo es un archivo en formato RDF (ver Figura 4.8), que contiene toda la información de la base relacional ahora en formato que un Store de triplets puede almacenar.



```
1 <?xml version="1.0"?>
2 <!DOCTYPE rdf:RDF [
3   !ENTITY terms "http://purl.org/dc/terms/" ;
4   !ENTITY foaf "http://xmlns.com/foaf/0.1/" ;
5   !ENTITY property "http://dbpedia.org/property/" ;
6   !ENTITY bibo "http://purl.org/ontology/bibo/" ;
7
8   <rdf:RDF xmlns="#"
9     xmlns:base="http://direccion.ontologia/rdf/"
10    xmlns:dc="http://dublincore.org/documents/dcmi-terms/#"
11    xmlns:foaf="http://xmlns.com/foaf/0.1/" ;
12
13   <owl:NamedIndividual rdf:about="#course00470">
14     <rdf:type rdf:resource="#teach:Course"/>
15     <ipd:KnowledgeArea rdf:datatype="#xsd:string">Philosophy</ipd:KnowledgeArea>
16     <docw:urlArea rdf:datatype="#xsd:string">http://ocw.nd.edu/philosophy</docw:urlArea>
17     <locwd:url rdf:datatype="#xsd:string">http://ocw.nd.edu/philosophy/ancient-and-medieval
18     <locwd:language rdf:datatype="#xsd:string">English</locwd:language>
19
20
21
22
23
24
25
26
27
28
29
```

Figura 4.8 Extracto de archivo final
Elaborado: Diana Molina, Paula Soto F.

El proceso de generación de datos RDF depende del origen y estructura de los mismos, con este fin se elabora un programa que describe una relación directa entre la BD y el modelado del vocabulario DOCW, generando un archivo RDF que contiene la información académica de los cursos OCW.

4.2.2. Validación del código RDF

Obtenido el archivo docw_individuals.rdf, se procede a validar (ver Figura 4.9) el código RDF utilizando el validador oficial de W3C llamada RDF Validator⁸⁰. La validación de datos consiste en verificar que el archivo RDF cumple con la sintaxis de una tripla para alcanzar datos enlazados de calidad.

⁸⁰ http://www.w3.org/2001/sw/wiki/RDF_Validator

The screenshot shows the 'Validation Service' interface. At the top, there are links for 'Home', 'Documentation', and 'Feedback'. Below that, a banner says 'Check and Visualize your RDF documents'. A text area contains the RDF code:

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF I
<!ENTITY terms "http://purl.org/dc/terms#" >
<!ENTITY foaf "http://xmlns.com/foaf/0.1/#" >
<!ENTITY property "http://dbpedia.org/property/" >
<!ENTITY bibo "http://purl.org/ontology/bibo/" >
<!ENTITY owl "http://www.w3.org/2002/07/owl#" >
<!ENTITY alias "http://vocab.org/alias/#names#" >
<!ENTITY docw "http://dixedicion.ontologia.rdf/" >
<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
```

Below the code are buttons: 'Parse RDF', 'Restore the original example', and 'Clear the textarea'. Under 'Display Result Options', there are dropdown menus for 'Triples and/or Graph' (set to 'Triples Only') and 'Graph format' (set to 'PNG - embedded').

Figura 4.9 Muestra del código docw_individuals.rdf usando RDF Validator para proceder a validarla

Elaborado: Diana Molina, Paula Soto F.

En la Figura 4.10 se observa el resultado positivo al procesar el documento RDF en la herramienta online RDF Validator

Validation Results		
Triples of the Data Model		
Number	Predicate	Object
1	http://dixedicion.ontologia.rdf#resource13879 http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2000/01/rdf-schema#Resource
2	http://dixedicion.ontologia.rdf#resource13879 http://www.w3.org/1999/02/22-rdf-syntax-ns#label	"Earth, Atmosphere, and Planetary Sciences"
3	http://dixedicion.ontologia.rdf#resource13879 http://www.w3.org/1999/02/22-rdf-syntax-ns#comment	"This course covers the fundamental methods used for exploring the indoor observations calculated on kilometric and diurnal scales. It also covers the basic concepts of atmospheric and planetary sciences."
4	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#isSubclassOf	"Earth, Atmosphere, and Planetary Sciences"
5	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasLanguage	"English"
6	http://dixedicion.ontologia.rdf#resource13879 http://www.w3.org/2002/07/owl#sameAs	"Indoor Air Quality"
7	http://dixedicion.ontologia.rdf#resource13879 http://www.w3.org/2002/07/owl#isDefinedBy	"http://www.w3.org/2000/01/rdf-schema#Resource"
8	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"greeting"
9	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"This course covers the fundamental methods used for exploring the indoor observations calculated on kilometric and diurnal scales. It also covers the basic concepts of atmospheric and planetary sciences."
10	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
11	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"This course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
12	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
13	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"Because of the nature of material to be treated, there is little time to assess its applications over studies. But examples from various branches of science will be discussed."
14	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
15	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
16	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
17	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
18	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
19	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
20	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
21	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
22	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
23	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."
24	http://dixedicion.ontologia.rdf#resource13879 http://dixedicion.ontologia.rdf#hasText	"The course is oriented at making scientifically sensible reductions from observations with dynamics and kinematics represented, generally, as non-overlapping central themes."

Figura 4.10 Resultado del documento RDF validado
Elaborado: Diana Molina, Paula Soto F.

En la Figura 4.11 presentamos el grafo de una Universidad con sus respectivas clases y propiedades utilizando RDF Validator.

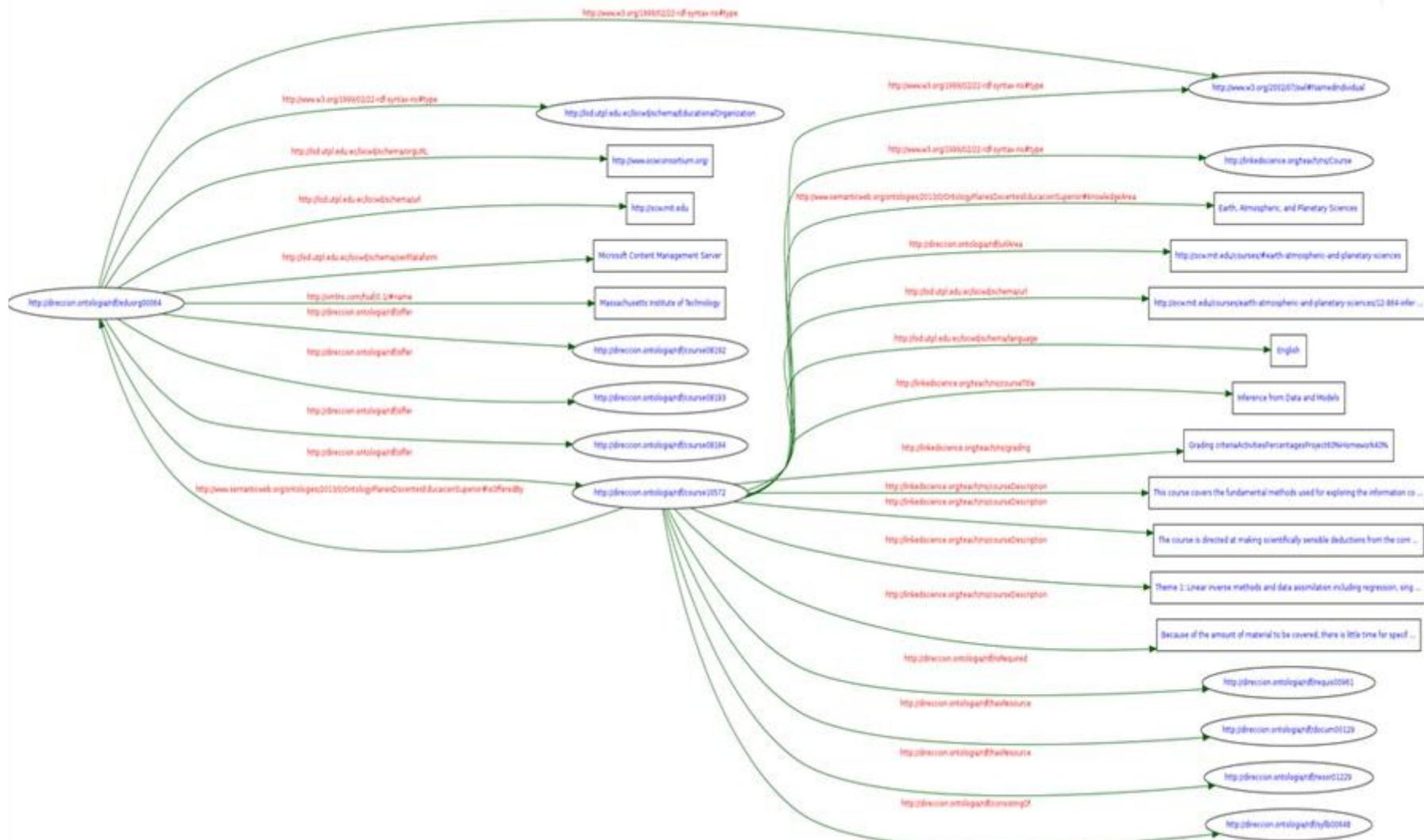


Figura 4.11 Grafo de una universidad utilizando RDF Validator
Elaborado: Diana Molina, Paula Soto F.

Al terminar la fase de implementación del vocabulario DOCW se argumenta que este vocabulario con sus clases y propiedades podrá ser reutilizado en futuros proyectos que tenga relación con REA de tipo OCW.

La información académica de los cursos OCW, previo proceso de segmentación y categorización en esta fase ya se encuentra disponible en un lenguaje que los ordenadores pueden entender, todas las tripletas que contiene el fichero RDF denominado docw_individuos.rdf están listas para ser utilizadas en buscadores semántico.

Los datos contenidos en el fichero RDF en el futuro podrán ser vinculados externamente con otros dataset que guardan relación específica con el vocablo denominado Syllabus.

CAPÍTULO V: ALMACENAMIENTO, ACCESO Y VISUALIZACIÓN DE DATOS

5. Almacenamiento, acceso y visualización de datos

Los datos RDF son fáciles de procesar por máquinas, pero puede resultar complicado para las personas comunes entender estos formalismos. Por esto se requieren herramientas que permitan acceder, consultar, visualizar y en general consumir esta información.

Hoy en día existen SGBD⁸¹ cuya orientación es semántica, por lo tanto están optimizadas para almacenar datos de este tipo, la principal característica es la capacidad que tienen para guardar tripletas, por esta razón se las llama bases de datos “triplestore”⁸².

Se destaca que los Datasets de las aplicaciones que se describen a futuro están enlazados internamente; se analizó los repositorios existentes y no se encontró información sobre Datasets que describan al Syllabus, lo cual dificultó el enlace externo de nuestros datos.

5.1. Almacenamiento de tripletas en TripleStore

Analizando las diferentes herramientas para almacenamiento de tripletas se prefiere utilizar Virtuoso⁸³ triplestore en su versión OpenSource⁸⁴, debido a que ofrece un sistema probado de almacenamiento eficiente, esta eficiencia involucra la carga de datos y la fiabilidad en cuanto a la integridad de dichos datos y sus relaciones; facilita la interacción con aplicaciones mediante peticiones HTTP, además contiene unEndPoint por medio del cual se recupera información en tripletas utilizando el lenguaje de consulta SPARQL.

Virtuoso-OpenSource forma parte de Virtuoso Universal Server el cual es un servidor multi-plataforma escalable que combina la gestión de datos SQL / RDF / XML, Gráfico, Gestión de datos y documentos con Web Application Server y la funcionalidad de la plataforma de servicios Web.

Una vez instalado⁸⁵ Virtuoso-OpenSource se debe definir un nombre para el grafo que se va a crear y donde se almacenará toda la información anteriormente procesada. En este caso el nombre del grafo será el siguiente:

<http://localhost:8890/descOCW>

⁸¹ SGBS: Sistema de gestión de base de datos [es.wikipedia.org/wiki/Sistema_de_gestión_de_bases_de_datos]

⁸² <http://en.wikipedia.org/wiki/Triplestore>

⁸³ <http://www.openlinksw.com/dataspace/doc/dav/wiki/Main/>

⁸⁴ Opensource: Código abierto [[http://es.wikipedia.org/wiki/C%C3%B3digo_abierto](https://es.wikipedia.org/wiki/C%C3%B3digo_abierto)]

⁸⁵ <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSUbuntuNotes>

Mediante la interfaz del Conductor de Virtuoso dentro de la pestaña Linked Data seguido de Quad Store Upload (ver Figura 5.1) se puede fácilmente, seleccionar el archivo y cargarlo a un grafo que nos permitirá realizar consultas mediante SPARQL.

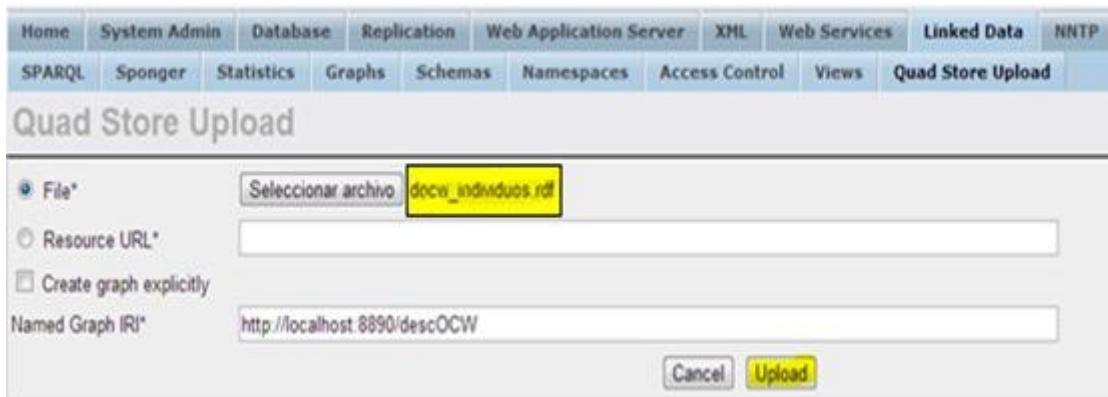


Figura 5.1 Carga de archivo RDF
Elaborado: Diana Molina, Paula Soto F

5.1.1. Explotación de datos mediante lenguaje de consulta SPARQL.

RDF es un formato de datos para grafos dirigidos y etiquetados para representar la información en la Web.

Esta especificación define la sintaxis y la semántica del lenguaje de consulta SPARQL para RDF.

SPARQL, lenguaje de consulta de la web semántica de W3C, 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 solicitadas basándose en la comparación de una tripleta pasada como parámetro de la consulta con todas las tripletas que componen el grafo RDF.

Para consumir los datos RDF generados mediante el proceso descrito en el capítulo anterior se realizan consultas mediante SPARQL desde el EndPoint de Virtuoso, y sus resultados, en este caso para mejor visualización son presentados en formato HTML.

A continuación presentamos algunas consultas de ejemplo.

- a. Consulta para extraer todas las universidades con sus respectivos nombres.

SPARQL

Default Data Set Name (Graph IRI)
<http://localhost:8890/descOCW>

Query Text

```

PREFIX docw: <http://direccion.ontologia/rdf/>
PREFIX locwd: <http://lod.utpl.edu.ec/locwd/schema/>
PREFIX teach: <http://linkedscience.org/teach/ns/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/#>

SELECT ?univers ?nombre
WHERE
{
    ?univers a locwd:EducationalOrganization;
              foaf:name ?nombre
}

```

Resultado:

univers	nombre
http://direccion.ontologia/rdf/eduorg00064	"Massachusetts Institute of Technology"^^<http://www.w3.org/2001/XMLSchema#string>
http://direccion.ontologia/rdf/eduorg00065	"University of Michigan"^^<http://www.w3.org/2001/XMLSchema#string>
http://direccion.ontologia/rdf/eduorg00066	"Universidad de Sevilla"^^<http://www.w3.org/2001/XMLSchema#string>
http://direccion.ontologia/rdf/eduorg00067	"Universidad Politécnica de Madrid"^^<http://www.w3.org/2001/XMLSchema#string>
http://direccion.ontologia/rdf/eduorg00068	"Universidad Carlos III de Madrid"^^<http://www.w3.org/2001/XMLSchema#string>
http://direccion.ontologia/rdf/eduorg00069	"Universidad de Murcia"^^<http://www.w3.org/2001/XMLSchema#string>
http://direccion.ontologia/rdf/eduorg00070	"Universidad de Cantabria"^^<http://www.w3.org/2001/XMLSchema#string>

- b. Consulta para extraer todos los cursos ofertados por la universidad Politécnica de Madrid.

SPARQL

Default Data Set Name (Graph IRI)
<http://localhost:8890/descOCW>

Query Text

```

PREFIX docw: <http://direccion.ontologia/rdf/>
PREFIX locwd: <http://lod.utpl.edu.ec/locwd/schema/>
PREFIX teach: <http://linkedscience.org/teach/ns/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/#>

SELECT ?universidad ?nomb_curso
WHERE
{
    ?docw:eduorg00067 docw:offer ?curso;
                         foaf:name ?universidad.
    ?curso teach:courseTitle ?nomb_curso
}

```

Resultado

universidad	nomb_curso
"Universidad Politécnica de Madrid"^^< http://www.w3.org/2001/XMLSchema#string >	"English for Professional and Academic Communication: Curso de Adaptación"^^< http://www.w3.org/2001/XMLSchema#string >
"Universidad Politécnica de Madrid"^^< http://www.w3.org/2001/XMLSchema#string >	"Kinantropometria"^^< http://www.w3.org/2001/XMLSchema#string >
"Universidad Politécnica de Madrid"^^< http://www.w3.org/2001/XMLSchema#string >	"Cálculo I"^^< http://www.w3.org/2001/XMLSchema#string >
"Universidad Politécnica de Madrid"^^< http://www.w3.org/2001/XMLSchema#string >	"Dibujo (Preparación para la Universidad)"^^< http://www.w3.org/2001/XMLSchema#string >
"Universidad Politécnica de Madrid"^^< http://www.w3.org/2001/XMLSchema#string >	"Física (Preparación para la Universidad)"^^< http://www.w3.org/2001/XMLSchema#string >
"Universidad Politécnica de Madrid"^^< http://www.w3.org/2001/XMLSchema#string >	"Introducción a los Sistemas de Tiempo Real"^^< http://www.w3.org/2001/XMLSchema#string >

Además, de las consultas básicas presentadas anteriormente, se han implementado otras consultas SPARQL (ver ANEXO 10), esto con el objetivo de contrastar cuantitativamente la información académica de los cursos OCW representada en el RDF.

- **Componentes OCW por total de Universidades**

La Figura 5.2 relaciona el número de universidades con los componentes de los cursos OCW (Syllabus, Academic Calendar, Subject, Learning Outcomes, Bibliography, Requisite).

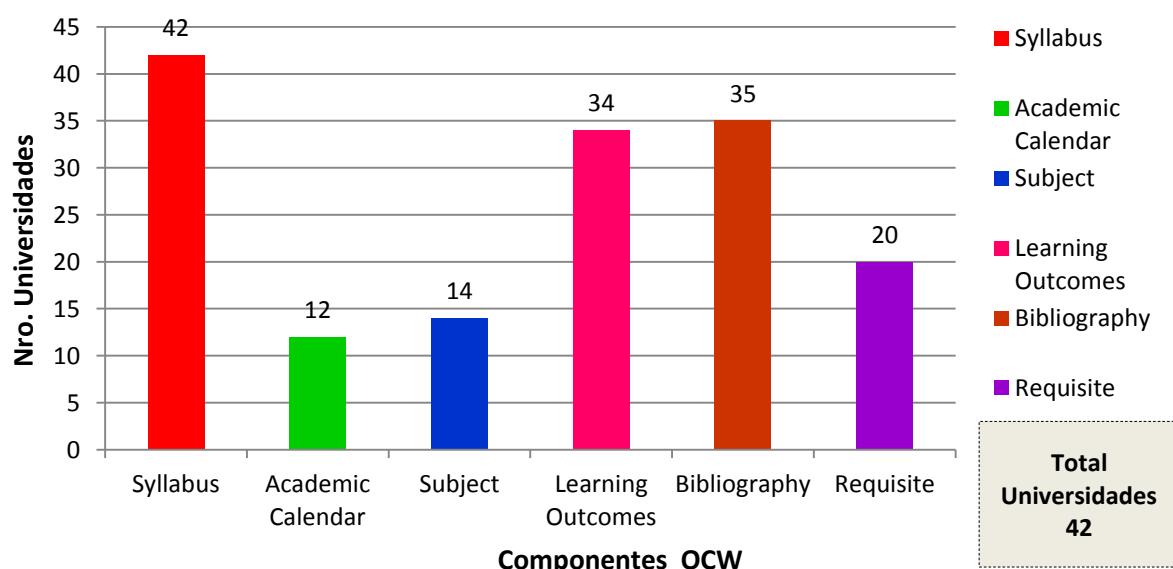


Figura 5.2 Componentes de cursos OCW por total de Universidades
Elaborado: Diana Molina, Paula Soto F

- **Componentes OCW por total de cursos**

En la figura 5.3 se observa el número de cursos que poseen los componentes de los cursos OCW (Syllabus, Academic Calendar, Subject, Learning Outcomes, Bibliography and Requisite).

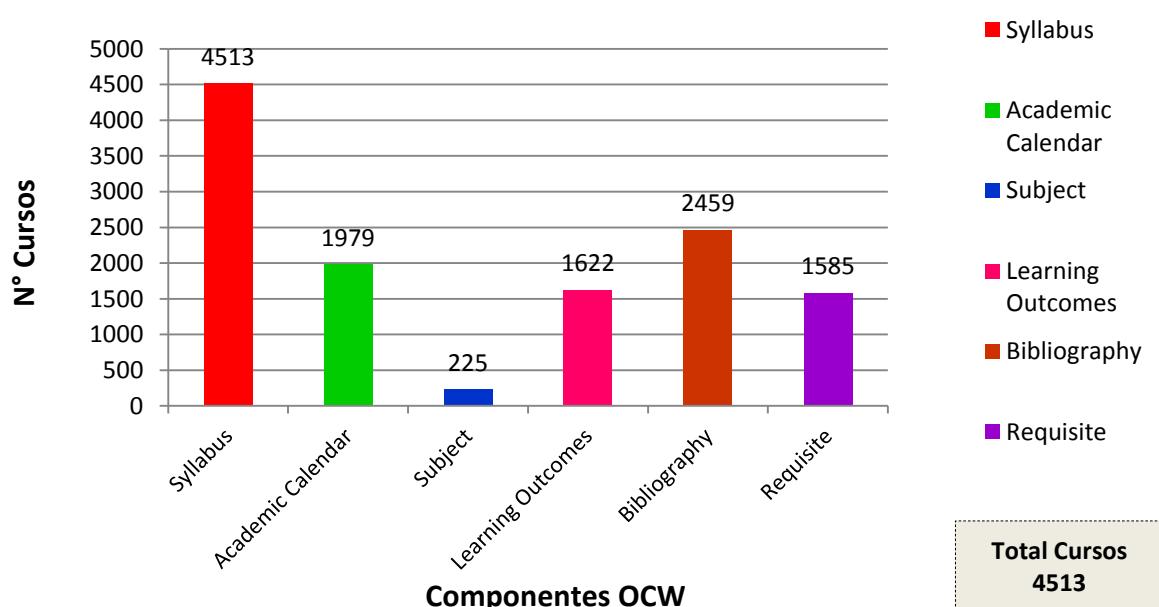


Figura 5.3 Componentes de cursos OCW por total de Cursos
Elaborado: Diana Molina, Paula Soto F

- **Componentes de cursos OCW por cada universidad.**

En la Figura 5.4 se observa la relación entre cada universidad y el número de componentes que tiene cada una. Para el gráfico se toma como muestra 7 universidades.

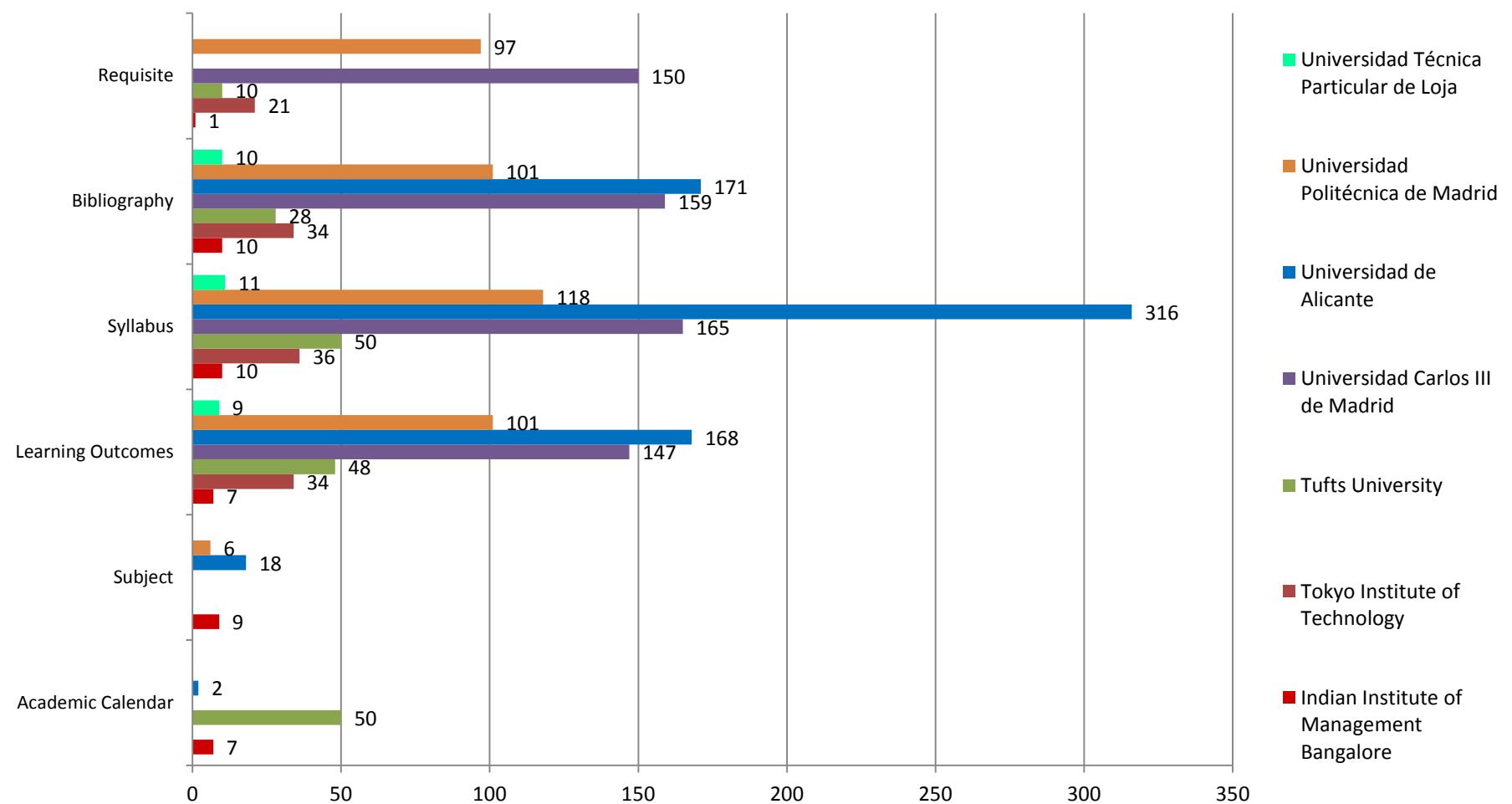


Figura 5.4 Comparativo de componentes de OCW por universidades (limitado el número de universidades)
Elaborado: Diana Molina, Paula Soto F

- **Número de cursos por área de conocimiento y componentes**

Por la extensa cantidad de cursos se ha tomado como referente un solo componente (Academic Calendar) y tres universidades con sus respectivas áreas de conocimiento (Figura 5.6).

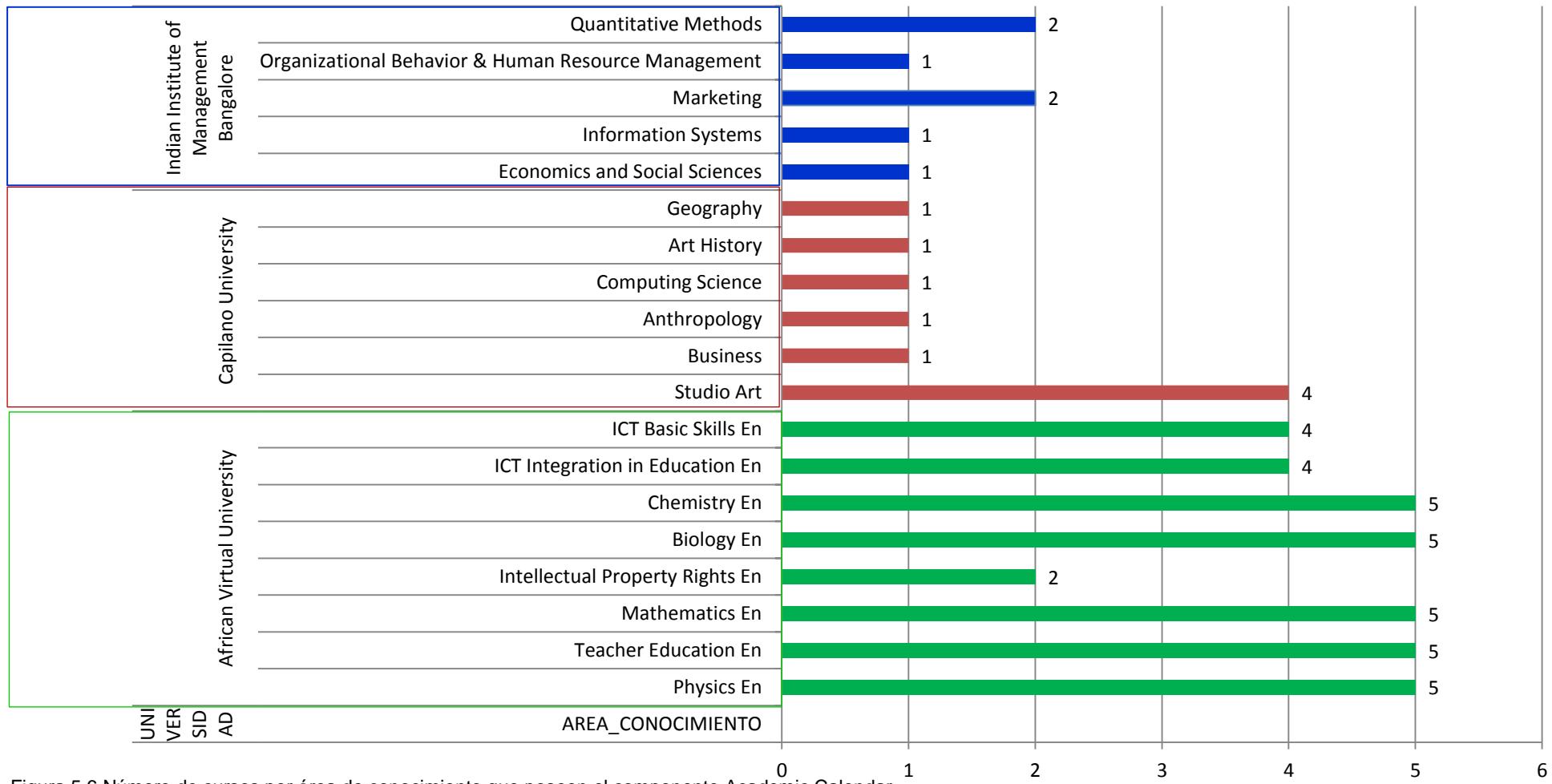


Figura 5.6 Número de cursos por área de conocimiento que poseen el componente Academic Calendar
Elaborado: Diana Molina, Paula Soto F.

Basándose en los datos generados por las consultas anteriores, se pueden realizar varias gráficas porcentuales sobre la estructura de los cursos OCW y sus componentes.

La Figura 5.7 presenta el porcentaje de los componentes OCW por total de las 42 universidades. Por ejemplo el 81% de universidades poseen Learning Outcomes (resultados de aprendizaje).

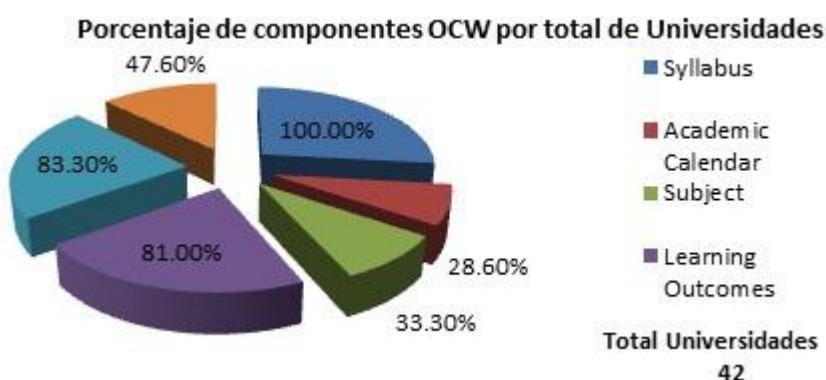


Figura 5.7 Porcentaje de componentes OCW por total de Universidades
Elaborado: Diana Molina, Paula Soto F.

En Figura 5.8 se presenta el porcentaje de los componentes OCW por total de los 4513 cursos. Por ejemplo el 54.44% de los cursos poseen Bibliography.

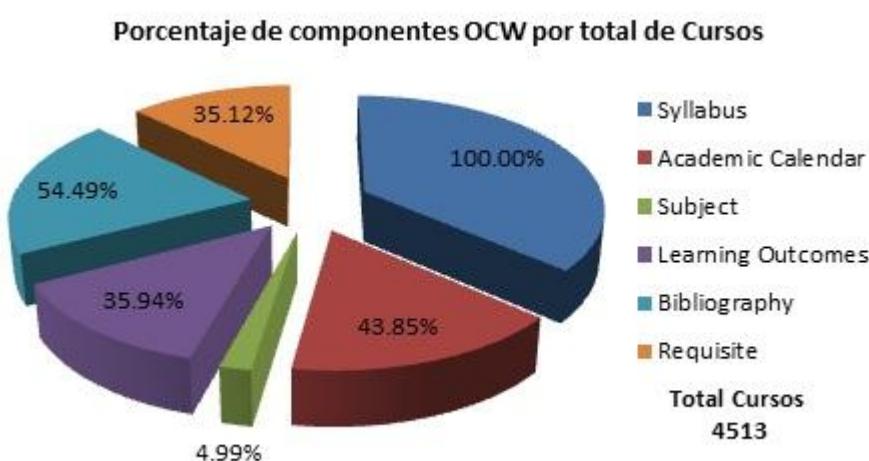


Figura 5.8 Porcentaje de componentes OCW por total de Cursos
Elaborado: Diana Molina, Paula Soto F.

En la Figura 5.9 se puede observar el porcentaje total de los componentes que posee dicha Universidad.

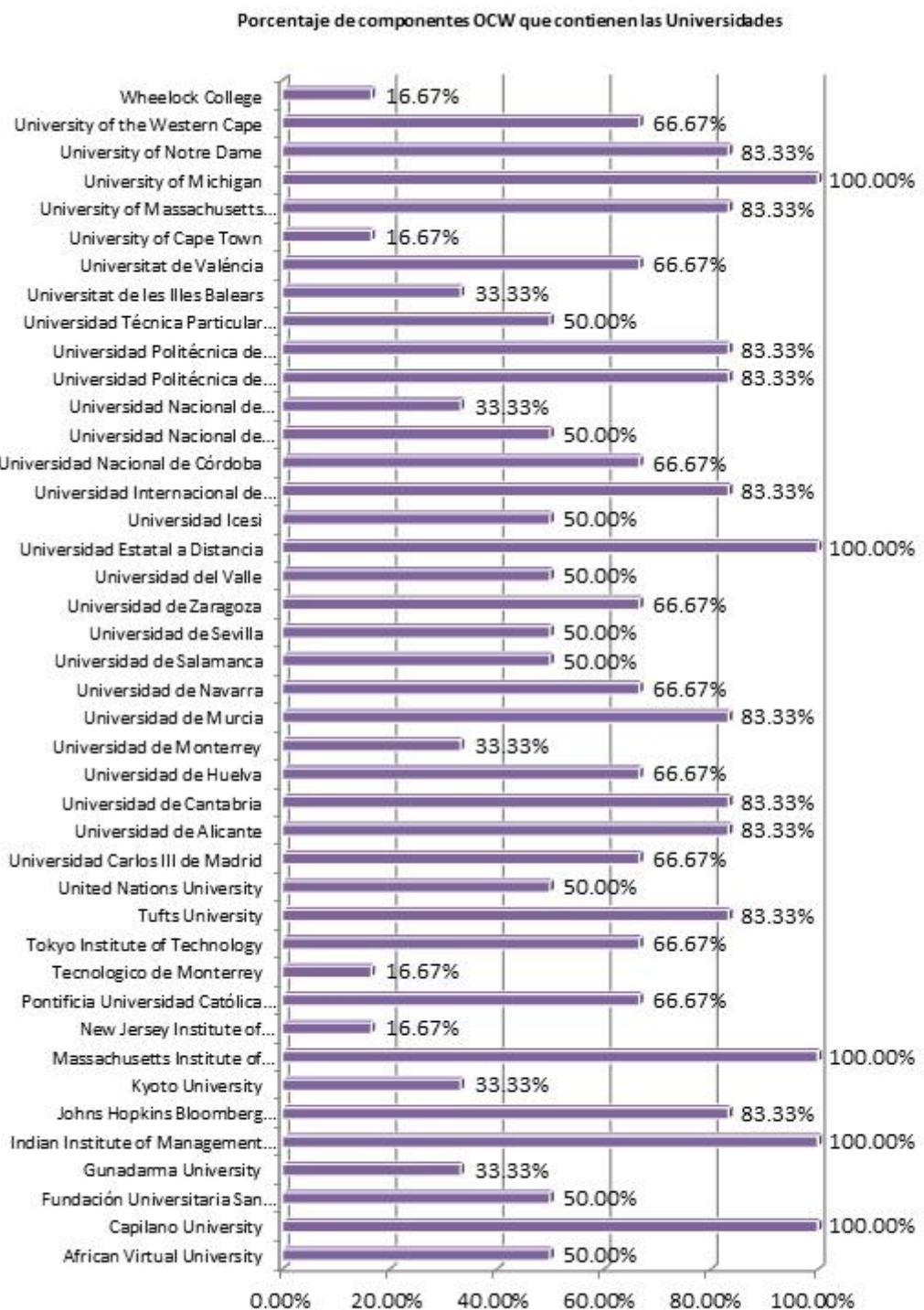


Figura 5.9 Porcentaje de componentes OCW que contienen las Universidades
Elaborado: Diana Molina, Paula Soto F.

La Figura 5.10 ayuda a establecer el valor porcentual de acuerdo al número de componentes. Es decir que si una universidad posee tres componentes esto equivale al 50% del global de componentes académicos. Esto es para tener una mejor lectura de la gráfica anterior (Figura 5.9).

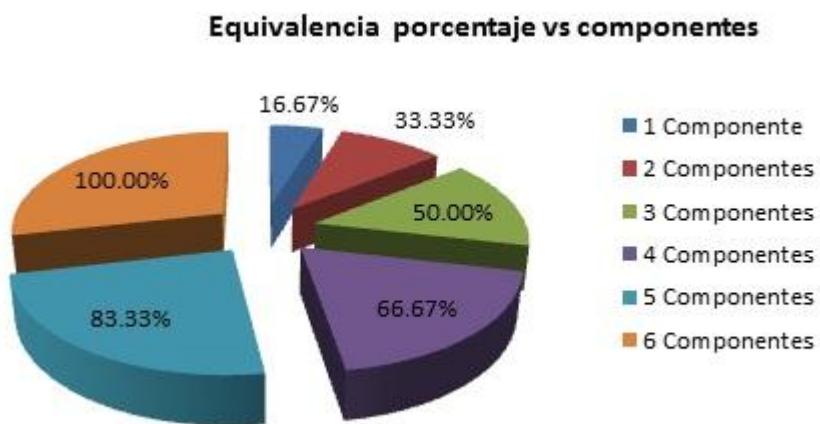


Figura 5.10 Equivalencia porcentaje VS componentes
Elaborado: Diana Molina, Paula Soto F.

En la Figura 5.11 se observa el número de componentes con respecto al total de universidades. Por ejemplo solo 4 universidades que corresponden al 10% del total de las 42 Universidades poseen un componente.

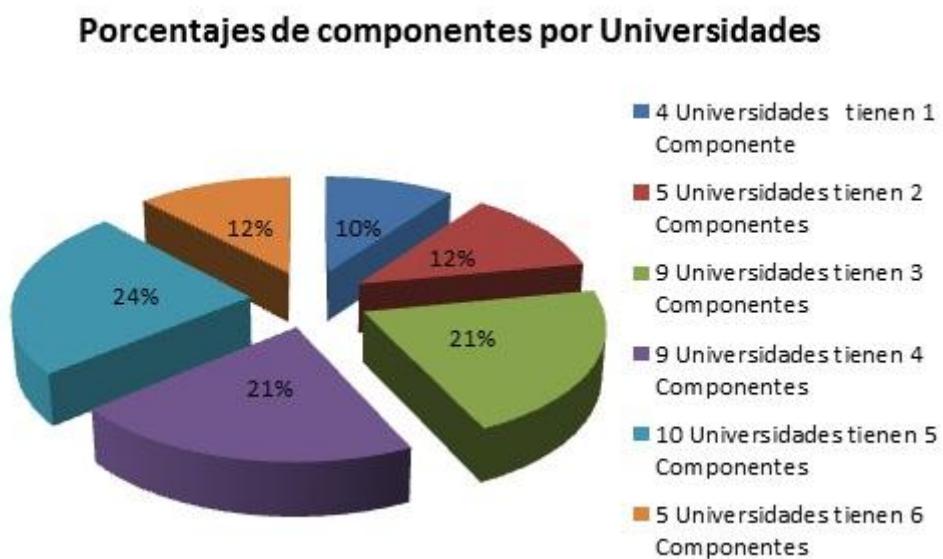


Figura 5.11 Porcentajes de componentes por universidades
Elaborado: Diana Molina, Paula Soto F.

De la exploración realizada sobre el fichero RDF mediante el lenguaje de consultas SPARQL y el posterior análisis de los datos recuperados por las consultas se concluye que:

- Los componentes académicos más representativos en los cursos OCW son: Syllabus, Bibliografía, Calendarios académicos, Resultados de aprendizaje, Requisitos y Criterios de evaluación.
- 5 universidades poseen el 100% de componentes de un syllabus: Capilano University, Indian Institute of Management Bangalore, Massachusetts Institute of Technology, Universidad Estatal a Distancia y University of Michigan.

Además, en cuanto al porcentaje de cumplimiento de los componentes de un syllabus, se encontró la siguiente distribución: 10 universidades tienen el 83.33%, 9 universidades tienen el 66.67%, 9 universidades tienen el 50%, 5 universidades poseen el 33.33% y 4 universidades tiene 16.67% de componentes del syllabus de un curso OCW.

5.2. Visualización de datos RDF

Como punto final para visualización de datos RDF, se escoge una herramienta que soporte la publicación y navegación por el modelo RDF mediante consultas SPARQL.

Hay varias formas de acceder a datos RDF, la más práctica y amigable es utilizar LODSPeaKr, este es un marco de código abierto para crear aplicaciones de Linked Data y publicar datos RDF de una manera menos compleja; entre las características importantes de este framework están la creación servicios basados en datos enlazados, provee la posibilidad de crear nuestro propio flujo de trabajo de ejecución de consultas SPARQL en múltiples EndPoints.

LODSPeaKr trabaja con servicios, de acuerdo a la petición que el usuario especifique en el servicio (consulta SPARQL), el núcleo de LODSPeaKr va contra elEndPoint del Triplestore (Virtuoso), realiza la consulta y devuelve la información en cualquiera de los contenidos que por defecto se pueden utilizar con LODSPeaKr (RDF, N-Triple, JSON, etc.). Ver Figura (5.12)

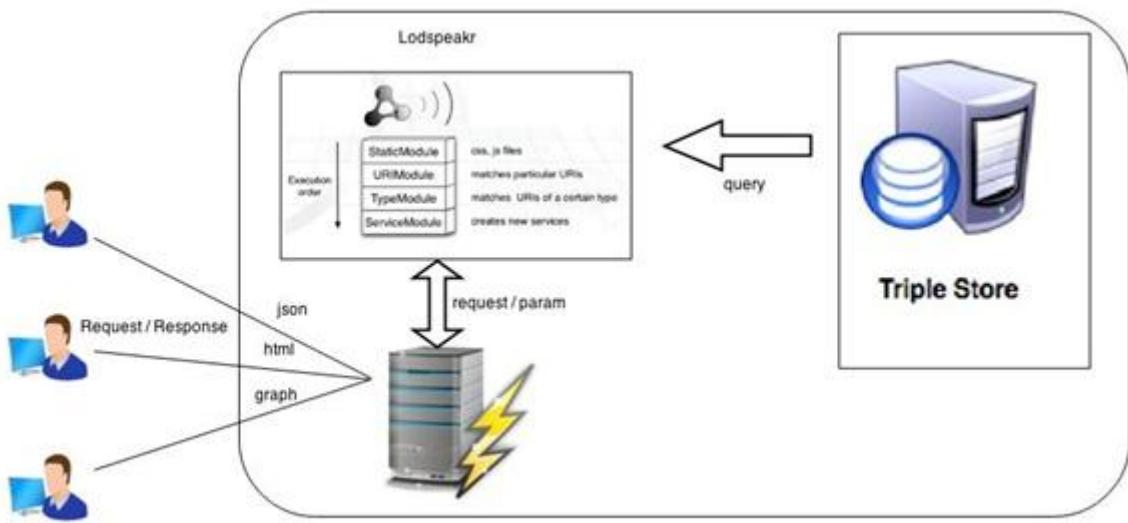


Figura 5.12 Esquema de trabajo de LODSPeakr
Elaborado: Diana Molina, Paula Soto F.

5.2.1. Creación de servicios usando LODSPeakr con la librería D3

Con LODSPeakr, se pueden crear tres tipos de componentes URIs, Type y Service, en nuestro caso creamos un componente de tipo Service ya que necesitamos que las aplicaciones nos devuelvan datos en el formato específico JSON.

LODSPeakr para poder graficar trabaja con: Google maps⁸⁶, gRaphael⁸⁷ y D3⁸⁸, debido a la naturaleza de nuestra información se utiliza D3.js ya que posee una galería diversa y además ha sido probada con éxito , mientras que las demás todavía se encuentran en fase de pruebas.

Como el objetivo del proyecto es mostrar la máxima información académica de los cursos OCW, se procedió a crear un servicio por cada clase presente en la ontología DOCW.

Para la presentación de los datos se crean cinco servicios principales, en cada una existe información relacionada directamente con la consulta SPARQL, por ejemplo el servicio topicCurso (Ver anexo 7) presenta los nombres de los cursos y este a su vez llamará al servicio detailCourse (Ver anexo 8), el cual mostrará los data property y data objective de ese curso, este a su vez al hacer click sobre sus data objective llamará a los servicios respectivos y así sucesivamente (Ver anexo 9).

⁸⁶ <https://maps.google.com.ec/>

⁸⁷ <http://g.raphaeljs.com/>

⁸⁸ <http://d3js.org/>

En la tabla 5.1 se detalla el nombre de los servicios creados y su descripción.

Tabla 5.1 Servicios creados con LODSPeakr

Nombre Servicio	Descripción
Topic	Muestra los nombres de los cursos que coincidan en el nombre del curso (parámetro).
topicCurso	Muestra los nombres de los cursos que coincidan en el syllabus o en el área de conocimiento (parámetro)
Universidad	Lista las universidades (parámetro).
Graficador	Lista las universidades.
graficalIndividuo	Muestra el vínculo entre Universidades y Cursos.
detailAcademicCalendar	Detalla los dataproperty y dataobjective de la clase a que su nombre hace referencia.
detailActivity	
detailCourse	
detailDocument	
detailRequirement	
detailResource	
detailSubject	
detailSyllabus	
detailTeach	
detailUniversity	

Elaborado: Diana Molina, Paula Soto F.

5.2.1.1. Creación de Servicios.

Para crear un servicio en general debemos ingresar al terminal y ejecutar el siguiente comando \$utils/lodspk.sh create service topicCourse con el respectivo nombre que se le va a dar al servicio.

Se crea automáticamente dentro de components/services/ la carpeta topicCourse, (ver Figura 5.13).

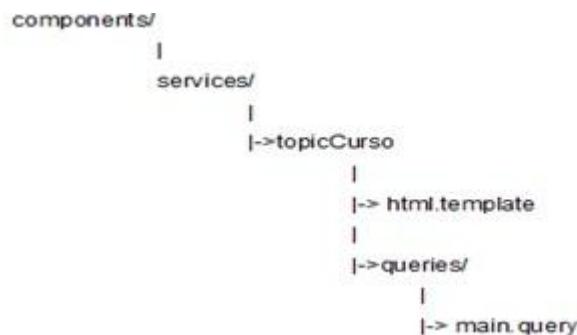


Figura 5.13 Estructura de la creación de un servicio
Elaborado: Diana Molina, Paula Soto F.

Por defecto, al momento de crear un servicio están alojados los siguientes archivos y carpeta:

- **Carpeta Queries:**

Dentro de esta carpeta encontramos un archivo denominado main.query; en este archivo se escribe la consulta SPARQL. Enfatizamos que por la gran cantidad de tripletas del RDF, se debe delimitar las consultas para evitar problemas con el Triplestore. LODSPeaKr permite utilizar argumentos⁸⁹ cuando necesitamos trabajar con parámetros en particular.

- **Archivo html.template**

Por la gran cantidad de tripletas es preciso exponer a los usuarios gráficas amigables e ilustrativas que ayuden a entender las relaciones existentes de la ontología creada, motivo por el cual se utilizó la librería de Java D3.

En el archivo html.template se diseña el estilo de cómo se desea visualizar los datos, para esto se copia el código fuente del diseño del gráfico escogido de la galería D3.

- **Archivo Json.template**

Este archivo es creado y codificado manualmente, contiene la estructura para generar los datos en formato Json de acuerdo a la consulta SPARQL.

5.2.1.2. Aplicaciones creadas con LODSPeaKr.

Finalizado el proceso de creación de servicios, a continuación se muestran las gráficas de los diferentes servicios.

- **Servicio TopicCurso**

Esta gráfica (ver Figura 5.14) detalla todos los cursos que contengan por ejemplo la palabra “MATH” en el Syllabus o en el Área de conocimiento.

El parámetro se lo envía por el navegador (<http://localhost/individuos/topicCurso/MATH>), esta gráfica no está restringida para un argumento en específico; es decir se puede enviar

⁸⁹ Argumento: Variable que puede ser recibida por una rutina o subrutina.

cualquier parámetro como física, literatura, ciencias, etc. Para la gráfica se utiliza el modelo Indented Tree de la galería d3.

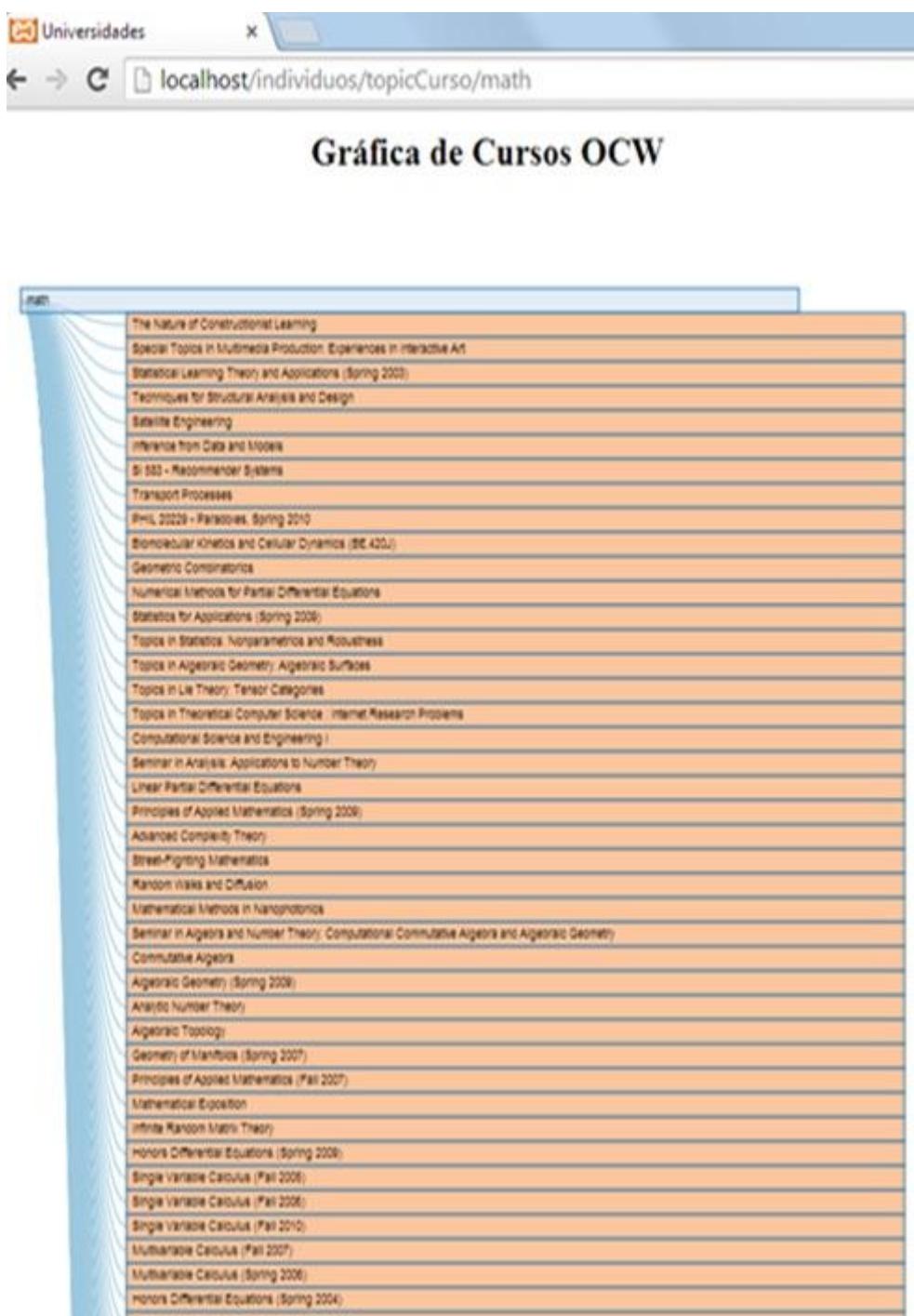


Figura 5.14 Gráfica topicCurso
Elaborado: Diana Molina, Paula Soto F.

- **Servicio Topic**

Esta gráfica (ver Figura 5.15) trabaja de igual manera que la visualización anterior, la diferencia radica en que el parámetro enviado busca en el nombre del curso; se utiliza el modelo Circle Packing de la galería d3.

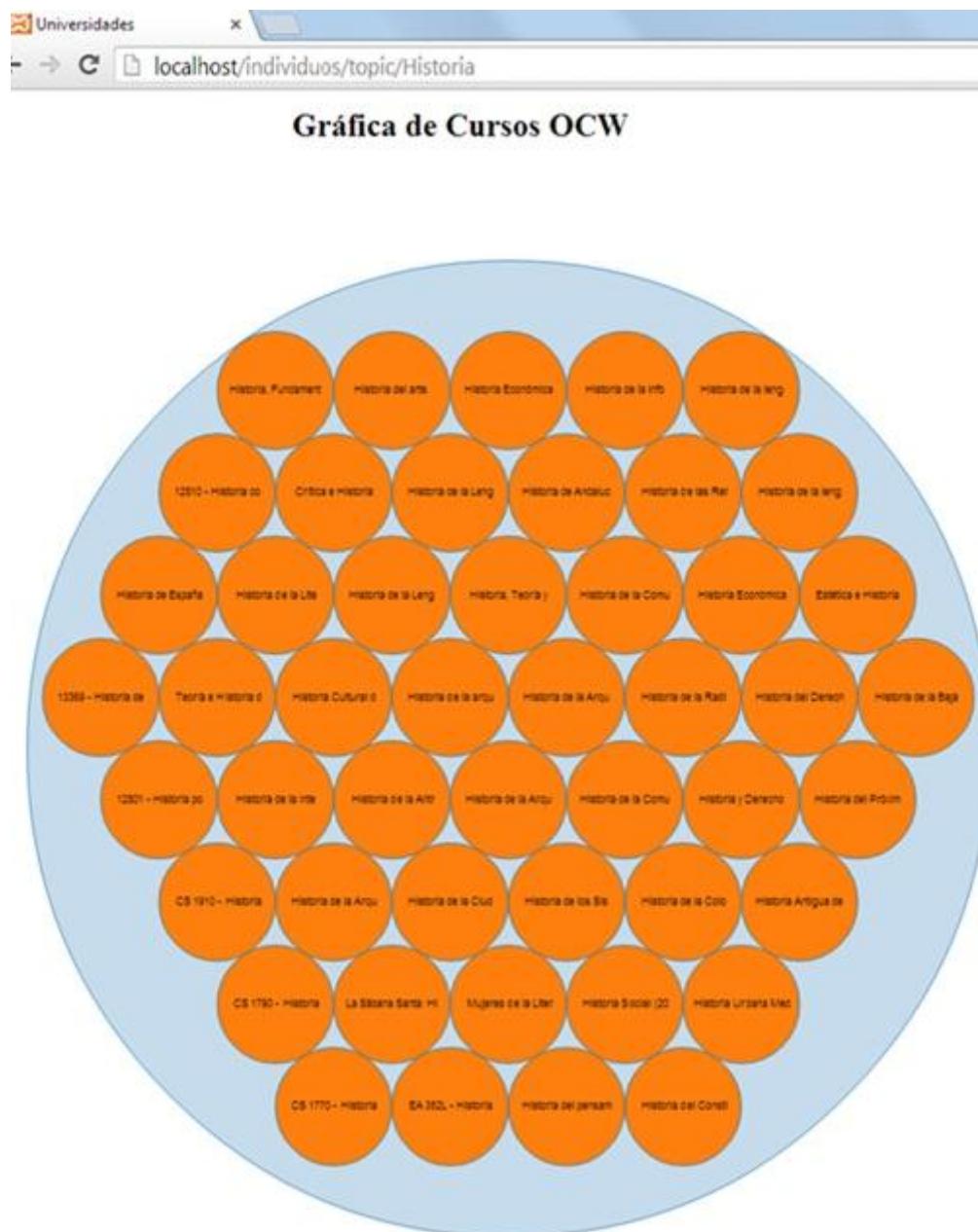


Figura 5.15 Gráfica Topic
Elaborado: Diana Molina, Paula Soto F

- **Servicio Universidad**

Este servicio (ver Figura 5.16) primero presenta una tabla con los nombres de las universidades que coincidan con el parámetro enviado, por ejemplo “Madrid”, luego al escoger la universidad deseada, se mostrará en otra gráfica los data property de la universidad y los cursos que ofrece.

Se utiliza el modelo d3.layout.cluster de la galería d3.

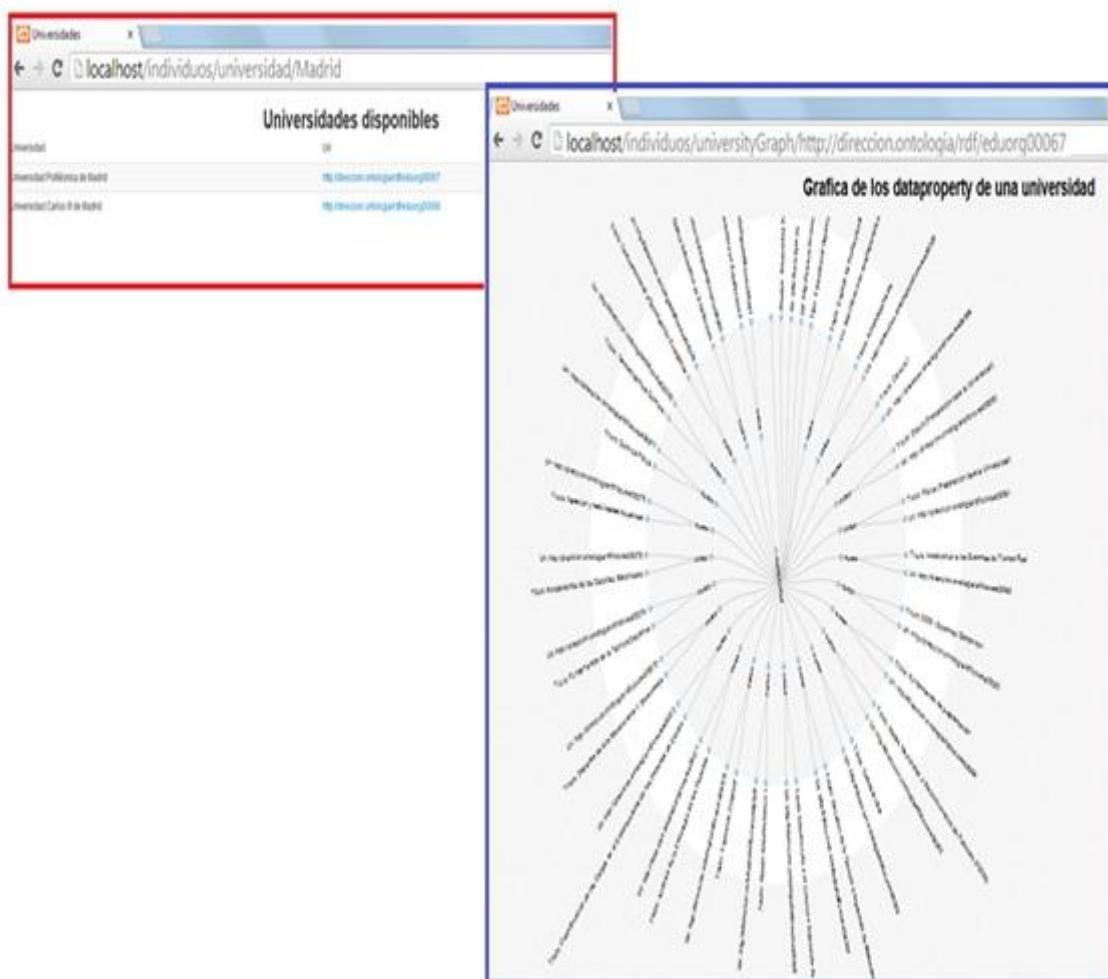


Figura 5.16 Gráfica Universidades
Elaborado: Diana Molina, Paula Soto F

- **Servicio graficaIndividuos**

Servicio (ver Figura 5.17) que muestra el vínculo entre universidades y cursos.

Se utiliza el modelo Collapsible Tree Layout d3.layout.cluster de la galería d3.

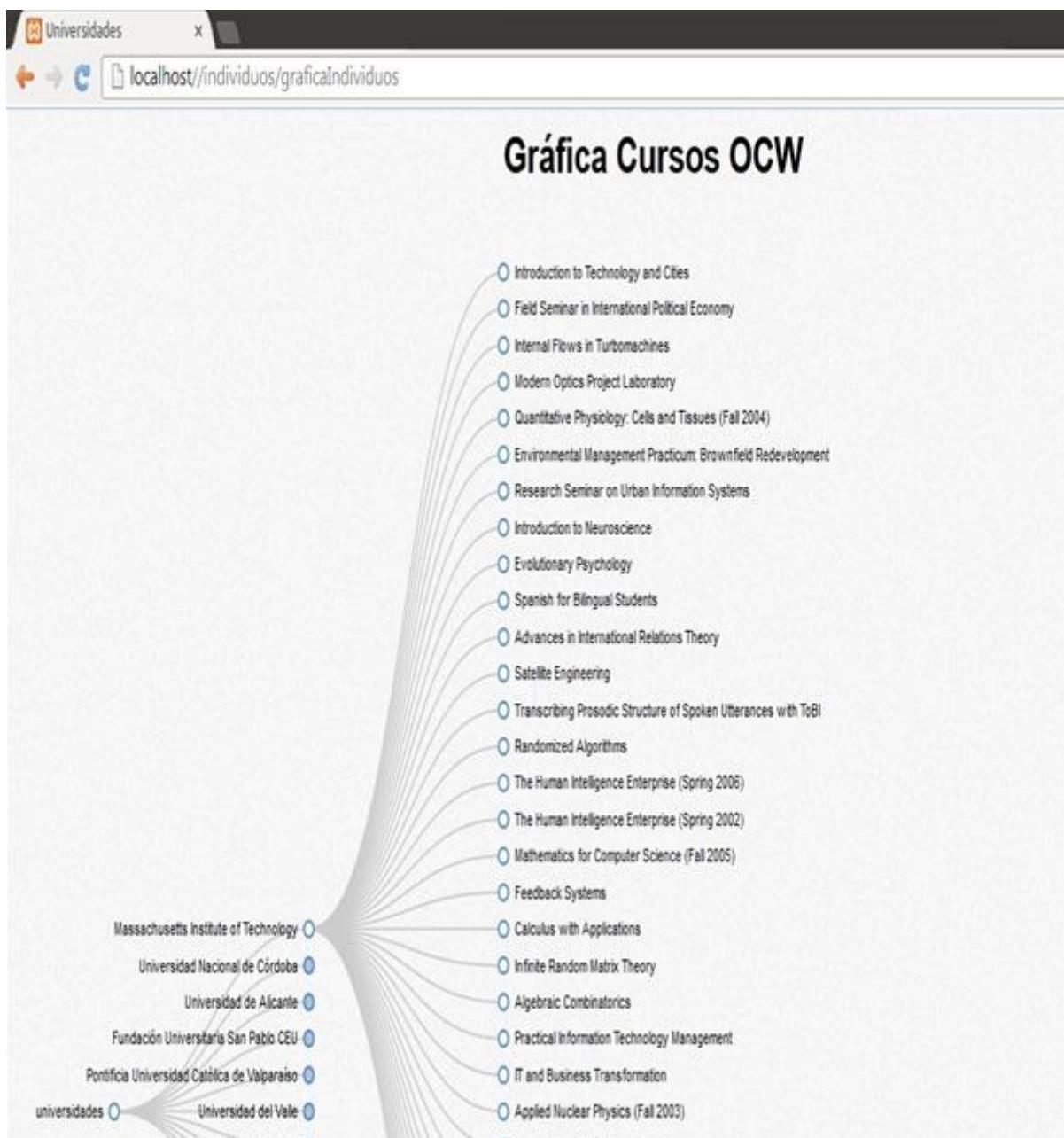


Figura 5.17 Gráfica individuos
Elaborado: Diana Molina, Paula Soto F

- **Servicio Graficador**

Servicio (ver Figura 5.18) que muestra el listado de las 42 universidades, este servicio no hace uso de parámetros, se utiliza el modelo Indented Tree.

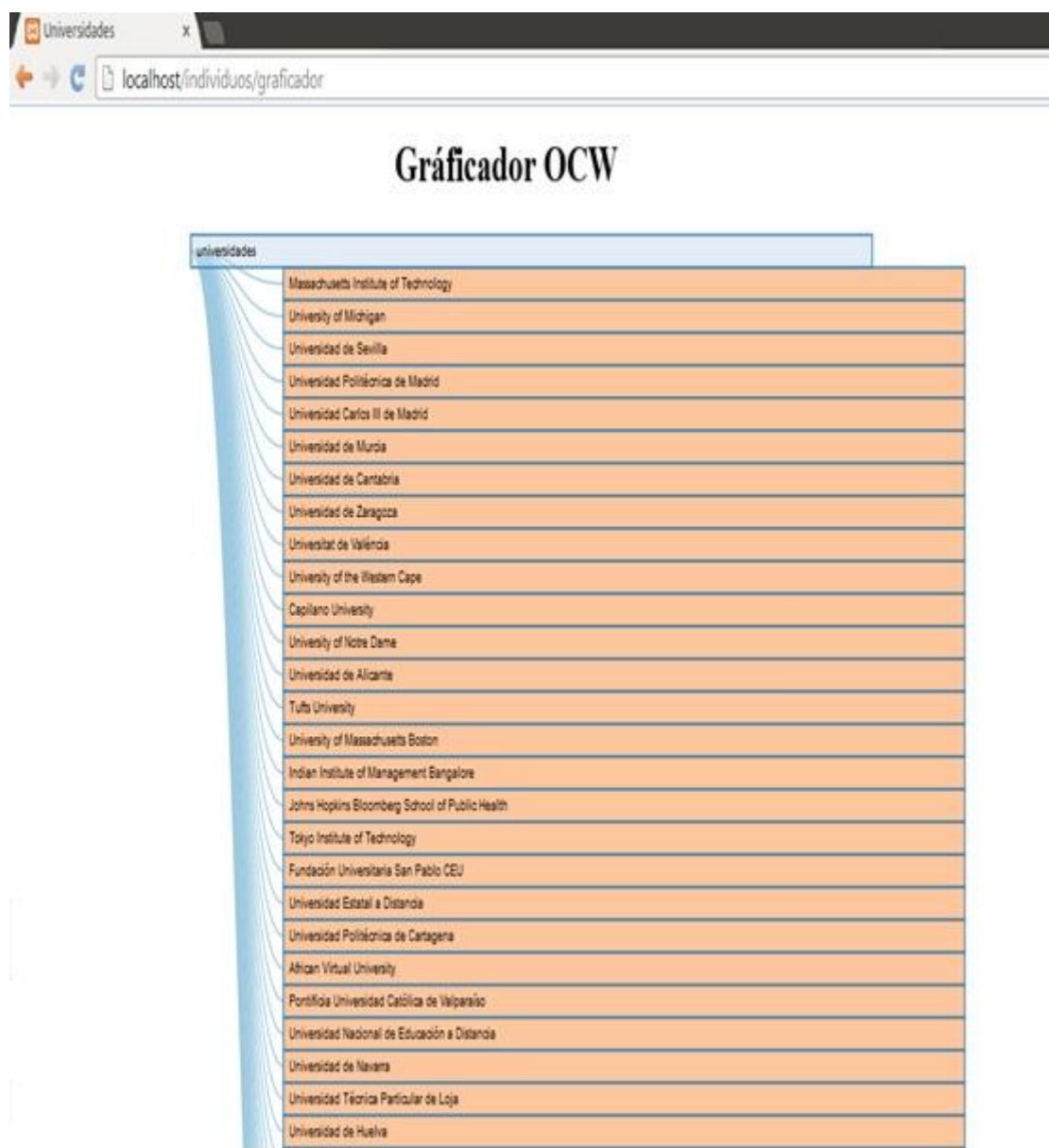


Figura 5.18 Gráfica del servicio Graficador
Elaborado: Diana Molina, Paula Soto F.

- **Servicios detail**

Una de las metas del presente trabajo es maximizar la utilización de un fichero RDF; se crearon 11 servicios de tipo detail, uno por cada clase de la ontología DOCW; cada uno de estos servicios presenta los DataProperty y los ObjectProperty de dicha clase; estos servicios fueron creados para poder ser reutilizados en futuros servicios o proyectos.

A diferencia de los servicios anteriores, los servicios detail muestran los datos en una tabla HTML; la Figura 5.18 describe los DataProperty (color negro) y los ObjectProperty (color azul) de un curso. Los ObjectProperty a su vez nos enlazaran con la información de ese objeto.

Property	Value
type	http://www.w3.org/2002/07/owl#NamedIndividual
type	http://linkedscience.org/teach/mis/Course
url	http://ocw.mit.edu/courses/brain-and-cognitive-sciences/9-520-statistical-learning-theory-and-applications-spring-2003/
KnowledgeArea	Brain and Cognitive Sciences
urlArea	http://ocw.mit.edu/courses/#brain-and-cognitive-sciences
language	English
courseTitle	Statistical Learning Theory and Applications (Spring 2003)
isOfferedBy	http://direccion.ontologia.net/eduorg00054
courseDescription	Focuses on the problem of supervised learning from the perspective of modern statistical learning theory starting with the theory of multivariate function approximation from sparse data. Develops basic tools such as Regularization including Support Vector Machines for regression and classification. Derives generalization bounds using both stability and VC theory. Discusses topics such as boosting and feature selection. Examines applications in several areas: computer vision, computer graphics, text classification and bioinformatics. Final projects and hands-on applications and exercises are planned, paralleling the rapidly increasing practical uses of the techniques described in the subject.
courseDescription	We introduce and motivate the main theme of the course, setting the problem of learning from examples as the problem of approximating a multivariate function from sparse data. We present an overview of the theoretical part of the course and sketch the connection between classical Regularization Theory and its algorithms including Support Vector Machines and Learning Theory, the two cornerstones of the course. We mention theoretical developments during the last few months that provide a new perspective on the foundations of the theory. We briefly describe several different applications ranging from vision to computer graphics, to finance and neuroscience.
courseDescription	Support vector machines have proven to be very useful in classification networks. These SVMs are now being used by drivers for pedestrian avoidance. This is one of the first truly universal applications of this technology. This course is for upper-level graduate students who are planning careers in computational neuroscience. The assignments focus on some of the functions needed to make problemsolving more efficient for computer systems. The project topics students can choose from are based on unsolved problems in the field today. By the conclusion of this course, students should be able to solve one or two of these problems, and should be able to frame an approach to the rest of them.
isRequired	http://direccion.ontologia.net/irrequisito00211

Figura 5.18 Gráfica del servicio detailCurso
Elaborado: Diana Molina, Paula Soto F.

Las gráficas realizadas para la presentación de datos en formato RDF sobre información académica de cursos OCW son de fácil entendimiento para personas que no tienen conocimientos sobre Web semántica, son gráficas que contienen terminología simple y precisa la cual da al usuario la pauta sobre lo que está leyendo u observando.

Al finalizar este capítulo VI denominado acceso y visualización de datos OCW, podemos determinar que con el objetivo de proporcionar acceso a la información generada en el marco del presente proyecto, se han desarrollado cinco servicios para compartir la información con otras aplicaciones, la visualización de datos está orientada para que el usuario final navegue por la información pedagógica de los cursos OCW analizados.

CONCLUSIONES

Basadas en la investigación realizada y el estudio de los componentes que conforman el presente proyecto se listan las siguientes conclusiones.

- Al analizar el proceso de extracción y segmentación de datos de cursos OCW se hace evidente la naturaleza heterogénea de la estructura HTML de los mismos, ya que cada universidad publica sus cursos en diferentes plataformas y organiza de diferente forma sus contenidos.
- Del proceso de categorización se puede exponer que cada institución OCW tiene diferente terminología para nombrar la información académica de los OCWs, dificultando la agrupación, búsqueda o procesamiento de datos, debido a que se realiza un trabajo de análisis de palabras para poder establecer y reunir la información de acuerdo a sus similares en concepto.
- El vocabulario DOCW ayuda a controlar la interoperabilidad de datos ya que se tiene acceso a la información académica OCW en formato máquina, información que puede ser utilizada y aprovechada en diferentes aplicaciones con diferentes fines.
- Los datos anteriores evidencian que el porcentaje de universidades con una estructura académica completa y organizada es muy bajo en relación al total de las 42 universidades investigadas. Sin embargo, alrededor del 80% de universidades exponen al menos el 50% de los componentes del syllabus de un curso.
- La importancia de la presencia de la información pedagógica en un curso puede impactar en su uso por parte de un aprendiz. Es así, que si el usuario no tiene acceso a la información sobre resultados/objetivos de aprendizaje, difícilmente podrá saber en qué competencias el recurso le formará. Lamentablemente menos del 40% de cursos exponen este tipo de información.
- LODSPeakr como marco de creación de aplicaciones basadas en Linked data simplifica el acceso y la publicación de datos RDF, ya que cuenta con un entorno de desarrollo amigable; la visualización gráfica de los servicios creados son de fácil entendimiento para cualquier usuario final que no tenga conocimientos sobre web semántica.

RECOMENDACIONES

Basadas en la investigación realizada y el estudio de los componentes que conforman el presente proyecto se listan las siguientes recomendaciones.

- Crear servicios de actualización periódica que explore los cambios ocurridos en los datos de los repositorios OCW debido a que las organizaciones educativas conscientes del creciente auge de los REA, compiten por publicar datos académico de alto nivel; trabajo que se ve reflejado en la constante actualización de los repositorios OCW.
- En lo relacionado a la extracción de datos se recomienda seguir trabajando en Python ya que es un lenguaje amigable y potente en lo referente a web scraping.
- Se recomienda la reutilización de vocabularios consensuados ya que evita la duplicación de términos para nombrar clases y propiedades.
- En lo relacionado a la creación de nuevos vocabularios se recomienda utilizar términos comunes enmarcados en el ámbito de desarrollo del mismo, de esta manera el vocabulario tendrá acogida y será reutilizado con facilidad.
- Recomendamos la utilización de Virtuoso ya que ofrece un sistema probado de almacenamiento eficiente, esta eficiencia involucra la carga de datos y la fiabilidad en cuanto a la integridad de dichos datos y sus relaciones.
- Se recomienda el uso de LODSPeaKr por su entorno de desarrollo amigable para la creación de aplicaciones basadas en Linked data y publicación de datos RDF.

BIBLIOGRAFÍA

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 4.
- Berners-Lee, T., & Handler, L. (18 de 06 de 2001). The Semantic Web. *Scientific American*.
- BREEN, M. P. (2002). Syllabus design. En R. CARTER, & D. NUNAN, *The Cambridge Guide to Teaching English to Speakers of Other Languages*. (pág. 151). Cambridge University Press.
- Cantera Fonseca, J. M., Hierro Sureda, J. J., & Romo Zabala, P. Á. (2007). *La Web Semántica, la siguiente generación de Webs*. Recuperado el 05 de 2013, de Artículos de la Sociedad de la Información, Fundación Telefónica:
http://telos.fundaciontelefonica.com/DYC/SHI/seccion=1188&idioma=es_ES&id=2009100116310013&activo=4.do?elem=4299
- Carter, R., & Nunan, D. (2002). *The Cambridge Guide to Teaching English to Speakers of Other Languages*. (Eds, Ed.) Cambridge University Press.
- Chicaiza, J. E., Piedra, N., & López, J. (2011). *Propuesta de un mecanismo para conectar e interoperar los planes docentes de EaD de la región usando tecnologías de Web Semántica*. Recuperado el 06 de 2013, de AIESAD 2011:
<http://memorias.utpl.edu.ec/sites/default/files/documentacion/aiasad2011/utpl-aiasad2011-propuesta-de-un-mecanismo.pdf>
- Corcho, O., & Gómez Pérez, A. (2010). *Mini-curso sobre Linked Data*. Obtenido de
<http://www.slideshare.net/ocorcho/linked-data-tutorial-florianopolis>
- Educación Superior Virtual Inclusiva – América Latina. (2013). *E1.1.2 Informe de estado del arte en tecnología de apoyo a la educación superior de personas con discapacidad*. ESVI-AL.
- Ferreiro , R. (29 de Octubre de 2011). *NOVA SouthEastern University*. Recuperado el 23 de 06 de 2013, de http://nsu-curso-moi.blogspot.mx/2011/11/el-syllabus_14.html
- Ferreiro, R. F. (29 de 10 de 2011). *NOVA Southeastern University Fischer School of Education and Human Services*. Recuperado el 06 de 2013, de http://nsu-curso-moi.blogspot.com/2011/11/el-syllabus_14.html
- Galeano Marín, M. E. (2003). Diseño de proyectos en la investigación cualitativa. Medellín: Universidad EAFIT.

García Peñalvo, F. J. (s.f.). *zarza.usual.es*. Recuperado el 10 de Mayo de 2013, de Departamento de Informática y Automática – Facultad de Ciencias Universidad de Salamanca: <http://zarza.usal.es/~fgarcia/doctorado/iuce/WSemantica.pdf>

García Ruiz, R. (2008). *TFC: XML y Web Semántica Estudio del Impacto de las aplicaciones comerciales basadas en tecnologías de la Web Semántica*. Recuperado el 06 de 2013, de <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/1861/1/42897.pdf>

Guzmán, J., Motz, R., & Rodrigue da Siva, A. (2011). *Hacia la publicación abierta de objetos de aprendizaje*. Obtenido de <http://isg.inesc-id.pt/alb/static/papers/2011/C98-CITA2011-jacqueline-v1.2-short-final.pdf>

Lamarca, M. J. (2011). *Metadatos Dublin Core*. Recuperado el 06 de 2013, de Hipertexto.info: http://www.hipertexto.info/documentos/dublin_core.htm

Leguizamo León, A. V. (2007). *The Web Semantic approach like tool for the e-learning*. Recuperado el 06 de 2013, de CEUR Workshop Proceedings: <http://ceur-ws.org/Vol-361/paper12.pdf>

López Guzmán, C., & García Peñalvo, F. (2008). *Los contenidos Educativos y la Web Semántica*. Recuperado el 05 de 2013, de Dirección Nacional de Bibliotecas E-mail Educativo: <http://www.revistas.unal.edu.co/ojs/index.php/email/issue/view/109>

Núñez Paris, F. (2008). *Teoría del currículum y didáctica de las lenguas extranjeras*. Recuperado el 06 de 2013, de Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas : http://www.nebrija.com/revista-linguistica/files/articulosPDF/articulo_531f2c7ea2c17.pdf

Oriol Borrás, G. (Diciembre de 2010). *Observatorio de plataformas para OCW*. Obtenido de <http://ocw.upm.es/documentacion/estudio-utilizacion-de-plataformas-para-opencourseware-2010>

Piedra, N., Tovar, E., Chicaiza, J., & López, J. (08 - 10 de 05 de 2013). *LOCWD: A vocabulary for OpenCourseWarebased on Linked Open Data Technologies*. Recuperado el 06 de 2013, de es.slideshare.net/OCWConsortium: <http://es.slideshare.net/OCWConsortium/ocwc-global-2013-locwd-a-vocabulary-for-ocw-based-on-linked-open-data-technologies>

Santacruz Valencia , L. P., Aedo, I., & Delgado Kloos, C. (2004). *Objetos de aprendizaje: Tendencias dentro de la web semántica*. Recuperado el 05 de 2013, de Rediris.es: <http://www.rediris.es/difusion/publicaciones/boletin/66-67/ponencia18.pdf>

Unesco. (2012). Obtenido de <http://www.unesco.org/new/es/communication-and-information/access-to-knowledge/open-educational-resources/>

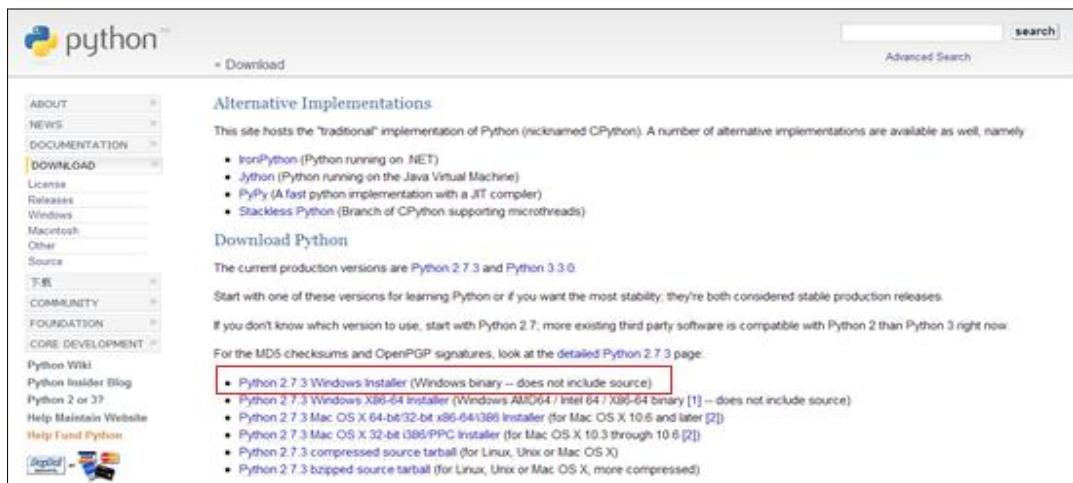
VIRTUAL UNAB. (s.f.). *Creando un Syllabus (Programa de cursos en líneas)*. Recuperado el 06 de 2013, de
https://my.laureate.net/Faculty/docs/Faculty%20Documents/Creando_un_Programa_de_Curso.pdf

ANEXOS

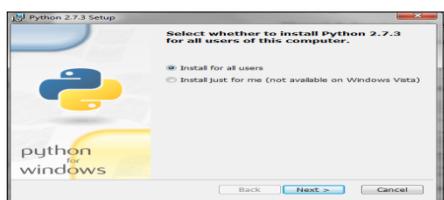
Anexo 1: MANUAL DE INSTALACIÓN DE PYTHON EN WINDOWS

Debemos tener presente cuales son los requerimientos de nuestro ordenador antes de instalar cualquier programa. En nuestro caso trabajaremos con el sistema operativo Windows 7 de 32 bits.

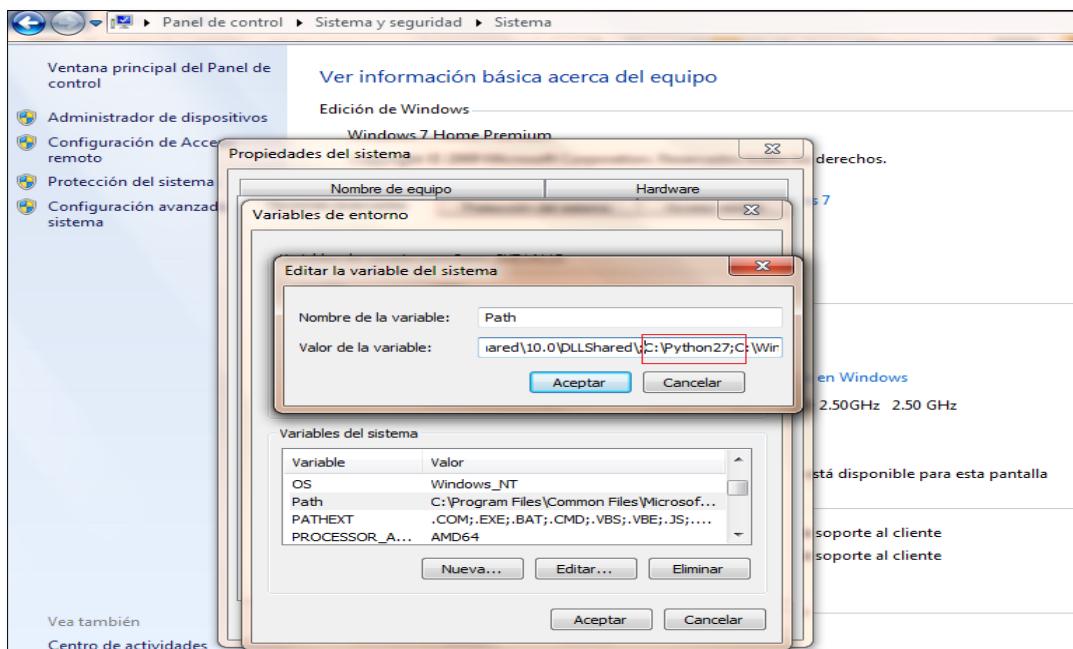
1. Buscamos la versión de [Python](http://python.org/download) que sea compatible con nuestro ordenador, en este caso vamos a descargar Python 2.7 para 32 bits disponible en <http://python.org/download>



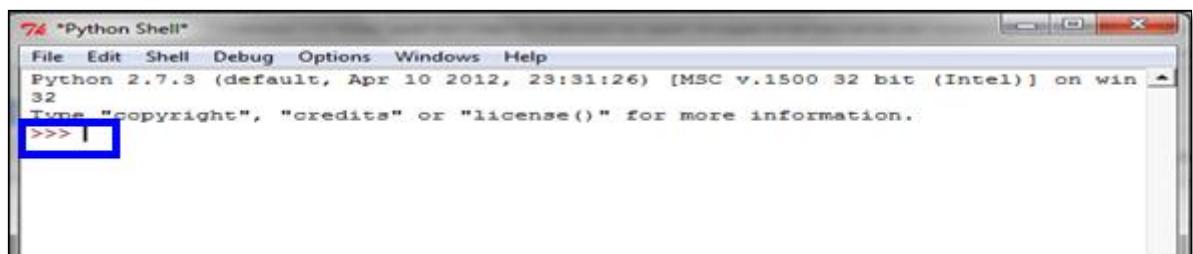
2. Finalizada la descarga, ejecutamos el paquete de Windows installer Python-2.7.3.



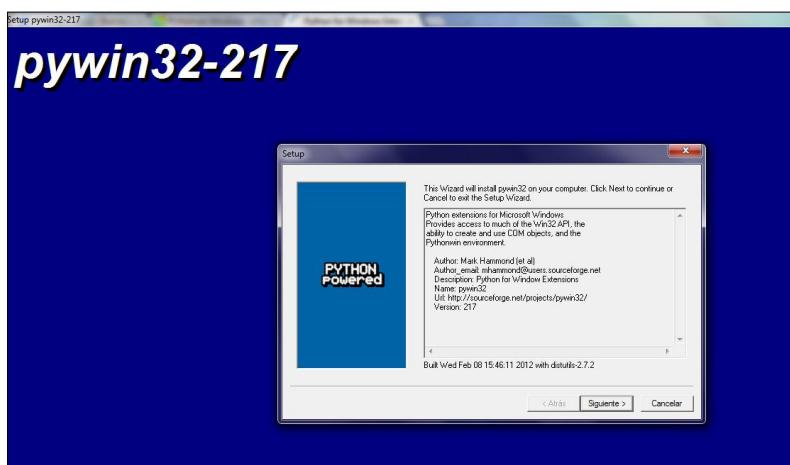
Antes de continuar con el siguiente paso, debemos cambiar los parámetros de la variable de entorno por defecto, para esto accedemos a Inicio->Panel de control/Sistema y seguridad/Sistema/Propiedad del sistema/Variables de entorno/Editar la variable del sistema/Path y añadimos al valor de la variable C:\Python27.



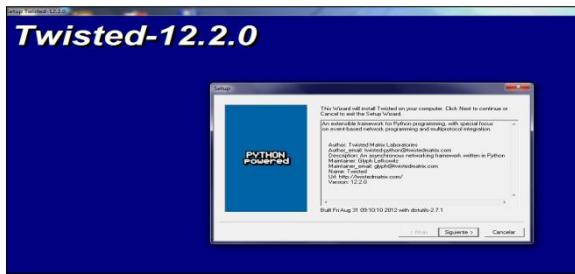
3. El siguiente paso es la ejecución de IDLE de Python que se encuentra en Inicio -> Todos los programas -> Python 2.7 -> IDLE (Python GUI); Al observar la ventana si aparece el prompt **>>>**; significa que está bien realizada la instalación.



4. Procedemos a instalar pywin32 para la versión de Python que instalamos anteriormente, en nuestro caso [pywin32-217](#), disponible en <http://sourceforge.net/projects/pywin32/files/pywin32/Build%20217/>



5. Descargamos [Twisted-12.2.0](#) en la siguiente dirección, esto de acuerdo a nuestros requerimientos y lo ejecutamos.



6. Necesitamos instalar Zope.Interface ya que esta junto con pywin32 son requeridos por Twisted, se procederá de la manera más sencilla mediante la sentencia **easy_install zope.interface**, esta instalación se la realiza directo desde consola bajo C:\.

Inmediatamente editamos la variables de entorno añadiendo C:\Python27;C:\Windows\System;C:\Windows;C:\Windows\System32;C:\Python27\Scripts;C:\Windows\System32;

En algunas ocasiones si se instala Python en Windows no se puede utilizar easy_install porque marca errores; para evitar este inconveniente lo que se tiene que hacer es dirigirse al link <http://blogtaco.blogspot.com/2011/10/usr-easyinstall-de-python-en-windows.html> y guardar el script que se señala en la página en un bloc de notas con el nombre ez_setup.py, guardándolo en el siguiente directorio C:\Python27\ y ejecutarlo.

```
C:\> C:\Windows\system32\cmd.exe
C:\Users\Sony_SUE14115>easy_install zope.interface
Searching for zope.interface
Reading http://pypi.python.org/simple/zope.interface/
Best match: zope.interface 4.0.1
Downloading http://pypi.python.org/packages/2.7/z/zope.interface/zope.interface-4.0.1-py2.7-win32.egg#md5=1e3a463d11aa98e9452bfaff01ab70f2
Processing zope.interface-4.0.1-py2.7-win32.egg
creating c:\python27\lib\site-packages\zope.interface-4.0.1-py2.7-win32.egg
Extracting zope.interface-4.0.1-py2.7-win32.egg to c:\python27\lib\site-packages
Adding zope.interface 4.0.1 to easy-install.pth file
Installed c:\python27\lib\site-packages\zope.interface-4.0.1-py2.7-win32.egg
Processing dependencies for zope.interface
Finished processing dependencies for zope.interface
C:\Users\Sony_SUE14115>
```

Se recomienda instalar `#easy_install -U pip` que es un gestor de descargas de paquetes de Python.

7. Para instalar [libxml2-python-2.7.7](#), se lo puede realizar de dos formas:

a) Buscando el archivo ejecutable en la web <http://xmlsoft.org/sources/win32/python/>, o



b) Se lo puede instalar desde la consola con el gestor de descargas utilizando el siguiente comando: `>easy_install libxml2-python`

8. Instalamos PyOpenSSL, al igual que libxml2 lo podemos descargar manualmente o directamente desde la consola, en este caso se lo realizó desde consola utilizando la línea de comando `>easy_install PyOpenSSL`

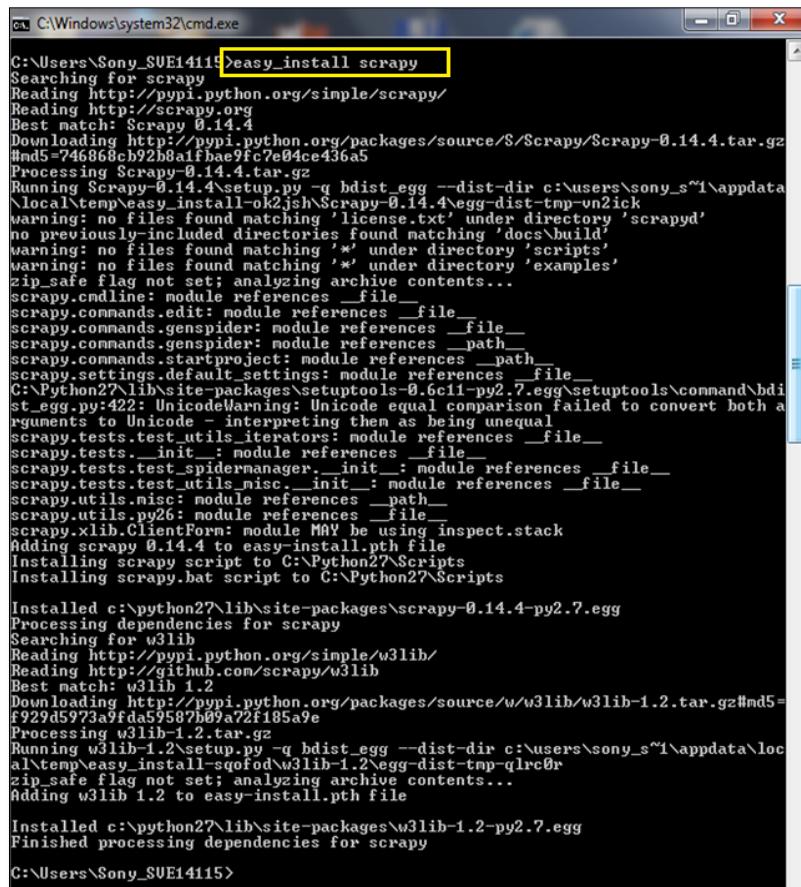
A screenshot of a Windows Command Prompt window titled "C:\Windows\system32\cmd.exe". The window shows the following command and its execution:

```
C:\Users\Sony_SUE14115>easy_install PyOpenSSL
Searching for PyOpenSSL
Reading http://pypi.python.org/simple/PyOpenSSL/
Reading http://pyopenssl.sourceforge.net/
Reading http://launchpad.net/pyopenssl
Best match: pyOpenSSL 0.13
Downloading http://pypi.python.org/packages/2.7/p/pyOpenSSL/pyOpenSSL-0.13-py2.7-win32.egg#md5=e0b7b0d2ccccca83ae38e1c3a105dfe4
Processing pyOpenSSL-0.13-py2.7-win32.egg
creating c:\python27\lib\site-packages\pyOpenSSL-0.13-py2.7-win32.egg
Extracting pyOpenSSL-0.13-py2.7-win32.egg to c:\python27\lib\site-packages
Adding pyOpenSSL 0.13 to easy-install.pth file

Installed c:\python27\lib\site-packages\pyOpenSSL-0.13-py2.7-win32.egg
Processing dependencies for PyOpenSSL
Finished processing dependencies for PyOpenSSL
```

Anexo 2: MANUAL DE INSTALACIÓN DE SCRAPY EN WINDOWS

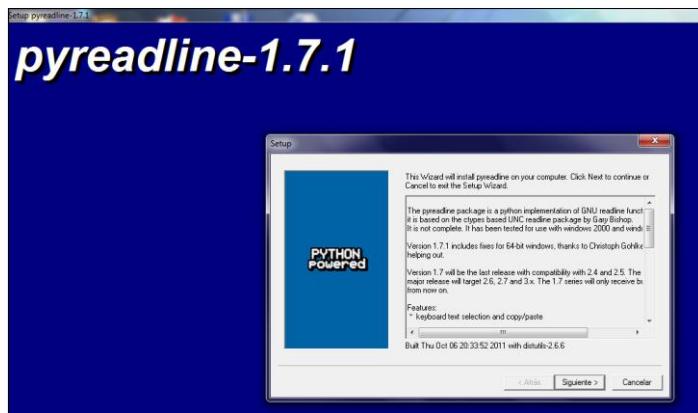
1. Instalamos Scrapy.exe desde la consola usando el comando >easy_install scrapy



```
C:\Windows\system32\cmd.exe
C:\Users\Sony_SUE14115>easy_install scrapy
Searching for scrapy
Reading http://pypi.python.org/simple/scrapy/
Reading http://scrapy.org/
Best match: Scrapy 0.14.4
Downloading http://pypi.python.org/packages/source/S/Scrapy/Scrapy-0.14.4.tar.gz
#md5=746868cb92b8a1fbae9fc7e04ce436a5
Processing Scrapy-0.14.4.tar.gz
Running Scrapy-0.14.4\setup.py -q bdist_egg --dist-dir c:\users\sony_s\appdata\local\temp\easy_install-0k2jsh\Scrapy-0.14.4\egg-dist-tmp-vn2ick
warning: no files found matching 'license.txt' under directory 'scrapyd'
no previously-included directories found matching 'docs\build'
warning: no files found matching '*' under directory 'scripts'
warning: no files found matching '*' under directory 'examples'
zip_safe flag not set; analyzing archive contents...
scrapy.cmdline: module references __file__
scrapy.commands.edit: module references __file__
scrapy.commands.genspider: module references __file__
scrapy.commands.genspider: module references __path__
scrapy.commands.startproject: module references __path__
scrapy.settings.default_settings: module references __file__
C:\Python27\lib\site-packages\setuptools-0.6c11-py2.7.egg\setuptools\command\bdist_egg.py:422: UnicodeWarning: Unicode equal comparison failed to convert both arguments to Unicode - interpreting them as being unequal
scrapy.tests.test_utils_iterators: module references __file__
scrapy.tests._init__: module references __file__
scrapy.tests.test_spidermanager._init__: module references __file__
scrapy.tests.test_utils_misc._init__: module references __file__
scrapy.utils.misc: module references __path__
scrapy.utils.py26: module references __file__
scrapy.xlib.ClientForm: module MAY be using __inspect.stack
Adding scrapy 0.14.4 to easy-install.pth file
Installing scrapy script to C:\Python27\Scripts
Installing scrapy.bat script to C:\Python27\Scripts

Installed c:\python27\lib\site-packages\scrapy-0.14.4-py2.7.egg
Processing dependencies for scrapy
Searching for w3lib
Reading http://pypi.python.org/simple/w3lib/
Reading http://github.com/scrapy/w3lib/
Best match: w3lib 1.2
Downloading http://pypi.python.org/packages/source/w/w3lib/w3lib-1.2.tar.gz#md5=f929d5973a9fda59587h09a72f185a9e
Processing w3lib-1.2.tar.gz
Running w3lib-1.2\setup.py -q bdist_egg --dist-dir c:\users\sony_s\appdata\local\temp\easy_install-sqofod\w3lib-1.2\egg-dist-tmp-qlrc0r
zip_safe flag not set; analyzing archive contents...
Adding w3lib 1.2 to easy-install.pth file
Installed c:\python27\lib\site-packages\w3lib-1.2-py2.7.egg
Finished processing dependencies for scrapy
C:\Users\Sony_SUE14115>
```

2. Debemos instalar Ipython y PyReadline esto con el propósito de mejorar la experiencia interactiva con el intérprete python mejorando las funciones de edición de línea. Siendo la más importante la implementación del tabulador y copiar y pegar. En el siguiente enlace <https://pypi.python.org/pypi/pyreadline/1.7.1> se descarga el archivo pyreadline-1.7.1 y se lo instala manualmente, Ipython se lo descarga dese consola utilizando la sentencia >easy_install ipython.



```
C:\Windows\system32\cmd.exe
C:\Users\Sony_SUE14115>easy_install ipython
Searching for ipython
Reading http://pypi.python.org/simple/ipython/
Reading https://pypi.python.org/pypi/ipython/
Reading https://pypi.python.org/pypi/dist
Reading https://pypi.python.org/pypi/dist/0.8.4
Reading https://pypi.python.org/pypi/dist/0.9.1
Reading https://pypi.python.org/pypi/dist/0.9.2
Reading https://github.com/ipython/ipython.org/release/0.12.1
Reading https://github.com/ipython/ipython.org/dist/old/0.9
Reading https://ipython.scipy.org/dist/0.10
Reading https://ipython.scipy.org/release/0.11/
Reading https://archive.ipython.org/release/0.12
Best match: ipython 0.13
Downloading http://pypi.python.org/packages/2.7/i/ipython/ipython-0.13-py2.7.e
Builds=69ace591616322bd09f17a47e2a61
Processing ipython-0.13-py2.7.egg
creating c:\python27\lib\site-packages\ipython-0.13-py2.7.egg
Extracting ipython-0.13-py2.7.egg to c:\python27\lib\site-packages
Installing ipcontroller.exe script to C:\Python27\Scripts
Installing ipcontroller.exe.manifest script to C:\Python27\Scripts
Installing iptest-script.py script to C:\Python27\Scripts
Installing iptest.exe script to C:\Python27\Scripts
Installing iptest-test-script.py manifest script to C:\Python27\Scripts
Installing ipcluster-script.py script to C:\Python27\Scripts
Installing ipccluster.exe script to C:\Python27\Scripts
Installing ipccluster.exe.manifest script to C:\Python27\Scripts
Installing ipython.exe script to C:\Python27\Scripts
Installing ipython.exe.manifest script to C:\Python27\Scripts
Installing pycolor-script.py script to C:\Python27\Scripts
Installing pycolor.exe manifest script to C:\Python27\Scripts
Installing pycolor.exe.manifest script to C:\Python27\Scripts
Installing iplogger-script.py script to C:\Python27\Scripts
Installing iplogger.exe script to C:\Python27\Scripts
Installing iplogger.exe.manifest script to C:\Python27\Scripts
Installing iprunner-script.py script to C:\Python27\Scripts
Installing iprunner.exe script to C:\Python27\Scripts
Installing ipengine-script.py script to C:\Python27\Scripts
Installing ipengine.exe script to C:\Python27\Scripts
Installing ipengine.exe.manifest script to C:\Python27\Scripts
Installed c:\python27\lib\site-packages\ipython-0.13-py2.7.egg
Processing dependencies for ipython
Finished processing dependencies for ipython
C:\Users\Sony_SUE14115>
```

3. Finalmente procedemos a trabajar en el Shell de Python, (**símbolo del sistema, consola o inicio -> cmd -> C:\>_**)
 - a. Creamos un proyecto por medio del comando **scrapy startproject ejemplo**

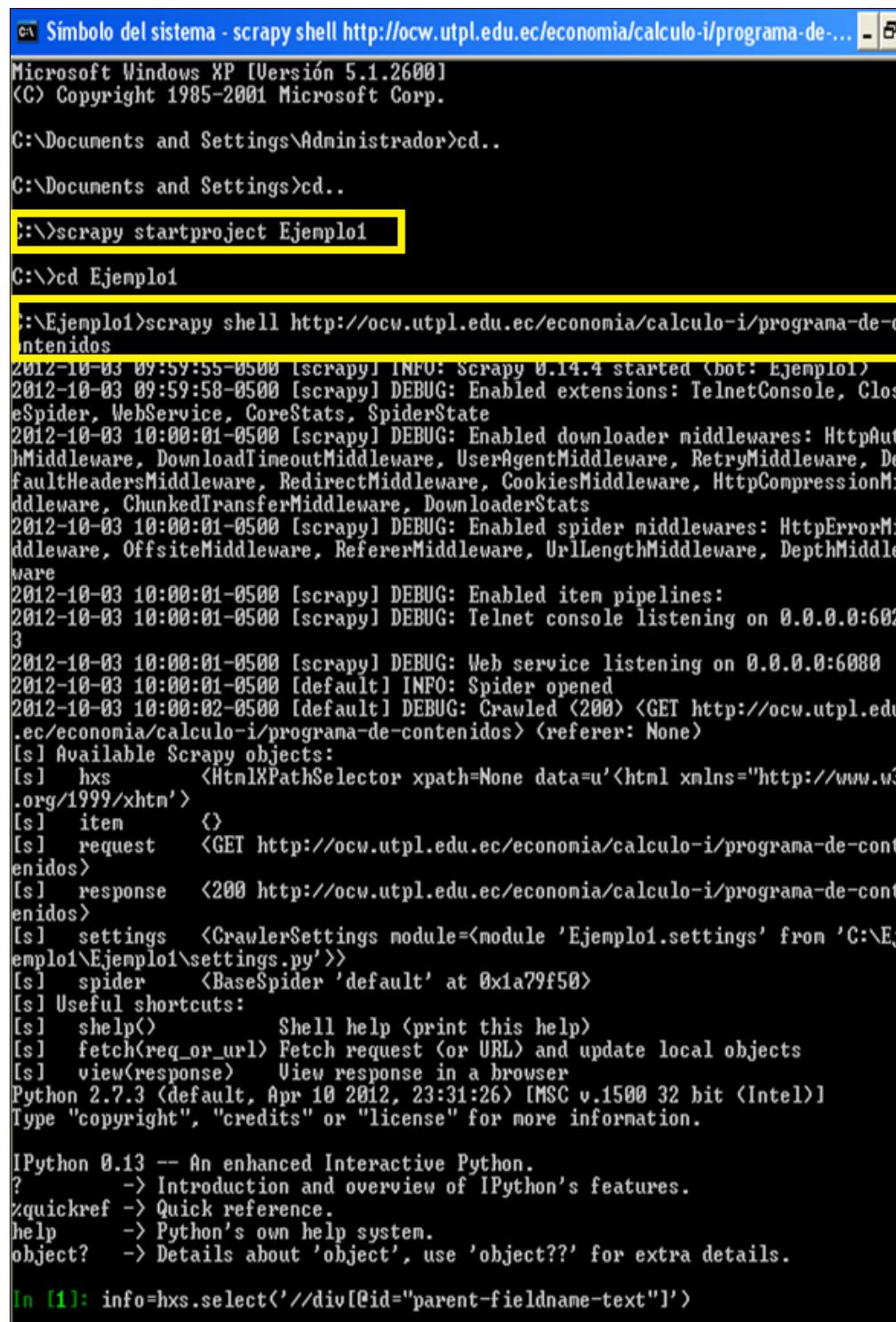
```
C:\Windows\system32\cmd.exe
C:\>scrapy startproject ejemplo
C:\>_
```

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

Ejemplo de scrapy

1. Ingresaremos al proyecto pues desde este nivel podremos ejecutar el Shell, mediante el comando:

```
>scrapy shellhttp://ocw.utpl.edu.ec/economia/calcu.../programa-de-contenidos
```



```
Microsoft Windows XP [Versión 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\Administrador>cd..
C:\Documents and Settings>cd..
C:\>scrapy startproject Ejemplo1
C:\>cd Ejemplo1
C:\Ejemplo1>scrapy shell http://ocw.utpl.edu.ec/economia/calcu.../programa-de-contenidos
2012-10-03 09:59:55-0500 [scrapy] INFO: Scrapy 0.14.4 started (bot: Ejemplo1)
2012-10-03 09:59:58-0500 [scrapy] DEBUG: Enabled extensions: TelnetConsole, Clos
eSpider, WebService, CoreStats, SpiderState
2012-10-03 10:00:01-0500 [scrapy] DEBUG: Enabled downloader middlewares: HttpAut
hMiddleware, DownloadTimeoutMiddleware, UserAgentMiddleware, RetryMiddleware, De
faultHeadersMiddleware, RedirectMiddleware, CookiesMiddleware, HttpCompressionMi
ddleware, ChunkedTransferMiddleware, DownloaderStats
2012-10-03 10:00:01-0500 [scrapy] DEBUG: Enabled spider middlewares: HttpErrorMi
ddleware, OffsiteMiddleware, RefererMiddleware, UrlLengthMiddleware, DepthMiddle
ware
2012-10-03 10:00:01-0500 [scrapy] DEBUG: Enabled item pipelines:
2012-10-03 10:00:01-0500 [scrapy] DEBUG: Telnet console listening on 0.0.0.0:602
3
2012-10-03 10:00:01-0500 [scrapy] DEBUG: Web service listening on 0.0.0.0:6080
2012-10-03 10:00:01-0500 [default] INFO: Spider opened
2012-10-03 10:00:02-0500 [default] DEBUG: Crawled <200> <GET http://ocw.utpl.edu
.ec/economia/calcu.../programa-de-contenidos> <referer: None>
[s] Available Scrapy objects:
[s]   hxs      <HtmlXPathSelector xpath=None data=u'<html xmlns="http://www.w3
.org/1999/xhtml">'>
[s]   item     {}
[s]   request   <GET http://ocw.utpl.edu.ec/economia/calcu.../programa-de-cont
enidos>
[s]   response   <200 http://ocw.utpl.edu.ec/economia/calcu.../programa-de-cont
enidos>
[s]   settings  <CrawlerSettings module=<module 'Ejemplo1.settings' from 'C:\Ej
emplo1\Ejemplo1\settings.py'>>
[s]   spider     <BaseSpider 'default' at 0x1a79f50>
[s] Useful shortcuts:
[s]   shell()      Shell help (print this help)
[s]   fetch(req_or_url) Fetch request (or URL) and update local objects
[s]   view(response) View response in a browser
Python 2.7.3 (default, Apr 10 2012, 23:31:26) [MSC v.1500 32 bit (Intel)]
Type "copyright", "credits" or "license" for more information.

IPython 0.13 -- An enhanced Interactive Python.
?          -> Introduction and overview of IPython's features.
zquickref -> Quick reference.
help       -> Python's own help system.
object?    -> Details about 'object', use 'object??' for extra details.

In [1]: info=hxs.select('//div[@id="parent-fieldname-text"]')
```

```

In [30]: info=hxs.select('//div[@id="parent-fieldname-text"]/blockquote/ul/li')
In [31]: info
Out[31]:
[<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Definici\xf3n</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Notaci\xf3n</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Composici\xf3n\r\nnde funciones</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Gr\xe1fica\r\nnde una funci\xf3n</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Funcione\r\nLineales</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Modelamiento\r\nen Econom\xeda</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Definici\xf3n\r\nnde limites</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Propiedades</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>C\xe1lculo\r\nanal\xedtico</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Limites\r\nunilaterales</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Continuidad\r\nen un punto</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Continuidad\r\nen un intervalo</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>Propiedad\r\ndel valor intermedio</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>As\xedntotas\r\nverticales</li>'],
<HtmlXPathSelector xpath='//div[@id="parent-fieldname-text"]/blockquote/ul/li'
data=u'<li>As\xedntotas\r\nhorizontales</li>']

In [34]: for i in info:
....:     print i.select("text()").extract()[0]
....:
Definición
Notación
Composición
de funciones
Gráfica
de una función
Funcione
Lineales
Modelamiento
en Economía
Definición
de límites
Propiedades
Cálculo
analítico
Límites
unilaterales
Continuidad
en un punto
Continuidad
en un intervalo
Propiedad
del valor intermedio
Asíntotas
verticales
Asíntotas
horizontales

```

Anexo 3: MANUAL DE INSTALACIÓN DE VIRTUALENV

Es recomendable instalar un entorno virtual para trabajar en Python, ya que el mismo ayuda a optimizar recursos de nuestra computadora, como la memoria y disco. Este entorno virtual hereda todas las librerías actuales ya que está vinculado directamente a Python. Al trabajar con esta herramienta podemos tener varios entornos aislados tanto para la misma versión de Python como para otras versiones.

1. Procedemos a la instalación de Virtualenv utilizando el comando >easy_install virtualenv

```
C:\Users\Sony_SUE14115>easy_install virtualenv
Searching for virtualenv
Reading http://pypi.python.org/simple/virtualenv/
Reading http://www.virtualenv.org/
Reading http://virtualenv.openplans.org/
Best match: virtualenv 1.8.2
Downloaded http://pypi.python.org/packages/source/v/virtualenv/virtualenv-1.8.2.tar.gz#md5=174ca075cb1a42c685415692ec4ce2e
Processing virtualenv-1.8.2.tar.gz
Running virtualenv-1.8.2\setup.py -q bdist_egg --dist-dir c:\users\sony_s\1\appdata\local\temp\easy_install-9xmqub\virtualenv-1.8.2\egg-dist-tmp-cekis6
warning: no previously-included files matching '*' found under directory 'docs\_templates'
warning: no previously-included files matching '*' found under directory 'docs\_build'
Adding virtualenv 1.8.2 to easy-install.pth file
Installing virtualenv-script.py script to C:\Python27\Scripts
Installing virtualenv.exe script to C:\Python27\Scripts
Installing virtualenv.exe.manifest script to C:\Python27\Scripts
Installing virtualenv-2.7-script.py script to C:\Python27\Scripts
Installing virtualenv-2.7.exe script to C:\Python27\Scripts
Installing virtualenv-2.7.exe.manifest script to C:\Python27\Scripts
Installed c:\python27\lib\site-packages\virtualenv-1.8.2-py2.7.egg
Processing dependencies for virtualenv
Finished processing dependencies for virtualenv
C:\Users\Sony_SUE14115>
```

2. Designamos un nombre a nuestro entorno en este caso **EntornoCCU** y luego lo activamos. Con el siguiente comando >**virtualenv --no-site-packages EntornoCCU**

```
C:\>virtualenv --no-site-packages EntornoCCU
New python executable in EntornoCCU\Scripts\python.exe
Installing setuptools.....done.
Installing pip.....done.
C:\>_
```

3. Activamos el entorno con : **nombreEntorno\Scripts\activate.bat**

En nuestro caso el nombre de la carpeta es:

EntornoCCU\Scripts\activate.bat

```
C:\>virtualenv --no-site-packages EntornoCCU
New python executable in EntornoCCU\Scripts\python.exe
Installing setuptools.....done.
Installing pip.....done.

C:\>EntornoCCU\Scripts\activate.bat
<EntornoCCU> C:\>deactivate.bat
C:\>
```

Para desactivar el entorno, se utiliza el siguiente comando >**deactivate.bat**, como lo mostramos en la Figura anterior.

4. Procederemos a trabajar en el Shell del entorno virtual, primero lo activamos y luego creamos un proyecto con el nombre **Scrapy CCU** por medio del comando: **/scrapy startproject ScrapyCCU**

```
C:\>EntornoCCU\Scripts\activate.bat
<EntornoCCU> C:\>scrapy startproject ScrapyCCU
```

Se procede a probar la funcionalidad de scrapy dentro del entorno de la misma manera que se lo realiza en el entorno de Python, dirigirse al numeral 3 del proceso de instalación de Python.

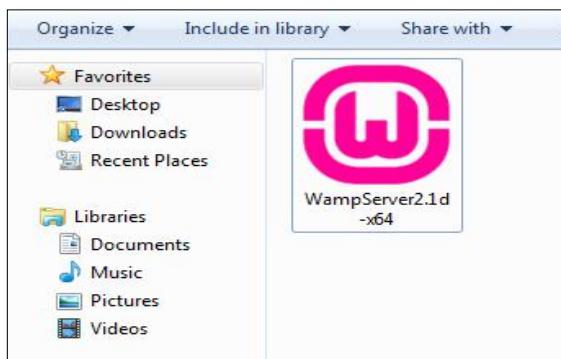
Anexo 4: MANUAL DE INSTALACIÓN DE WAMPSERVER 2.1

Este tutorial nos ayudara a instalar WampServer 2.1 en un equipo con Windows 7.

1. Nos dirigimos al sitio oficial de Wampserver <http://www.wampserver.com/>, se encontrará dos opciones de descarga de WampServer para Windows 7.
 - [Descarga WampServer 2.1d \(64 bits\)](#)
 - [Descargar WampServer 2.1e \(32 bits\)](#)



2. Cuando la descarga de WampServer 2.1 está completa, buscamos el ícono WampServer en nuestra carpeta de Descargas



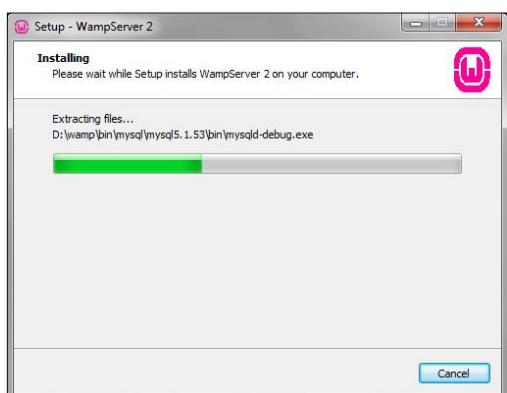
3. Es el momento de instalar WampServer. Se recibirá una advertencia de seguridad después de abrir el archivo WampServer.exe.



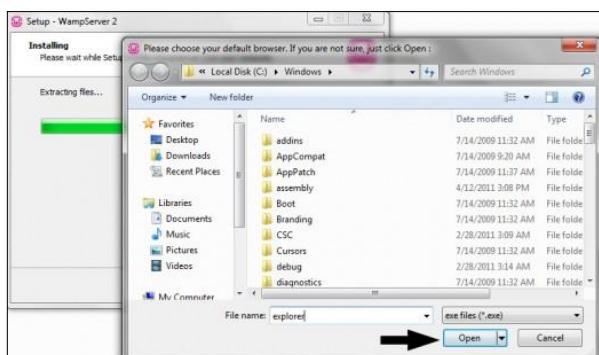
- Aparecerá un asistente de instalación estándar de Windows después de hacer clic en Ejecutar botón del cuadro de diálogo de advertencia de seguridad.



- Seguir las instrucciones del asistente hasta que se empiece a instalar WampServer en el ordenador.



- Aparecerá un cuadro de diálogo para elegir el navegador predeterminado para WampServer. Puede elegir su navegador favorito o dejar el navegador de Internet Explorer por defecto.

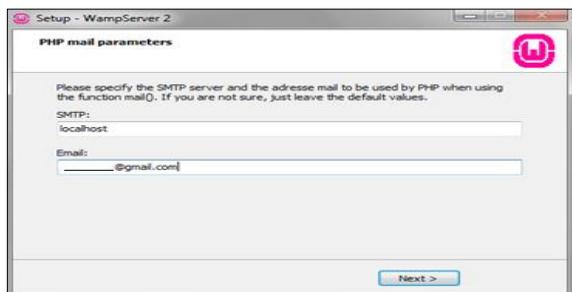


- Emergerá un "Firewall de Windows" de diálogo estándar, para la configuración de Apache por WampServer. (Usted no puede observar esto, si el servidor de seguridad de Windows no está activo). Haga clic en "Permitir acceso" para dejar las opciones por defecto.



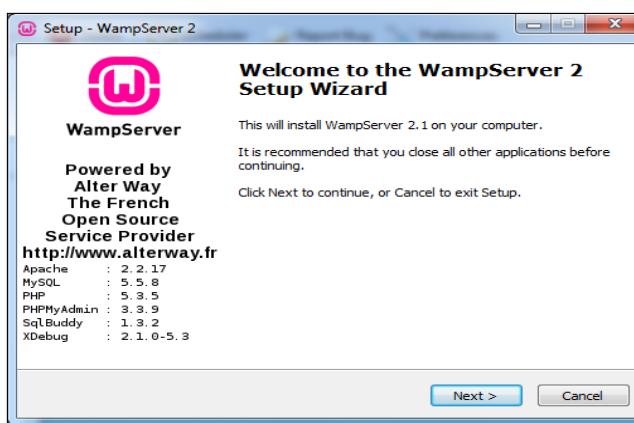
8. Una vez permitido el acceso al servidor Apache, se puede especificar el servidor SMTP y la dirección de correo electrónico para ser utilizado por PHP cuando se utiliza la función de correo electrónico. Se recomienda los siguientes parámetros,
 - SMTP: localhost
 - Email: Tu dirección de correo electrónico.

Hacer clic en "Siguiente" para el diálogo de la instalación final.



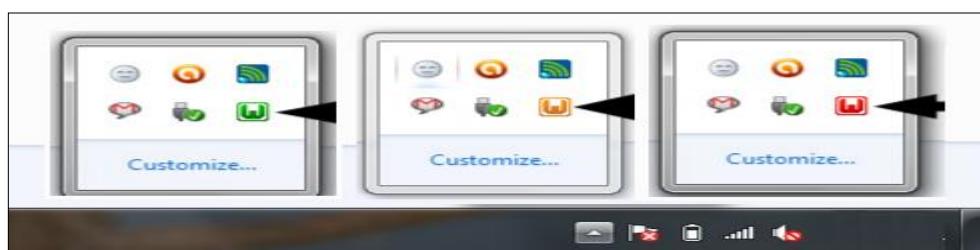
9. Se ha instalado WampServer 2.1 junto con Apache, MySQL, PHP, phpMyAdmin en nuestro ordenador.

Hacer clic en "Finalizar" para iniciar WampServer junto con otros servicios.



10. Ahora podemos ubicar el icono de WampServer en la barra de tareas. El icono de WampServer tiene tres estados.

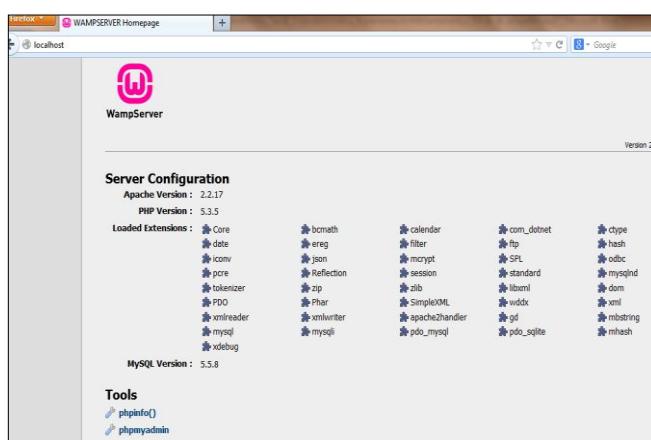
- Rojo.- El icono de color rojo muestra que WampServer está en línea y que el servicio no se ha iniciado todavía, es decir Apache, PHP y MySQL.
- Amarillo. El icono de color amarillo muestra que se han detenido los servicios de WampServer.
- Verde.- Este color muestra que Wampserver está funcionando normal, junto con todos los servicios y el servidor de desarrollo está listo para su uso.



11. Es el momento de probar WampServer con los servicios instalados de Apache, PHP, MySQL y phpMyAdmin. A continuación los pasos para probar WampServer en Windows 7.

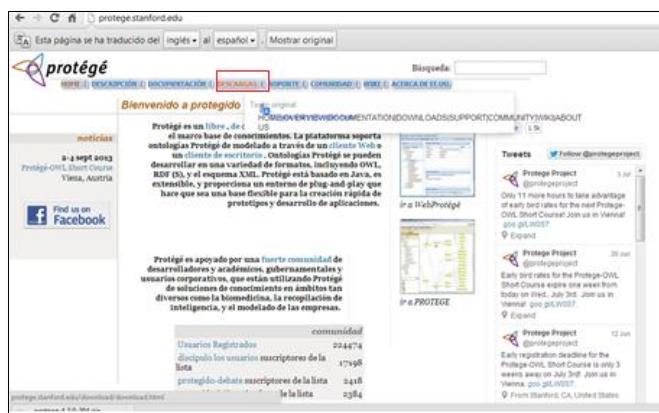
- Asegúrese, que el color de icono de WampServer está en verde en la barra de tareas
- Abrir el navegador predeterminado en la instalación e introducir la dirección <http://localhost/>
- Pulsar Enter

Si se ha instalado correctamente WampServer en el computador, se visualizará la pantalla de bienvenida de WampServer con la respectiva configuración.

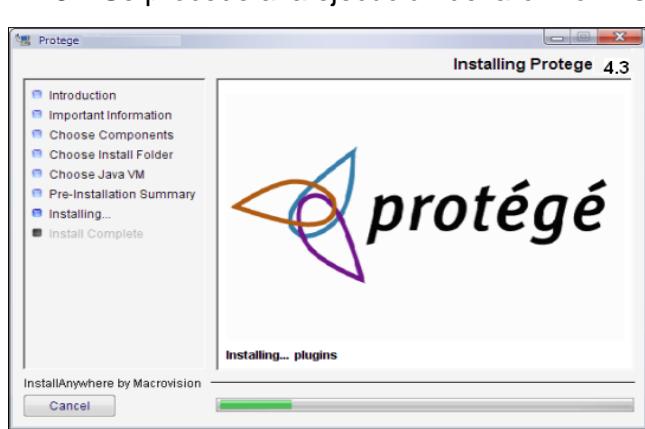


Anexo 5: MANUAL DE INSTALACIÓN DE PROTEGÉ

1. Procedemos a descargar Protégé de forma gratuita desde la dirección <http://protege.stanford.edu/>. Se descarga la versión 4.3 procederemos a la instalación, para la cual solo es necesario que tengamos una máquina virtual de Java instalada en nuestro sistema versión 1.5 o superior, de lo contrario deberemos bajarnos el instalable que la incluye.



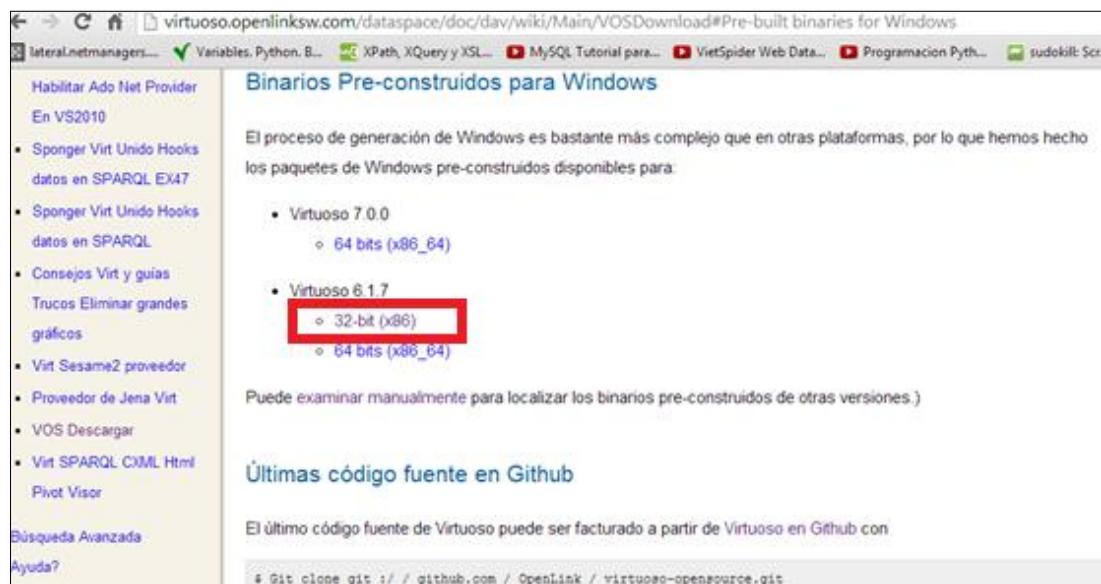
2. Antes de proceder con la instalación es necesario que tengamos una máquina virtual de Java instalada en nuestro sistema versión 1.5 o superior, de lo contrario deberemos bajarnos el instalable, direccionarse al siguiente link <http://java.com/en/download/chrome.jsp>
3. Se procede a la ejecución del archivo **install_protege_4.3.exe**:



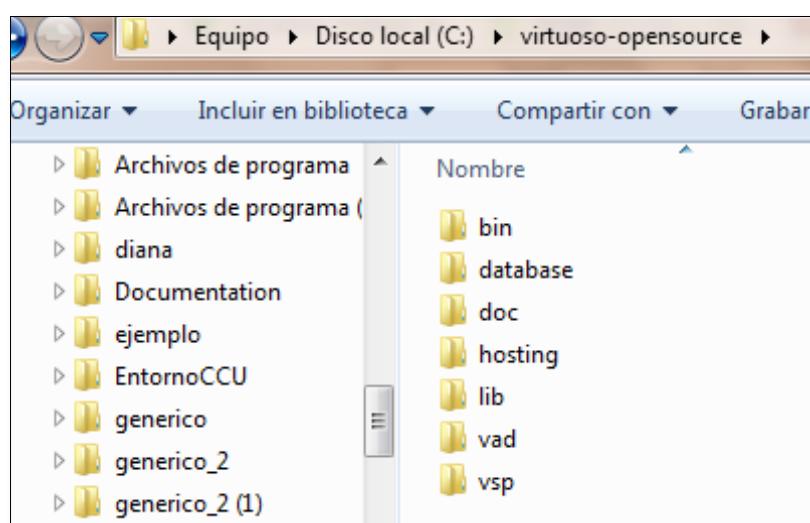
Anexo 6: INSTALACION DE VIRTUOSO EN LINUX

Descargamos virtuoso 6.17 para 32 bits de la siguiente dirección:

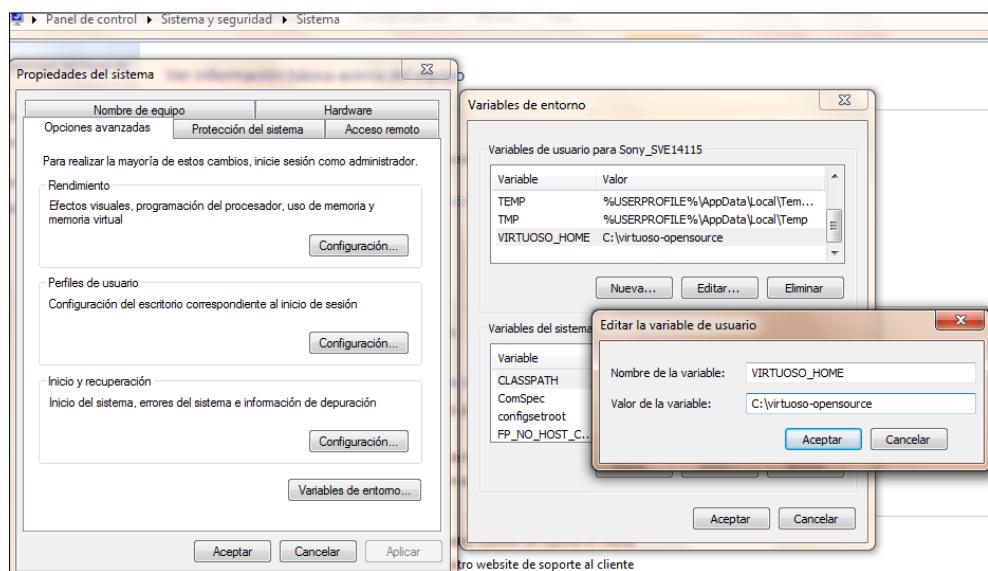
<http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSDownload#Pre-built binaries for Windows>



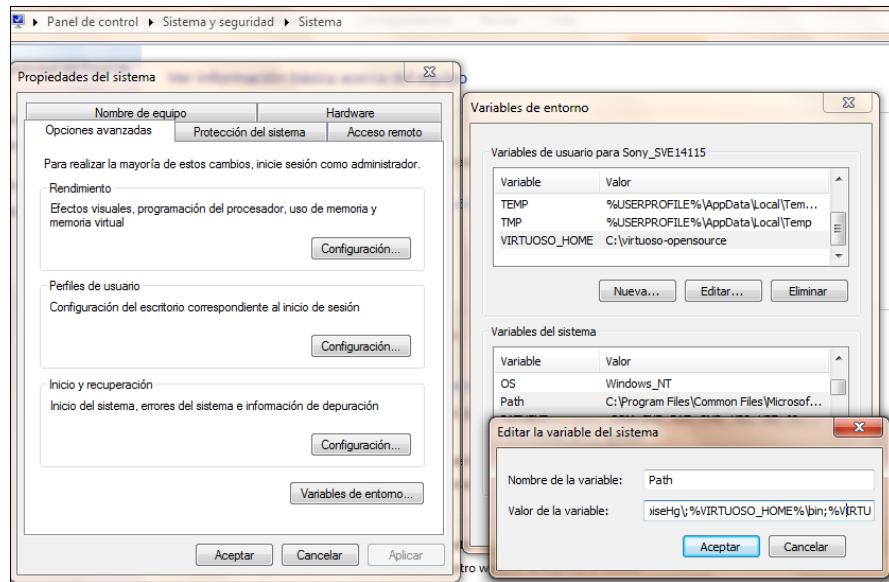
1. Nos arroja una archivo .zip, se lo descomprime en la raíz de C: y creará un directorio denominado “virtuoso-opensource”, el mismo que contiene 7 subcarpetas: bin, database, doc, hosting, lib, vad y vsp, como se muestra en la siguiente imagen.



- Para el correcto funcionamiento del servicio de la base de datos predeterminada se debe crear una variable de entorno denominada VIRTUOSO_HOME, con el valor de la variable C:\virtuoso-opensource.



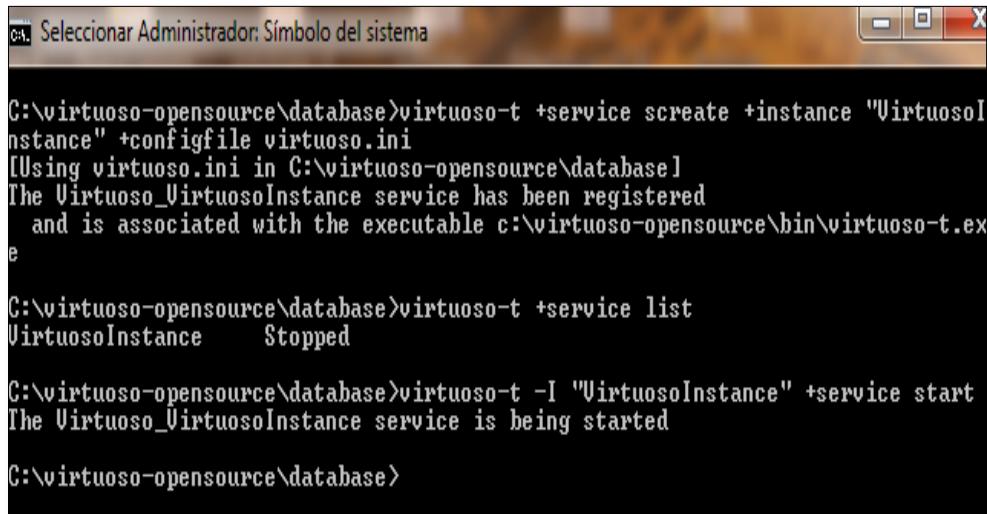
- Así mismo modificamos nuestro path añadiendo C:\virtuoso-opensource\bin;C:\virtuoso-opensource\lib.



- Vamos al símbolo del sistema, y ejecutamos los siguientes comandos para verificar que el servicio que ofrece virtuoso funciona adecuadamente.

```
C:\virtuoso-opensource\database>virtuoso-t +service screate +instance
```

```
C:\virtuoso.opensource\database>virtuoso-t +service list
C:\virtuoso.opensource\database>virtuoso-t -I "VirtuosoInstance" +service start
C:\virtuoso.opensource\database>
```



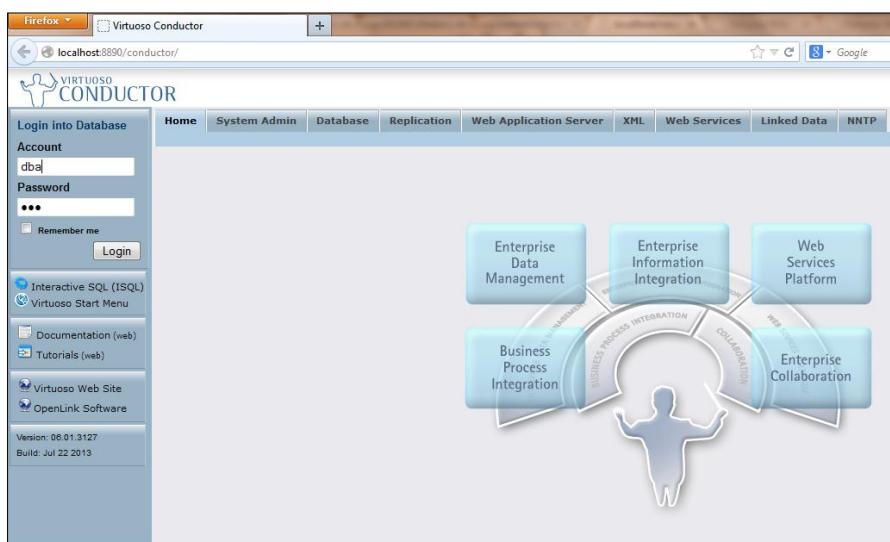
```
C:\virtuoso.opensource\database>virtuoso-t +service list
VirtuosoInstance      Stopped

C:\virtuoso.opensource\database>virtuoso-t -I "VirtuosoInstance" +service start
The Virtuoso_VirtuosoInstance service is being started

C:\virtuoso.opensource\database>
```

- Vamos al conductor de virtuoso, por defecto ejecuta el puerto HTTP "8890" y accedemos a la url <http://localhost:8890/conductor>.

Nos pide nombre de usuario predeterminado para el administrador y la contraseña, por defecto son “dba”, “dba”, para mayor seguridad es recomendable cambiar estos parámetros.



Anexo 7: CÓDIGO FUENTE DEL SERVICIO TOPICCOURSE

- **MAIN.QUERY**

```
SELECT DISTINCT ?curso ?tituloCurso WHERE
{ ?curso
<http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#KnowledgeArea> ?areaConocimiento ;
    <http://linkedscience.org/teach/ns/courseTitle> ?tituloCurso ;
    rdf:type <http://linkedscience.org/teach/ns/Course> ;
    <http://direccion.ontologia/rdf/consistingOf> ?syllabus .
?syllabus
<http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#Activity> ?activity .
FILTER ( regex(str(?activity), "{{lodspk.args.arg0}}", "i") || regex(str(?areaConocimiento), "{{lodspk.args.arg0}}", "i") )
LIMIT 100
```

- **HTML.TEMPLATE**

```
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8" />
<title>Universidades</title>
<script type="text/javascript" src="{{lodspk.home}}js/jquery.js"></script>
    <script src="{{lodspk.home}}js/d3.v3.min.js"></script>
<style type="text/css">
    .node rect {
        cursor: pointer;
        fill: #fff;
        fill-opacity: .5;
        stroke: #3182bd;
        stroke-width: 1.5px;
    }

    .node text {
        font: 10px sans-serif;
        pointer-events: none;
    }

    path.link {
        fill: none;
        stroke: #9ecae1;
        stroke-width: 1.5px;
    }
</style>
</head>
<body>
<h1><center>Gráfica de Cursos OCW</center></h1>
```

```

<div id="content"></div>

<script type="text/javascript">

var root, tree, margin, i = 0,
    duration = 400, svg, diagonal;
jQuery(function($){
    $.ajax({
        url : '/individuos/topicCurso.json/{{lodspk.args.arg0}}',
        async : false,
        cache :false,
        type : 'get',
        method : 'get',
        dataType : 'html'
    }).error(function(data){
        //console.log(data);
    }).done(function(data){
        var cont = JSON.stringify(data.toString());
        var dt = data.replace(/\n|\t|r/g, "");
        var objParse = $.parseJSON(dt);
        //-----Inicia proceso de agrupar los datos y eliminar los
repetidos-----
        /*
        ** Nodo principal
        */
//-----Inicia a graficar los resultados-----

margin = {top: 30, right: 20, bottom: 30, left: 20},
width = 960 - margin.left - margin.right,
barHeight = 20,
barWidth = width * .8;

tree = d3.layout.tree()
.size([0, 100]);

diagonal = d3.svg.diagonal()
.projection(function(d) { return [d.y, d.x]; });

svg = d3.select("#content").append("svg")
.attr("width", width + margin.left + margin.right)
.append("g")
.attr("transform", "translate(" + margin.left + "," + margin.top + ")");

objParse.x0 = 0;
objParse.y0 = 0;
update(root = objParse);

});
});

function update(source) {

```

```

// Compute the flattened node list. TODO use d3.layout.hierarchy.
var nodes = tree.nodes(root);

var height = Math.max(500, nodes.length * barHeight + margin.top + margin.bottom);

d3.select("svg")
  .attr("height", height);

d3.select(self.frameElement)
  .style("height", height + "px");

// Compute the "layout".
nodes.forEach(function(n, i) {
  n.x = i * barHeight;
});

// Update the nodes...
var node = svg.selectAll("g.node")
  .data(nodes, function(d) { return d.id || (d.id = ++i); });

var nodeEnter = node.enter().append("g")
  .attr("class", "node")
  .attr("transform", function(d) { return "translate(" + source.y0 + "," + source.x0 + ")"; })
  .style("opacity", 1e-6);

// Enter any new nodes at the parent's previous position.
nodeEnter.append("rect")
  .attr("y", -barHeight / 2)
  .attr("height", barHeight)
  .attr("width", barWidth)
  .style("fill", color)
  .on("click", click);

nodeEnter.append("text")
  .attr("dy", 3.5)
  .attr("dx", 5.5)
  .text(function(d) { return d.name; });

// Transition nodes to their new position.
nodeEnter.transition()
  .duration(duration)
  .attr("transform", function(d) { return "translate(" + d.y + "," + d.x + ")"; })
  .style("opacity", 1);

node.transition()
  .duration(duration)
  .attr("transform", function(d) { return "translate(" + d.y + "," + d.x + ")"; })
  .style("opacity", 1)
  .select("rect")
  .style("fill", color);

// Transition exiting nodes to the parent's new position.
node.exit().transition()
  .duration(duration)

```

```

    .attr("transform", function(d) { return "translate(" + source.y + "," + source.x + ")"; })
    .style("opacity", 1e-6)
    .remove();

// Update the links...
var link = svg.selectAll("path.link")
    .data(tree.links(nodes), function(d) { return d.target.id; });

// Enter any new links at the parent's previous position.
link.enter().insert("path", "g")
    .attr("class", "link")
    .attr("d", function(d) {
        var o = {x: source.x0, y: source.y0};
        return diagonal({source: o, target: o});
    })
    .transition()
    .duration(duration)
    .attr("d", diagonal);

// Transition links to their new position.
link.transition()
    .duration(duration)
    .attr("d", diagonal);

// Transition exiting nodes to the parent's new position.
link.exit().transition()
    .duration(duration)
    .attr("d", function(d) {
        var o = {x: source.x, y: source.y};
        return diagonal({source: o, target: o});
    })
    .remove();

// Stash the old positions for transition.
nodes.forEach(function(d) {
    d.x0 = d.x;
    d.y0 = d.y;
});
}

// Toggle children on click.
function click(d) {
    detailCourse(d.url);
    /*
    if (d.children) {
        d._children = d.children;
        d.children = null;
    } else {
        d.children = d._children;
        d._children = null;
    }
    update(d);
    */
}
}

```

```

        function detailCourse(uri){
            window.location.href = '/individuos/detailCourse/'+uri;
        }

function color(d) {
    return d._children ? "#3182bd" : d.children ? "#c6dbef" : "#fd8d3c";
}

        </script>
</body>
</html>

```

- **JSON.TEMPLATE**

```
{
  "name" : "{{lodspk.args.arg0}}",
  "children": [
    {% for row in models.main %}
    {
      "name": "{{row.tituloCurso.value}}", "size" : 3000, "url": "{{row.curso.value}}"
    }{%if forloop.first%}, {%else%} {%endif%} {%if !forloop.last%},{%endif%}

    {% endfor %}
    {
      "name": "", "size" : 0, "url": ""
    }
  ]
}
```

Anexo 8: CÓDIGO FUENTE DEL SERVICIO DETAILCOURSE

- **MAIN.QUERY**

```
SELECT DISTINCT ?p ?o
WHERE {
  ?s a <http://linkedscience.org/teach/ns/Course>;
  ?p ?o .
  filter (str(?s) = "http://{{lodspk.args.all|uripath}}")
}
```

- **HTML.TEMPLATE**

```
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8" />
<title>Detalle Curso
</title>
<link href="{{lodspk.home}}css/bootstrap.min.css" rel="stylesheet" type="text/css"
media="screen" />
<link href="{{lodspk.home}}css/bootstrap-responsive.min.css" rel="stylesheet" type="text/css"
media="screen" />
<script type="text/javascript" src="{{lodspk.home}}js/jquery.js"></script>
<script type="text/javascript" src="{{lodspk.home}}js/bootstrap.min.js"></script>
<script type="text/javascript" src="{{lodspk.home}}js/bootstrap-typeahead.js"></script>
<script src="{{lodspk.home}}js/jquery-1.10.2.min.js"></script>
<script src="{{lodspk.home}}js/sammy.js"></script>
</head>
<style>
.nolink{
  pointer-events: none;
  cursor: default;
  color: #333333;
  text-decoration : none;
}
</style>
<script type="text/javascript">
jQuery(function($){
  $('.container-fluid a').each(function(index, d){
    if($(d).attr("href").indexOf("resor") == -1 &&
$(d).attr("href").indexOf("syllb0") == -1
      && $(d).attr("href").indexOf("teach0") == -1 &&
$(d).attr("href").indexOf("requis0") == -1
      && $(d).attr("href").indexOf("docum0") == -1 &&
$(d).attr("href").indexOf("activ0") == -1
      && $(d).attr("href").indexOf("eduorg0") == -1){
        $(d).removeAttr("class");
        $(d).attr("class", "nolink");
      }
  });
  $('.link').click(function(e){
    var bandera = false;
```

```

var url = "";
var text = $(this).text();
if(text.indexOf("resor") != -1 ){
    url = '/individuos/detailResource/'+$(this).text();
    bandera = true;
}
else{
    if(text.indexOf("syllib0") != -1){
        url = '/individuos/detailSyllabus/'+$(this).text();
        bandera = true;
    }
    else{
        if(text.indexOf("teach0") != -1){
            url =
'/individuos/detailTeach/'+$(this).text();
            bandera = true;
        }
        else{
            if(text.indexOf("requis0") != -1){
                url =
'/individuos/detailRequirement/'+$(this).text();
                bandera = true;
            }
            else{
                if(text.indexOf("docum0") != -1){
                    url =
'/individuos/detailDocument/'+$(this).text();
                    bandera = true;
                }
                else{
                    if(text.indexOf("activ0") != -1){
                        url =
'/individuos/detailActivity/'+$(this).text();
                        bandera = true;
                    }
                    else{
                        if(text.indexOf("eduorg0") != -1){
                            url =
'/individuos/detailUniversity/'+$(this).text();
                            bandera
= true;
                        }
                    }
                }
            }
        }
    }
}
if(bandera){
    window.location.href = url;
}

```

```

        /*
        jQuery.ajax({
        url : url,
        async : false,
        cache :false,
        type : 'get',
        method : 'get',
        dataType : 'html'
        }).error(function(data){
        }).done(function(data){
                $("#content").html(data);
        });
        */
    }

    return false;
});

</script>
<body>
<div class="container-fluid">
<h1><center>Detalle de Curso</center></h1>
    <table class="table table-striped">
        <thead>
            <tr>
                <td align="center">
                    Property
                </td>
                <td align="center">
                    Value
                </td>
            </tr>
        </thead>
        <tbody>
            {% for row in models.main %}
                <tr>
                    <td>
                        {{row.p.localname|explode:"/"|pop|explode:"#|pop}}
                    </td>
                    <td>
                        <a class="link" href='{{row.o.value}}'>{{row.o.value}}</a>
                    </td>
                </tr>
            {% endfor %}
        </tbody>
        </table>
    </div>
    <a href="javascript:history.back(1)">Volver Atrás</a>
</body>
</html>

```

Anexo 9: CÓDIGO FUENTE DEL SERVICIO DETAILSyllabus

- **MAIN.QUERY**

```
SELECT DISTINCT ?p ?o
WHERE {
  ?s a
  <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#Syllabus>;
  ?p ?o .
  filter (str(?s) = "http://{{lodspk.args.all|uripath}}")
}
```

- **HTML.TEMPLATE**

```
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8" />
<title>Detalle Syllabus
</title>
<link href="{{lodspk.home}}css/bootstrap.min.css" rel="stylesheet" type="text/css" media="screen" />
<link href="{{lodspk.home}}css/bootstrap-responsive.min.css" rel="stylesheet" type="text/css" media="screen" />
<script type="text/javascript" src="{{lodspk.home}}js/jquery.js"></script>
<script type="text/javascript" src="{{lodspk.home}}js/bootstrap.min.js"></script>
<script type="text/javascript" src="{{lodspk.home}}js/bootstrap-typeahead.js"></script>
<script src="{{lodspk.home}}js/jquery-1.10.2.min.js"></script>
</head>
<style>
.nolink{
  pointer-events: none;
  cursor: default;
  color: #333333;
  text-decoration : none;
}
</style>
<script type="text/javascript">
jQuery(function($){
  $('.container-fluid a').each(function(index, d){
    if(($(d).attr("href").indexOf("lerot") == -1 &&
      $(d).attr("href").indexOf("acadc0") == -1 && $(d).attr("href").indexOf("acadc1") == -1 &&
      $(d).attr("href").indexOf("acadc2") == -1 && $(d).attr("href").indexOf("acadc3") == -1 &&
      $(d).attr("href").indexOf("subj0") == -1 ) ){
      $(d).removeAttr("class");
      $(d).attr("class", "nolink");
    }
  })
})
```

```

    });
    $('.link').click(function(e){
        var bandera = false;
        var url = "";
        var text = $(this).text();
        if(text.indexOf("lerot0") != -1){
            url = '/individuos/detailOutcomes/'+$(this).text();
            bandera = true;
        }
        else {
            if(text.indexOf("acadc0") != -1 || text.indexOf("acadc1") != -1 || text.indexOf("acadc2") != -1 || text.indexOf("acadc3") != -1){
                url =
                    '/individuos/detailAcademicCalendar/'+$(this).text();
                bandera = true;
            }
            else{
                if(text.indexOf("subj0") != -1){
                    url =
                        '/individuos/detailSubject/'+$(this).text();
                    bandera = true;
                }
            }
        }

        if(bandera){
            window.location.href = url;
            /*
            jQuery.ajax({
                url : url,
                async : false,
                cache :false,
                type : 'get',
                method : 'get',
                dataType : 'html'
            }).error(function(data){
            }).done(function(data){
                $("#content").html(data);
            });
            */
        }
    });

    return false;
});

});
</script>
<body>
<div class="container-fluid">
<h1><center>Detalle de Syllabus</center></h1>
    <table class="table table-striped">
        <thead>
            <tr>
                <td align="center">

```

```

Property
</td>
<td align="center">
Value
</td>
</tr>
</thead>
<tbody>
{% for row in models.main %}
<tr>
<td>
{{row.p.localname|explode:"/"|pop|explode:"#|pop}}
</td>
<td>
<a class="link" href='{{row.o.value}}'>{{row.o.value}}</a>
</td>
</tr>
{% endfor %}
</tbody>
</table>
</div>
<a href="javascript:history.back(1)">Volver Atrás</a>
</body>
</html>

```

Anexo 10: CONSULTAS SPARQL

- **Componentes OCW por total de Universidades**

Consulta que nos devuelve el valor de universidades que poseen el componente Syllabus.

Texto de la consulta
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> select count(distinct ?univer) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso. ?curso docw:consistingOf ?syllabus }</pre>

Consulta que nos devuelve el valor de universidades que poseen el componente Academic Calendar.

Texto de la consulta
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> select count(distinct ?univer) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso. ?curso docw:consistingOf ?syllabus. ?syllabus docw:timeline ?acadCal }</pre>

Consulta que nos devuelve el valor de universidades que poseen el componente Subject.

Texto de la consulta
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> select count(distinct ?univer) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso. ?curso docw:consistingOf ?syllabus. ?syllabus docw:evaluationCriteria ?Subj }</pre>

Consulta que nos devuelve el valor de universidades que poseen el componente Learning Outcomes

Texto de la consulta

```
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>

select count(distinct ?univer)
where
{
    ?univer a locwd:EducationalOrganization;
            docw:offer ?curso.
    ?curso docw:consistingOf ?syllabus.
    ?syllabus docw:outcomes ?objetivos
}
```

Consulta que nos devuelve el valor de universidades que poseen el componente Bibliography.

Texto de la consulta

```
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>

select count(distinct ?univer)
where
{
    ?univer a locwd:EducationalOrganization;
            docw:offer ?curso.
    ?curso docw:hasResource ?biblio
}
```

Consulta que nos devuelve el valor de universidades que poseen el componente Requisite.

Texto de la consulta

```
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>

select count(distinct ?univer)
where
{
    ?univer a locwd:EducationalOrganization;
            docw:offer ?curso.
    ?curso docw:isRequired ?requerim
}
```

- Componentes OCW por total de cursos

Consulta que nos devuelve el número de cursos que poseen Syllabus.

Texto de la consulta
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> prefix teach: <http://linkedscience.org/teach/ns/> select count(distinct ?curso) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso. ?curso docw:consistingOf ?syllabus; teach:courseTitle ?titulo }</pre>

Consulta que nos devuelve el número de cursos que poseen Academic calendar

Texto de la consulta
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> prefix teach: <http://linkedscience.org/teach/ns/> select count(distinct ?curso) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso. ?curso docw:consistingOf ?syllabus; teach:courseTitle ?titulo. ?syllabus docw:timeline ?acadCal }</pre>

Consulta que nos devuelve el número de cursos que poseen Subject.

Query Text
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> prefix teach: <http://linkedscience.org/teach/ns/> select count(distinct ?curso) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso. ?curso docw:consistingOf ?syllabus; teach:courseTitle ?titulo. ?syllabus docw:evaluationCriteria ?Subj }</pre>

Consulta que nos devuelve el número de cursos que poseen Learning Outcomes

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix teach: <http://linkedscience.org/teach/ns/>

select count(distinct ?curso)
where
{
    ?univer a locwd:EducationalOrganization;
              docw:offer ?curso.
    ?curso docw:consistingOf ?syllabus;
              teach:courseTitle ?titulo.
    ?syllabus docw:outcomes ?objetivos
}
```

Consulta que nos devuelve el número de cursos que poseen Bibliography

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix teach: <http://linkedscience.org/teach/ns/>

select count(distinct ?curso)
where
{
    ?univer a locwd:EducationalOrganization;
              docw:offer ?curso.
    ?curso docw:hasResource ?biblio;
              teach:courseTitle ?titulo.
}
```

Consulta que nos devuelve el número de cursos que poseen Requisite.

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix teach: <http://linkedscience.org/teach/ns/>

select count(distinct ?curso)
where
{
    ?univer a locwd:EducationalOrganization;
              docw:offer ?curso.
    ?curso docw:isRequired ?requerim;
              teach:courseTitle ?titulo.
}
```

- **Componentes de cursos OCW por cada universidad.**

Consultas que nos devuelve la cantidad de Syllabus que tiene cada universidad.

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombrUniv AS ?UNIVERSIDAD)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
             docw:offer ?curso;
             foaf:name ?nombrUniv.
    ?curso docw:consistingOf ?syllabus;
            ipd:KnowledgeArea ?areaCon
}
GROUP BY ?nombrUniv
```

Consultas que nos devuelve la cantidad de Academic Calendar que tiene cada universidad.

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombrUniv AS ?UNIVERSIDAD)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
             docw:offer ?curso;
             foaf:name ?nombrUniv.
    ?curso docw:consistingOf ?syllabus;
            ipd:KnowledgeArea ?areaCon.
    ?syllabus docw:timeline ?acadCal
}
GROUP BY ?nombrUniv
```

Consultas que nos devuelve la cantidad de Subject que tiene cada universidad.

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombrUniv AS ?UNIVERSIDAD)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
             docw:offer ?curso;
             foaf:name ?nombrUniv.
    ?curso docw:consistingOf ?syllabus;
            ipd:KnowledgeArea ?areaCon.
    ?syllabus docw:evaluationCriteria ?Subj
}
GROUP BY ?nombrUniv
```

Consultas que nos devuelve la cantidad de Learning Outcomes que tiene cada universidad.

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombUniv AS ?UNIVERSIDAD)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
             docw:offer ?curso;
             foaf:name ?nombUniv.
    ?curso docw:consistingOf ?syllabus;
             ipd:KnowledgeArea ?areaCon.
    ?syllabus docw:outcomes ?objetivos
}
GROUP BY ?nombUniv
```

Consultas que nos devuelve la cantidad de Bibliography que tiene cada universidad.

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombUniv AS ?UNIVERSIDAD)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
             docw:offer ?curso;
             foaf:name ?nombUniv.
    ?curso ipd:KnowledgeArea ?areaCon;
             docw:hasResource ?biblio.
}
GROUP BY ?nombUniv
```

Consultas que nos devuelve la cantidad de Requisite que tiene cada universidad.

```
Query Text
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombUniv AS ?UNIVERSIDAD)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
             docw:offer ?curso;
             foaf:name ?nombUniv.
    ?curso ipd:KnowledgeArea ?areaCon;
             docw:isRequired ?requerim
}
GROUP BY ?nombUniv
```

- Número de cursos por Área de conocimiento de cursos OCW.

Syllabus

Query Text
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> prefix foaf: <http://xmlns.com/foaf/0.1/#> prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#> select (?nombUniv AS ?UNIVERSIDAD) (?areaCon AS ?AREA_CONOCIMIENTO) (count(distinct ?curso) AS ?CANTIDAD_CURSOS) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso; foaf:name ?nombUniv. ?curso docw:consistingOf ?syllabus; ipd:KnowledgeArea ?areaCon } GROUP BY ?nombUniv ?areaCon</pre>

Academic Calendar

Query Text
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> prefix foaf: <http://xmlns.com/foaf/0.1/#> prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#> select (?nombUniv AS ?UNIVERSIDAD) (?areaCon AS ?AREA CONOCIMIENTO) (count(distinct ?curso) AS ?CANTIDAD_CURSOS) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso; foaf:name ?nombUniv. ?curso docw:consistingOf ?syllabus; ipd:KnowledgeArea ?areaCon. ?syllabus docw:timeline ?acadCal }</pre>

Subject

Query Text
<pre>prefix docw: <http://direccion.ontologia/rdf/> prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/> prefix foaf: <http://xmlns.com/foaf/0.1/#> prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#> select (?nombUniv AS ?UNIVERSIDAD) (?areaCon AS ?AREA_CONOCIMIENTO) (count(distinct ?curso) AS ?CANTIDAD_CURSOS) where { ?univer a locwd:EducationalOrganization; docw:offer ?curso; foaf:name ?nombUniv. ?curso docw:consistingOf ?syllabus; ipd:KnowledgeArea ?areaCon. ?syllabus docw:evaluationCriteria ?Subj } GROUP BY ?nombUniv ?areaCon</pre>

Learning Outcomes

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombUniv AS ?UNIVERSIDAD)
       (?areaCon AS ?AREA_CONOCIMIENTO)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
            docw:offer ?curso;
            foaf:name ?nombUniv.
    ?curso docw:consistingOf ?syllabus;
            ipd:KnowledgeArea ?areaCon.
    ?syllabus docw:outcomes ?objetivos
}
GROUP BY ?nombUniv ?areaCon
```

Bibliography

```
Query Text
prefix docw: <http://direccion.ontologia/rdf/>
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombUniv AS ?UNIVERSIDAD)
       (?areaCon AS ?AREA_CONOCIMIENTO)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
            docw:offer ?curso;
            foaf:name ?nombUniv.
    ?curso ipd:KnowledgeArea ?areaCon;
            docw:hasResource ?biblio.
}
GROUP BY ?nombUniv ?areaCon
```

Requisite

```
Query Text
prefix locwd: <http://lod.utpl.edu.ec/locwd/schema/>
prefix foaf: <http://xmlns.com/foaf/0.1/#>
prefix ipd: <http://www.semanticweb.org/ontologies/2013/0/OntologyPlanesDocentesEducacionSuperior#>

select (?nombUniv AS ?UNIVERSIDAD)
       (?areaCon AS ?AREA_CONOCIMIENTO)
       (count(distinct ?curso) AS ?CANTIDAD_CURSOS)
where
{
    ?univer a locwd:EducationalOrganization;
            docw:offer ?curso;
            foaf:name ?nombUniv.
    ?curso ipd:KnowledgeArea ?areaCon;
            docw:isRequired ?requerim
}
GROUP BY ?nombUniv ?areaCon
```