



**UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA**

*La Universidad Católica de Loja*

**ÁREA TÉCNICA**

TÍTULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y  
COMPUTACIÓN

**Desarrollo de un modelo de representación común para Repositorios  
de Recursos Educativos Abiertos**

PROYECTO DE TITULACION

**AUTOR:** Contreras Mendieta Jairo Antonio

**DIRECTOR:** Sucunuta España Manuel Eduardo Ing.

LOJA-ECUADOR

2017



*Esta versión digital, ha sido acreditada bajo la licencia Creative Commons 4.0, CC BY-NY-SA: Reconocimiento-No comercial-Compartir igual; la cual permite copiar, distribuir y comunicar públicamente la obra, mientras se reconozca la autoría original, no se utilice con fines comerciales y se permiten obras derivadas, siempre que mantenga la misma licencia al ser divulgada. <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>*

*Septiembre, 2017*

## **APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN**

Ingeniero.

Sucunuta España Manuel Eduardo

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de titulación Desarrollo de un modelo de representación común para Repositorios de Recursos Educativos Abiertos, realizado por: Jairo Antonio Contreras Mendieta, ha sido orientado y revisado durante su ejecución, para que se aprueba la presentación del mismo.

Loja, Marzo de 2017

F.....

Ing. Manuel Eduardo Sucunuta España

## DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo Jairo Antonio Contreras Mendieta declaro ser autor (a) del presente trabajo de titulación: Desarrollo de un modelo de representación común para Repositorios de Recursos Educativos Abiertos, de la Titulación en Sistemas Informáticos y Computación, siendo el Ing. Manuel Eduardo Sucunuta España Director del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 88 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través, o con el apoyo financiero, académico o institucional (operativo) de la Universidad”

F.....

Autor: Jairo Antonio Contreras Mendieta

Cédula: 1104062524

## **DEDICATORIA**

La presente tesis está dedicada primeramente a Dios por la vida, la salud y las fuerzas necesarias que me permitieron llegar hasta estas instancias y lograr uno más de mis objetivos planteados, a mis queridos padres que con su amor, apoyo, motivación, consejos y buenos valores han sido un pilar fundamental en mi vida para cumplir con todas mis metas y a mis hermanos que de distintas formas me brindaron su apoyo incondicional en el transcurso de mi vida.

## **AGRADECIMIENTO**

Al culminar con éxito el presente proyecto de tesis quiero dejar constancia de mi sincero agradecimiento a la Universidad Técnica Particular de Loja, carrera de Sistemas Informáticos y Computación que viene proporcionando excelentes profesionales, a todos mis docentes que a lo largo de mi ciclo universitario me supieron brindar su aporte académico, sus sugerencias y criterio profesional que me ayudaron a cumplir con uno de mis objetivos planteados.

Finalmente agradezco a Dios por brindarme la fuerza y sabiduría necesaria para día a día luchar por alcanzar esta meta, y a mis padres los que nunca desconfiaron de mí y supieron apoyarme en cada momento.

## INDICE DE CONTENIDO

<b>APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN .....</b>	<b>ii</b>
<b>DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....</b>	<b>iii</b>
<b>DEDICATORIA.....</b>	<b>iv</b>
<b>AGRADECIMIENTO.....</b>	<b>v</b>
<b>INDICE DE CONTENIDO .....</b>	<b>vi</b>
<b>INDICE FIGURAS .....</b>	<b>ix</b>
<b>INDICE DE TABLAS .....</b>	<b>x</b>
<b>RESUMEN.....</b>	<b>1</b>
<b>ABSTRACT .....</b>	<b>2</b>
<b>INTRODUCCIÓN.....</b>	<b>3</b>
<b>PERSPECTIVA GENERAL.....</b>	<b>5</b>
<b>CAPITULO I:.....</b>	<b>6</b>
<b>ESTADO DEL ARTE .....</b>	<b>6</b>
<b>1.1 Marco Conceptual.....</b>	<b>7</b>
1.1.1 Recursos Educativos Abiertos.....	7
1.1.2 Condiciones Generales.....	7
1.1.3 Características Globales.....	8
1.1.4 Repositorios de Recursos Educativos Abiertos (RREA).....	9
1.1.5 Condiciones generales que debe cumplir un RREA.....	9
1.1.6 Funciones que debe cumplir un repositorio.....	10
<b>1.2 Metadatos.....</b>	<b>10</b>
1.2.1 Criterios para valorar el nivel de calidad de los metadatos que describe un Recurso.....	11
1.2.2 Principios para el diseño de esquema de metadatos.....	14
1.2.3 Estándares de metadatos para Recursos Educativos Abiertos.....	15
1.2.3.1 LOM (Learning Object Metadata = Metadatos de Objetos de Aprendizaje).....	15
1.2.3.2 Estándar Dublin Core.....	18
1.2.3.3 OBAA.....	19
1.2.3.4 Referencias entre Dublin Core y LOM.....	20
<b>1.3 Indicadores de calidad para determinar la importancia de un repositorio.....</b>	<b>21</b>
1.3.1 Indicadores según Westell.....	22
1.3.2 Indicadores según Alma Swan.....	22
1.3.4 Indicadores según Yong Ho King y Hyun Hee Kim.....	23
1.3.5 Indicadores según María Cassella.....	23
1.3.1 Indicadores según RECOLECTA (Recolector de Ciencia Abierta).....	24
1.3.2 Indicadores según Paul Vierkant.....	24
1.3.3 Indicadores según Javiera Atenas Rivera.....	25
<b>1.4 Extracción de metadatos.....</b>	<b>29</b>
1.4.1 Extracción utilizando SAXEF.....	29
1.4.1.1 Análisis en el nivel bajo.....	31
1.4.1.2 Análisis de texto en el nivel medio.....	31

1.4.1.3	Análisis en el nivel alto.....	31
1.4.2	Extracción utilizando Looking4LO.....	31
1.4.3	Extracción utilizando SCRAPY.....	33
1.4.4	Comparativa de métodos de extracción.....	36
<b>1.5</b>	<b>Sistemas Recomendadores.....</b>	<b>37</b>
1.5.1	Características Generales.....	37
1.5.2	Técnicas algorítmicas utilizadas en los sistemas de recomendación.....	38
1.5.3	Clasificación de los sistemas de recomendación basada en filtros.....	38
1.5.3	Sistemas de Recomendación basado en Latent Dirichlet Allocation (LDA).....	39
1.5.3.1	Funcionamiento y modelo de representación.....	40
1.5.3.2	Métricas de evolución de sistema recomendador en el contexto de LDA.....	42
1.5.3.2.1	Precisión.....	43
1.5.3.2.2	Recall o exhaustividad.....	43
1.5.4	Arquitectura de software.....	44
1.5.4.1	Cliente-Servidor.....	45
1.5.4.2	Basado en componentes de arquitectura.....	45
1.5.4.3	Arquitectura basada en capas.....	45
1.5.4.4	Arquitectura orientada a objetos.....	46
1.5.4.5	Arquitectura orientada a servicios.....	46
1.5.4.6	Arquitectura basada en micro servicios.....	46
<b>1.6</b>	<b>Resultados obtenidos.....</b>	<b>47</b>
 <b>CAPITULO 2:.....</b>		<b>49</b>
 <b>DESARROLLO DEL MODELO DE REPRESENTACIÓN.....</b>		<b>49</b>
<b>2.1</b>	<b>Indicadores de calidad para valorar la importancia de los repositorios.....</b>	<b>50</b>
<b>2.2</b>	<b>Análisis de repositorios en base a los indicadores de calidad establecidos.....</b>	<b>51</b>
2.2.1	MERLOT (Multimedia Educational Resources for Learning and Online Teaching).....	52
2.2.2	OER COMMONS (Open Educational Resources Commons).....	54
2.2.3	REPOSITORIO DIGITAL UTPL.....	55
2.2.4	UNESCO.....	56
2.2.5	Repositorio de objetos de aprendizaje (ROAP).....	58
2.2.6	LA FLOR (Federación Latinoamericana de Repositorios de Objetos de Aprendizaje).....	59
2.2.7	FEB (Federación de Repositorios Educa Brasil).....	60
2.2.8	PROCOMON.....	61
2.2.9	Comparación y selección de Repositorios.....	63
<b>2.3</b>	<b>Análisis de Repositorios seleccionados.....</b>	<b>65</b>
2.3.1	Metadatos MERLOT.....	65
2.3.2	Metadatos OER COMMONS.....	66
2.3.3	Metadatos UNESCO.....	66
2.3.4	Metadatos REPOSITORIO DIGITAL UTPL.....	67
<b>2.4</b>	<b>Análisis de metadatos en los repositorios.....</b>	<b>68</b>
2.4.1	Repositorio MERLOT.....	68
2.4.2	Repositorio UNESDOC.....	68
2.4.3	Repositorio OER COMMONS.....	69
2.4.4	Repositorio Digital UTPL.....	70
<b>2.5</b>	<b>Descripción del modelo común de metadatos de representación de recursos.....</b>	<b>70</b>

<b>CAPITULO 3:</b> .....	<b>74</b>
<b>IMPLEMENTACION DEL MODELO DE REPRESENTACION COMÚN</b> .....	<b>74</b>
<b>3.1 Diseño</b> .....	<b>75</b>
3.1.1 Arquitectura tres capas .....	75
3.1.2 Diagrama de secuencias de la herramienta.....	77
3.1.3 Interfaz del Buscador de REA.....	78
<b>3.2 Codificación</b> .....	<b>79</b>
3.2.1 Extracción de metadatos .....	79
3.2.2 Extracción de metadatos de Repositorio Merlot .....	80
<b>3.3 Resultados Obtenidos</b> .....	<b>83</b>
<b>CAPÍTULO 4:</b> .....	<b>84</b>
<b>VALIDACION DEL MODELO DE REPRESENTACION IMPLEMENTADO A TRAVÉS DE UN SISTEMA RECOMENDADOR</b> .....	<b>84</b>
4.1 Valoración de recomendaciones con algoritmo LDA .....	85
4.2 Precisión y Recall .....	92
<b>CONCLUSIONES</b> .....	<b>95</b>
<b>RECOMENDACIONES</b> .....	<b>96</b>
<b>REFERENCIAS</b> .....	<b>97</b>
<b>ANEXOS</b> .....	<b>101</b>

## INDICE FIGURAS

Figura 1: Definición de Repositorio.....	9
Figura 2: Representación de los elementos de LOM.....	17
Figura 3: Representación del nivel simple DublinCore.....	18
Figura 4: Representación de los elementos de OBAA .....	20
Figura 5: Arquitectura SAXEF .....	30
Figura 6: SAFEX Implementación.....	30
Figura 7: Visión general del Sistema Looking4LO .....	32
Figura 8: Arquitectura de Scrappy .....	34
Figura 9: Esquema del funcionamiento básico de un sistema recomendador. ....	37
Figura 10: Representación gráfica de LDA .....	40
Figura 11: Resumen capítulo 2 .....	50
Figura 12: Estructura de los metadatos de los recursos en el Repositorio Merlot.....	65
Figura 13: Estructura de los metadatos de los recursos en el Repositorio OER COMMONS .....	66
Figura 14: Estructura de los metadatos de los recursos en el Repositorio UNESDOC	67
Figura 15: Estructura de los metadatos de los recursos en el Repositorio Digital UTP .....	67
Figura 16: Porcentajes de Metadatos en Repositorios.....	71
Figura 17: Modelo metadatos para la representación común de los Recursos Educativos Abiertos .....	71
Figura 18: Resumen capítulo tres.....	75
Figura 19: Arquitectura tres capas.....	76
Figura 20: Diagrama de Secuencias de la aplicación.....	77
Figura 21: Página principal del buscador .....	78
Figura 22: Página de Resultados de la búsqueda .....	78
Figura 23: Metadatos del recurso educativo seleccionado y recursos relacionados ....	79
Figura 24: Estructura de la Base de datos .....	83
Figura 25: Diccionario LDA .....	85
Figura 26: Descripciones de los recursos educativos abiertos .....	86
Figura 27: Resultado de tópicos .....	87
Figura 28: Resultado de tópico_word.....	88

## INDICE DE TABLAS

Tabla 1: Repositorios de REA revisados para el cálculo de pesos de metadatos.....	12
Tabla 2: Metadatos con mayor peso .....	12
Tabla 3: Descripción de metadatos LOM .....	13
Tabla 4: Comparaciones para evaluar la coherencia .....	14
Tabla 5: Categoría General - LOM .....	17
Tabla 6: Dublin Core .....	19
Tabla 7: Referencias entre DublinCore y LOM.....	20
Tabla 8: Indicadores de Calidad según Westell .....	22
Tabla 9: Indicadores de Calidad según Alma Swan .....	22
Tabla 10: Indicadores de Calidad según Yong Ho King y Hyun Hee Kim.....	23
Tabla 11: Indicadores de Calidad según María Cassella.....	23
Tabla 12: Indicadores de Calidad RECOLECTA.....	24
Tabla 13: Indicadores de Calidad según Rivera.....	25
Tabla 14: Tabla comparativa de indicadores de calidad propuestos por varios autores .....	27
Tabla 15: Comparativa de los métodos de extracción .....	36
Tabla 16: Comparativa de los sistemas Recomendadores basados en filtros .....	39
Tabla 17: Estilos de Arquitecturas .....	47
Tabla 18: Criterios de calidad definidos para evaluar los repositorios.....	51
Tabla 19: Análisis de Merlot en base a indicadores de calidad .....	52
Tabla 20: Análisis de OER COMMONS en base a indicadores de calidad .....	54
Tabla 21: Análisis de Repositorio Digital UTPL en base a indicadores de calidad.....	55
Tabla 22: Análisis del Repositorio UNESDOC en base a indicadores de calidad .....	57
Tabla 23: Análisis de Repositorio ROAP en base a indicadores de calidad.....	58
Tabla 24: Análisis de Repositorio La FLOR en base a indicadores de calidad.....	59
Tabla 25: Análisis de Repositorio FEB en base a indicadores de calidad .....	61
Tabla 26: Análisis de Repositorio PROCOMUN en base a indicadores de calidad .....	62
Tabla 27: Escala de Valoración.....	63
Tabla 28: Incidencias de indicadores de calidad en repositorios .....	64
Tabla 29: Selección de los repositorios de acuerdo al peso .....	64
Tabla 30: Incidencia de metadatos en Repositorio Merlot.....	68
Tabla 31: Incidencia de metadatos en Repositorio UNESDOC .....	69
Tabla 32: Incidencia de metadatos en Repositorio OER COMMONS.....	69
Tabla 33: Incidencia de metadatos en Repositorio Digital UTPL .....	70
Tabla 34: Modelo de metadatos para la representación común de los Recursos Educativos Abiertos .....	72
Tabla 35: Topic_World = n_topics= 10, n_iter= 100 .....	88
Tabla 36: Topic_World = n_topics= 10, n_iter= 200 .....	89
Tabla 37: Topic_World = n_topics= 10, n_iter= 5.....	89
Tabla 38: Topic_World = n_topics= 20, n_iter= 100 .....	89
Tabla 39: Topic_World = n_topics= 20, n_iter= 200 .....	90
Tabla 40: Topic_World = n_topics= 20, n_iter= 5.....	90
Tabla 41: Doc_Topic.....	91
Tabla 42: Tabla de Contingencia .....	93

## RESUMEN

El presente trabajo de titulación está basado en la necesidad de la búsqueda de información en los diferentes repositorios, teniendo en cuenta que existen factores que dificultan el encontrar estos recursos, estos problemas se dan principalmente debido a las diferentes estructuras de metadatos que utilizan distintos estándares, se debe tomar en cuenta también que los estándares de calidad como ISO, CEN O ADL entre otros no siempre se encuentran presentes en los repositorios. Es por esta razón que existe la necesidad de construir una sola estructura de metadatos en la cual se pueda unir los recursos de diferentes repositorios, que facilite al usuario en las búsquedas de recursos, por esta razón se plantea un modelo de representación común para Repositorios de Recursos Educativos Abiertos que se encuentra implementado en una aplicación web.

Palabras Claves: Recursos Educativos Abiertos, Repositorios Digitales, Metadatos, Estándares de Calidad.

## **ABSTRACT**

The present titling work is based on the need of information search in the different repositories, taking into account that there are factors that make it difficult to find these resources, these problems are mainly due to the different metadata structures that use different standards, It should also be taken into account that quality standards such as ISO, CEN or ADL among others are not always present in repositories. It is for this reason that there is a need to construct a single metadata structure in which the resources of different repositories can be joined, to facilitate the user in the search of resources, for this reason a common representation model for Repositories is proposed. Open Educational Resources that is implemented in a web application.

Keywords: Open Educational Resources, Digital Repositories, Metadata, Quality Standards

## INTRODUCCIÓN

Con el paso del tiempo, el avance tecnológico ha permitido utilizar distintos métodos de estudios en línea que exige mayor exclusividad de la información proporcionada, en la sociedad académica y científica los repositorios son una fuente de información especializada de mucha importancia para el desarrollo de trabajos científicos o para ampliar conocimientos de forma libre a lectores y estudiantes de diversas Área e incluso presenta una idea de innovación para las instituciones educativas. Estos avances tecnológicos se ven reflejados en la búsqueda de información que conlleva desde la coincidencia de palabras, a darle sentido a las palabras para obtener un mejor resultado.

A nivel mundial una característica notoria son el favorecimiento de estos avances implementados como recursos de aprendizaje, considerados como una de las innovaciones tecnológicas más destacadas, no solo por presentar una infraestructura para almacenamiento y acceso a cantidades masivas de información sino también por representar un medio para aprender desde cualquier lugar y momento, convirtiendo así, una sociedad de información en una sociedad del conocimiento, todo va relacionado a ser una ayuda tanto en contenidos didácticos y prácticos como investigaciones, proyectos y la educación en general. La era del conocimiento representará el cambio tanto en el ámbito económico como social actual, frente a ese pensamiento se consideran los recursos digitales como un requerimiento fundamental ajustable a diferentes modelos educativos, disciplinas y niveles de estudio, además una característica adicional de los repositorios es que posee recursos reutilizables y están enmarcados generalmente a la iniciativa Open Access que permite el acceso libre y abierto sin restricciones legales y económicas.

El presente Trabajo de Titulación se realizó con el propósito de desarrollar un modelo de representación común para los repositorios de Recursos Educativos Abiertos, para lo cual se plantearon los siguientes objetivos:

- Analizar los diferentes modelos de representación de REA en los repositorios.
- Analizar y determinar los criterios de relevancia para valorar la importancia de los repositorios y recursos
- Desarrollar un modelo de representación común para REA que utilizan diferentes estándares, basados en criterios de búsqueda del usuario.

- Implementar una herramienta para extraer recursos en base al modelo de representación propuesto
- Validar el modelo implementada mediante pruebas

Para atender estos objetivos se desarrollaron los siguientes capítulos. En el capítulo 1 Estado del arte en donde se encuentran las definiciones que permitirán comprender los temas con los que se trabajara en el desarrollo de la tesis como repositorio, indicadores de calidad para determinar la importancia de un repositorio, recurso educativo abierto, metadatos, estándares de metadatos, sistemas de extracción y sistemas recomendadores .

En el capítulo 2 Desarrollo del modelo de representación primeramente se determina indicadores de calidad para así seleccionar repositorios de recursos educativos abiertos entre varias opciones brevemente analizadas, para posteriormente después de un análisis de cada repositorio diseñar un modelo de representación común de metadatos a utilizar.

En el capítulo 3 Implementación del modelo de representación común se desarrollará una herramienta que nos permita realizar la extracción de metadatos de recursos educativos abiertos de los repositorios previamente seleccionados, para luego plasmarlos bajo un aplicativo web que se encuentra estructurado en base al modelo común de metadatos definido en el capítulo anterior.

En el capítulo 4 Validación del modelo de representación a través de un Sistema Recomendador, se elaborará un Sistema de Recomendación en el contexto LDA que nos permitirá validar el modelo planteado, además de utilizar métricas como Precisión y Recall que nos ayudarán para evaluar el sistema recomendador.

## PERSPECTIVA GENERAL

### Problemática

Enfocándose desde la perspectiva en que los Repositorios de Recursos Educativos Abiertos (REA) ayudan a la investigación en amplios campos, y teniendo en cuenta que cada repositorio maneja un modelo de metadatos diferente, surge la necesidad de crear un modelo común para que exista congruencia entre un repositorios, facilitando la presentación de información y así darle mayor veracidad a los recursos como resultado.

Los REA contienen cursos y programas curriculares, módulos didácticos, guías para el estudiante, libros, artículos, videos, herramientas de evaluación, materiales interactivos, bases de datos, software, aplicaciones y cualquier otro material educativo diseñado para el uso de la enseñanza y el aprendizaje.

Todos estos recursos están agrupados en los repositorios de las instituciones que los han desarrollado y los que utilizan diferentes tipos de estándares de metadatos para describirlos, para que puedan ser aprovechados por profesores y estudiantes.

La diversidad de estas herramientas ha hecho que los recursos no se manejen de forma común y que resulten difíciles de localizar en un determinado contexto por parte de los interesados. Es necesario organizar y estructurar los diferentes metadatos y acoplar a un modelo común independiente de los elementos tecnológicos, arquitectura o idioma.

El propósito del trabajo de titulación es definir y desarrollar un modelo que permita representar los diferentes estándares que utilizan los repositorios de Recursos Educativos Abiertos. Además incluye el desarrollo de una herramienta que implementa el modelo de representación en base a la extracción y análisis de los metadatos.

**CAPITULO I:**  
**ESTADO DEL ARTE**

## 1.1 Marco Conceptual

### 1.1.1 Recursos Educativos Abiertos

Según (UNESCO, 2011) el concepto de Recurso Educativo Abierto (REA), describe cualquier tipo de recurso (incluyendo planes curriculares, materiales de los cursos, libros de texto, vídeo, aplicaciones multimedia, secuencias de audio, y cualquier otro material que se haya diseñado para su uso en los procesos de enseñanza y aprendizaje) que están plenamente disponibles para ser utilizados por parte de educadores y estudiantes, sin la necesidad de pago alguno por derechos o licencias para su uso.

La fundación “William and Flora Hewlett Foundation” define los REA como “recursos destinados para la enseñanza, el aprendizaje y la investigación que residen en el dominio público o que han sido liberados bajo un esquema de licenciamiento que protege la propiedad intelectual y permite su uso de forma pública y gratuita o permite la generación de obras derivadas por otros. Los Recursos Educativos Abiertos se identifican como cursos completos, materiales de cursos, módulos, libros, videos, exámenes, software y cualquier otra herramienta, materiales o técnicas empleadas para dar soporte al acceso de conocimiento” (Atkins, Seely, y Hammond, 2007; p.4).

Los REA muchas veces implican cuestiones de los derechos de propiedad intelectual, los libros por ejemplo se encuentran protegidos por los derechos de autor, no obstante existen otras opciones de licencias alternativas que son más flexibles que han aparecido como resultado de Creative Commons que ofrece acuerdos que son menos restrictivos.

### 1.1.2 Condiciones Generales.

Según (Ministerio de Educación Nacional de Colombia, 2012) los REA deben de cumplir con 3 condiciones, ser educativo, digital y abierto:

- **Educativo:** el recurso debe apoyar el proceso de enseñanza/aprendizaje facilitando la comprensión y entendimiento de conceptos lo que permite aumentar el conocimiento, además de desarrollar y promover en el individuo competencias, habilidades y capacidades que le permitan utilizarlas en cualquier ámbito ya sea social, cultural, tecnológico entre otros.
- **Digital:** todo recurso debe estar programado en código binario lo que facilita la producción, almacenamiento, distribución, adaptación, intercambio, modificación y disposición de un recurso en un entorno.
- **Abierto:** el autor del recurso otorga todos los permisos necesarios mediante un sistema de licenciamiento para el uso, modificación y adaptación de dicho recurso de forma gratuita.

### 1.1.3 Características Globales.

Partiendo de referentes conceptuales con reconocimiento internacional como: Institute of Electrical and Electronics Engineers<sup>1</sup>, World Wide Web Consortium<sup>2</sup> e International Electrotechnical Commission<sup>3</sup>, un REA debe tener las siguientes características:

- **Accesible:** cualidad que busca garantizar que el REA pueda ser consultado y/o utilizado por el mayor número de personas.
- **Adaptable:** propiedad de un REA que le permite ser modificado, ajustado o personalizado de acuerdo con los intereses, necesidades o expectativas del usuario.
- **Durable:** cualidad que garantiza su vigencia y validez en el tiempo, la cual se logra con el uso de estándares y tecnologías comunes y reconocidas para ese fin.
- **Flexible:** característica que posee un REA para responder e integrarse con facilidad a diferentes escenarios digitales de usuario final, de modo que este último pueda configurar su uso según sus preferencias
- **Interoperable:** propiedad que le permite a un REA contar con las condiciones, y estar en capacidad de ser implementado en diversos entornos digitales (ambientes, plataformas, canales y medios), bajo un conjunto de estándares o especificaciones reconocidas que permitan su plena funcionalidad. Esta característica es transparente para el usuario final.
- **Modular:** capacidad que le permite interactuar o integrarse con otros, en igual o diferentes condiciones y contextos.
- **Portable:** particularidad de REA en la cual son diseñados, construidos y ensamblados para poder ser empleados en una o más plataformas. Además, es una cualidad que promueve el uso del recurso y mejora sus posibilidades de almacenamiento y distribución.
- **Usable:** propiedad que garantiza la correcta interacción del REA con el usuario, con el fin de procurar una experiencia cómoda, fácil y eficiente.
- **Reusable:** cualidad que permite que el REA sea utilizado en diferentes contextos y con distintas finalidades educativas, permitiendo la adaptación o modificación de sus componentes

---

<sup>1</sup> Institute of Electrical and Electronics Engineers :<https://www.ieee.org/index.html>

<sup>2</sup> World Wide Web Consortium : <http://www.w3c.es/>

<sup>3</sup> International Electrotechnical: <http://www.iec.ch/>

#### 1.1.4 Repositorios de Recursos Educativos Abiertos (RREA)

La palabra repositorio se deriva del latín “Repositorium” que significa armario y el diccionario de la Real Academia Española<sup>4</sup> define como: “lugar donde se guarda algo”.

En un Repositorio se almacenan componentes que tienen identidad propia, y por lo tanto deben ser localizables. Éstos son tan variados como gráficos, imágenes, textos, videos, sonido, entre otros. (Temesio, 2012)

Los RREA son considerados como sitios web en donde se almacena información digital, cuyo objetivo es el de organizar, archivar, preservar y difundir la producción intelectual resultante de la actividad investigadora de la entidad a la que pertenece el repositorio y de esta manera apoyar a prácticas educativas abiertas como la búsqueda y recuperación. Para entender más esta definición se trató de plasmar la figura 1 explicando este concepto gráficamente.



**Figura 1: Definición de Repositorio**

Fuente: (Morales & Vicari, 2013)

#### 1.1.5 Condiciones generales que debe cumplir un RREA.

Un repositorio debe de cumplir con las siguientes condiciones para que se lo pueda denominar un RREA, condiciones que permitan dar una veracidad y confiabilidad a la información de manera que sea segura y el usuario tenga una conformidad con lo que va a utilizar.

- El repositorio pertenezca y esté registrado en una Institución de Educación Superior (IES) que proteja y garantice la calidad y propiedad del recursos educativo. (Ministerio de Educación Nacional de Colombia, 2012)
- Los documentos almacenados en el repositorio deben haber sido creados por la institución o alguno de sus miembros como producto de sus investigaciones. Si en el repositorio se mezclan diferentes tipos de colecciones (científica, docente, cultural, institucional), cada una de ellas deberá estar claramente

<sup>4</sup> Real Academia Española : <http://www.rae.es/>

diferenciada tanto en el repositorio como en los metadatos que se pongan a disposición de recolectores. (Azorín et al., 2014)

- Los contenidos depositados en un repositorio deben ser de libre acceso es decir libre de la mayoría de las barreras impuestas por las licencias y los derechos de autor.
- Los documentos en un RREA están descritos utilizando un número suficiente de metadatos basados en estándares internacionales y mínimamente normalizados, organizados mediante la aplicación de alguna clasificación de contenidos; que sirven para responder a las necesidades de sus usuarios y prestarles servicios de valor añadido (Azorín et al., 2014).
- El repositorio no es un fin en sí mismo, sino que su verdadero potencial se descubre cuando sus contenidos se integran en un nivel superior de agregación desde donde se puedan prestar servicios especializados a comunidades concretas. Por lo tanto, es fundamental la interoperabilidad del repositorio a través de la utilización de protocolos como el OAI-PMH. (Azorín et al., 2014)

#### 1.1.6 Funciones que debe cumplir un repositorio

(Cristobal, 2010) propone un conjunto básico de funciones que los repositorios deben proveer a fin de dar acceso a los objetos de aprendizaje en un ambiente seguro. Estas funciones son:

- **Buscar/encontrar:** localizar un objeto de aprendizaje apropiado
- **Recuperar:** obtener un objeto de aprendizaje previamente encontrado.
- **Enviar:** entregar a un repositorio un objeto de aprendizaje para ser almacenado.
- **Almacenar:** poner dentro de un registro de datos un objeto, con un identificador único que le permita ser localizado.
- **Colectar:** almacenar metadatos de los objetos de otros repositorios mediante búsquedas federadas.
- **Publicar:** presentar los metadatos del objeto de aprendizaje.

## 1.2 Metadatos

Debido a la gran cantidad de fuentes y recursos en Internet, es necesario establecer un conjunto estructurado de datos que nos permitan etiquetar, catalogar, describir y

clasificar los recursos presentes en el internet con el fin de facilitar posteriormente la búsqueda y recuperación de la información.

Los metadatos son datos que describen a otros datos; son datos estructurados que pueden describir colecciones de objetos, así como también, las relaciones entre los mismos, describen el contenido, la calidad, la condición de los datos. (Flores, 2009)

Surge así el concepto de metadatos. Literalmente, los metadatos son «datos acerca de los datos», y su misión es explicar el significado de los datos. Es decir, ayudan a los usuarios de los datos a entender mejor el significado que estos tienen y la información que guardan. (Blake & Olaya, 2016)

Algunos autores como (Sheldon, 2001) proponen la definición “informaciones sobre datos”, (Steinacker, Ghavam, & Steinmetz, 2001) “datos sobre informaciones”, o (Swick, 2002) “informaciones sobre informaciones”.

Una vez analizado e interpretado las definiciones de metadatos por los autores antes mencionados se tomara a un metadatos como:” un conjunto organizado de datos que se emplean para describir otros datos, su estructura, sus funcionalidad es y servicios.”(Manso, 2009)

### **1.2.1 Criterios para valorar el nivel de calidad de los metadatos que describe un Recurso.**

Según (Tobares, Duque, Cadavid, Ovalle, & Vicari, 2013) los criterios para determinar el nivel de calidad de los metadatos que describe los recursos abiertos son:

- **Compleitud:** indica si los metadatos describen los objetos tanto como sea posible. Esta característica es evaluada para comprobar si cada elemento de los metadatos es efectivamente parte del estándar correspondiente (Morales & Vicari, 2013a). De acuerdo a (Morales & Vicari, 2013b) la calidad de los metadatos disminuye con la ausencia de elementos básicos que se utilizan para la búsqueda y recuperación, por lo tanto el criterio de completitud se centra en determinar cuantitativamente la calidad de los metadatos (Temesio, 2015)

Por su parte en (Rodríguez, Isaza, & Duque, 2011) se concluye que los metadatos usados principalmente con fines de búsqueda, su importancia radica en aquellos asociados con la información general, el ciclo de vida y las características educacionales.

De acuerdo a (Tobares et al., 2013) menciona que todos los campos no poseen la misma importancia, de tal manera que propone el uso de un sistema

de pesos para considerar la importancia relativa de los mismos. Estos autores para determinar dichos pesos revisaron 12 Repositorios de REA diferentes, presentados en la Tabla 1.

**Tabla 1: Repositorios de REA revisados para el cálculo de pesos de metadatos**

#	Nombre	URL
1	MERLOT	http://www.merlot.org/
2	GATEWAY	http://www.thegateway.org/
3	MARICOPA	http://www.mcli.dist.maricopa.edu/mlx/
4	GLOBE	http://www.globe-info.org/
5	ARIADNE	http://ariadne.cs.kuleuven.be/AriadneFinder/
6	FEB	http://feb.ufrgs.br
7	SMETE	http://www.smete.org/smete/
8	AGREGA	http://www.proyectoagrega.es
9	FUNES	http://funes.uniandes.edu.co/cgi/search/simple
10	LA FLOR	http://laflor.laclo.org/
11	Colombia Aprende	http://64.76.190.172/drupalM/
12	ROAp	http://roap.medellin.unal.edu.co/

Fuente: (Tobares et al., 2013)

En base a los 12 RREA se evaluó cuáles son los metadatos usados en las búsquedas avanzadas y cuáles eran exhibidos a los usuarios en los resultados de las búsquedas. En el [Anexo 1: Resultados de la revisión de ROAs para cálculo de peso de metadatos](#), se puede observar la suma de los valores binarios por cada uno de los metadatos de acuerdo a las búsqueda avanzada y los resultados, llevando luego este valor a porcentaje respecto a la totalidad de los metadatos, elevando previamente al cuadrado para enfatizar la importancia relativa, y por último los valores en porcentajes por cada uno de los metadatos analizados. A continuación en la tabla 2 se recoge los cinco metadatos con mayor peso según su utilización en los repositorios analizados.

**Tabla 2: Metadatos con mayor peso**

Metadatos LOM	Pesos
General.Title	22,18%
General.Keyword	19,50%
General.Description	12,48%
LifeCycle.Contribute.Entity	12,48%
Educational.LearningResourceType	8,67%

Fuente: (Tobares et al., 2013)

En base a estudios realizados por diversos trabajos se determina un subconjunto de metadatos según la importancia relativa que consideren ciertos campos (Bui & Park, 2006), (Hughes, 2004) y (Ochoa & Duval, 2006), evalúan sólo algunos campos, los cuales son considerados “los más frecuentemente utilizados”. Por su parte en (Rodríguez et al., 2011) se concluye que los metadatos usados principalmente con fines de búsqueda, por tanto de interés para la completitud, son aquellos asociados con la información general, el ciclo de vida y las características educacionales.

- **Consistencia:** consiste en que los metadatos se acoplen al estándar o estándares utilizados por el repositorio. Como ejemplo LOM aplica dos reglas de consistencia que los metadatos deben cumplir y dos niveles de conformidad: conformidad estricta es cuando el recurso contiene únicamente metadatos definidos por el estándar y la conformidad básica en la que los metadatos pueden incluir elementos adicionales a los especificados por el estándar. (Tobares et al., 2013)

Los autores antes mencionados indican que la consistencia la evalúan a partir de la especificación del estándar de metadatos, el cual establece si un determinado metadato puede tomar valores libres, o si por el contrario existe una lista de valores posibles.

Los valores posibles de los metadatos que así lo establecen, tal como se muestra en la Tabla 3. Para calcular la métrica de consistencia se revisa para un OA si existe algún valor para cada uno de dichos metadatos y, en caso afirmativo, se asigna un valor de 1 si cumple con el valor permitido o de 0 en caso contrario.

**Tabla 3: Descripción de metadatos LOM**

Metadatos LOM	Valores Posibles
General.Agregation Level	1, 2, 3, 4
General.Structure	atomic, collection, networked, hierarchical, linear
LifeCycle.Contribute.Role	author, publisher, unknown, initiator, terminator, validator, editor, graphical designer, technical implementer, content provider, technical validator, educational validator, script writer, instructional designer, subject matter expert
LifeCycle.Status	draft, final, revised, unavailable
Meta-Metadata.Contribute.Role	creator, validator
Educational.Interactivity Type	active, expositive, mixed
Educational.Learning Resource Type	exercise, simulation, questionnaire, diagram, figure, graph, index, slide, table, narrative text, exam, experiment, problem, statement, self assessment, lecture
Educational.Interactivity Level	very low, low, medium, high, very high
Educational.Semantic Density	very low, low, medium, high, very high
Educational.Intended End User Role	teacher, author, learner, manager
Educational.Context	school, higher education, training, other
Educational.Difficulty	very easy, easy, medium, difficult, very difficult
Rights.Copyright and Other Restrictions	yes, no
Rights.Cost	yes, no
Classification.Purpose	discipline, idea, prerequisite, educational objective, accessibility, restrictions, educational level, skill level, security level, competency

Fuente: (Tobares et al., 2013)

- **Coherencia:** tiene que ver con el grado en la que los metadatos describen al recurso, es decir si los metadatos son bien utilizados al momento en que se describe el recurso.

(Tobares et al., 2013) comparan y revisan el valor de un metadato en relación con el valor de otro y de esta forma determinan si existen metadatos que tienen una alta correlación y así determinar su coherencia, A diferencia de la completitud y consistencia se establece un valor escalonado que da cuenta de la pertinencia de cada combinación de valores. En la tabla 4 se presenta las comparaciones para evaluar la coherencia

**Tabla 4: Comparaciones para evaluar la coherencia**

#	Metadatos y Valores		Resultado
1	<b>General.Structure</b>	<b>General.AgregationLevel</b>	
	atomic	1	1
	atomic	2	0,5
	atomic	3	0,25
	atomic	4	0,125
	collection, networked, hierarchical, linear	1	0,5
	collection, networked, hierarchical, linear	2, 3, 4	1
2	<b>Educational.InteractivityType</b>	<b>Educational.InteractivityLevel</b>	
	active, mixed	very high, high, medium, low, very low	1
	expositive	very high, high	0
	expositive	medium	0,5
	expositive	low, very low	1
3	<b>Educational.InteractivityType</b>	<b>Educational.Learning Resource Type</b>	
	active	exercise, simulation, questionnaire, exam, experiment, problem statement, self assessment	1
	active	diagram, figure, graph, index, slide, table, narrative text, lecture	0
	expositive	exercise, simulation, questionnaire, exam, experiment, problem statement, self assessment	0
	expositive	diagram, figure, graph, index, slide, table, narrative text, lecture	1
	mixed	exercise, simulation, questionnaire, exam, experiment, problem statement, self assessment, diagram, figure, graph, index, slide, table, narrative text, lecture	1

Fuente: (Tobares et al., 2013)

### 1.2.2 Principios para el diseño de esquema de metadatos.

Estudios realizados por (Testa & Degiorgi, 2013) mencionan que en el año 2001 se realizó una convención en Ottawa, donde se establecieron los principios comunes que pueden servir de base para el diseño de cualquier esquema de metadatos o aplicación. Estos principios son:

- **Modularidad:** la arquitectura de los metadatos debe ser flexible de manera tal que se puedan combinar de manera interoperable, sintáctica (por ej. XML) y semánticamente elementos provenientes de diferentes esquemas

preestablecidos. Cada componente del esquema debe incluir la funcionalidad y los requisitos específicos para una determinada aplicación.

- **Extensibilidad:** los esquemas de metadatos deben ser lo suficientemente flexibles como para poder ajustarse a las necesidades específicas del repositorio, sin perder de vista la interoperabilidad proporcionada por el esquema base. Así pues, el elemento creador va estar presente en todos los esquemas, como así también el identificador, mientras que otros serán específicos para una aplicación particular.
- **Refinamiento:** el esquema de metadatos debe permitir un nivel de detalle más exhaustivo mediante, por ejemplo, la adición de calificadores a los elementos. Esto hará más específico su significado. Por ejemplo en el elemento colaborador se puede agregar un calificador que aclare su función: ilustrador, prologuista, editor, compilador, director, escenógrafo, etc. También el uso de vocabularios controlados ayuda al refinamiento dado que mejora la precisión de las descripciones. Este principio es muy importante para lograr una recuperación más profunda y precisa de la información contenida en el repositorio.
- **Plurilingüismo:** los esquemas de metadatos deben respetar la diversidad lingüística y cultural y la internacionalización de la Web, poniendo a disposición del usuario los recursos en su idioma nativo.

### **1.2.3 Estándares de metadatos para Recursos Educativos Abiertos**

Los estándares de metadato proporcionan, a los productores de datos, el formato y el contenido que describen convenientemente los datos, permitiendo a los usuarios evaluar la utilidad de los mismos para sus necesidades específicas.

Existen varias iniciativas de estandarización de metadatos dependiendo del dominio de aplicación por tal motivo se usa los estándares que nos permitan la descripción y recuperación de Recursos Educativos Abiertos.

#### **1.2.3.1 LOM (*Learning Object Metadata = Metadatos de Objetos de Aprendizaje*)**

LOM se centran en el conjunto mínimo de propiedades que permiten que los objetos educacionales sean gestionados, ubicados y evaluados. Por lo que se diseñó los siguientes objetivos (Barker, 2005):

- Crear descripciones bien estructuradas para recursos educativos. Estas descripciones deben facilitar el descubrimiento, localización, evaluación y adquisición de los recursos educativos por parte de los estudiantes, docentes o procesos de aplicaciones automáticas.
- Compartir descripciones de recursos educativos entre sistemas de descubrimiento de recursos. Esto conducirá a una reducción en el costo para proporcionar servicios basados en descripciones de calidad.
- Adecuación de las descripciones de recursos para cumplir las necesidades especializadas de una comunidad. Esto podría incluir la elección de vocabularios controlados para clasificación, reducir el número de elementos descritos o adicionar nuevos provenientes de otros esquemas de descripción de recursos.
- Los creadores y editores pueden usar LOM junto con otras especificaciones para "etiquetar" recursos con una descripción que pueda ser asociada con el recurso. Esto proporcionará información en un formato estándar similar al encontrado en la cubierta y portada de un libro de texto.

LOM es un modelo de datos usados para describir REA y otros recursos digitales similares usados para el apoyo al aprendizaje. En la Figura 2 se presenta los elementos de LOM.

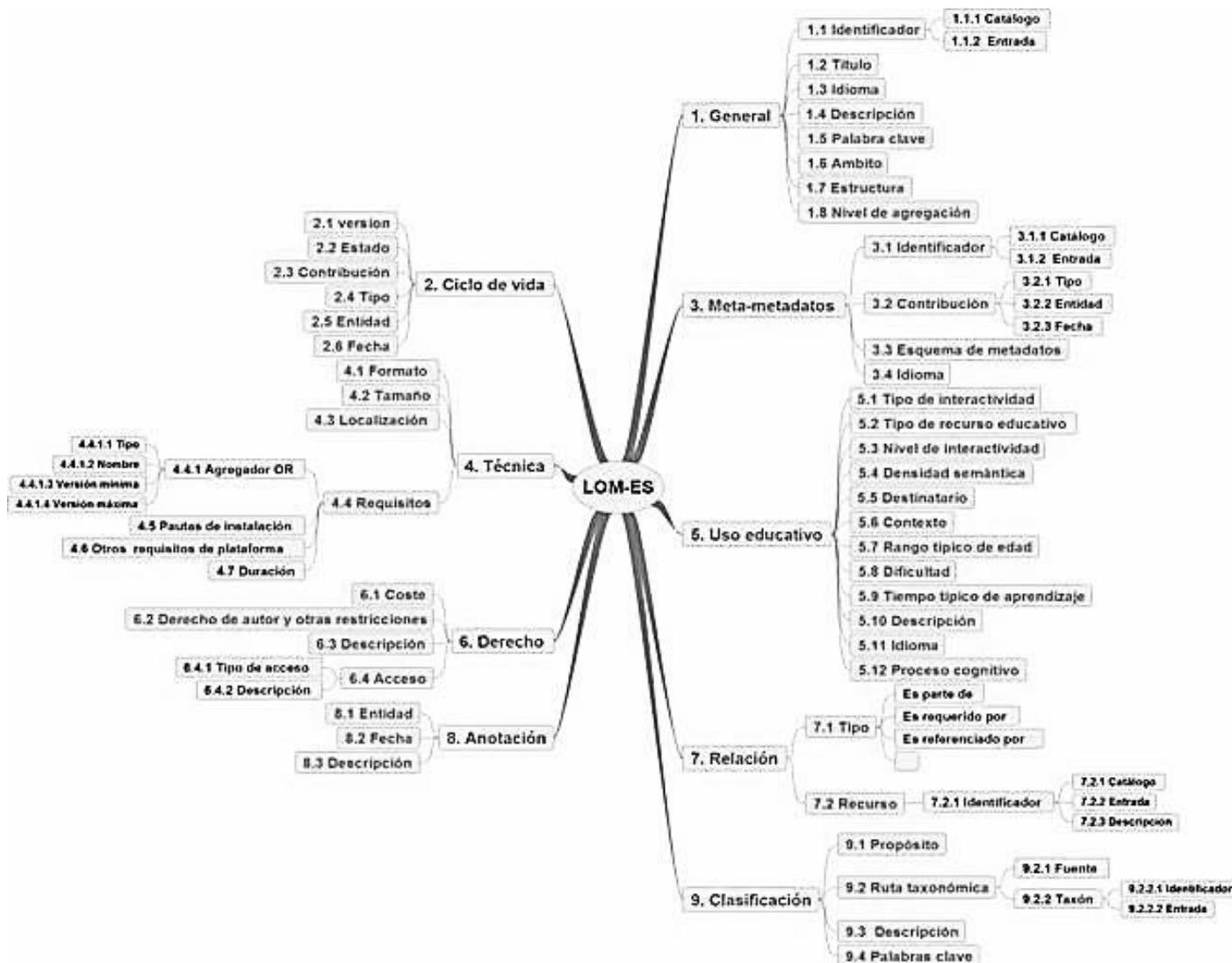


Figura 2: Representación de los elementos de LOM

Fuente: (Temesio, 2015)

Dentro del estándar los elementos pueden ser opcionales, obligatorios y recomendados, en la tabla 5 se definen los elementos de la categoría 1-General de LOM los que deben ser obligatorios y recomendados:

Tabla 5: Categoría General - LOM

Categoría General	Obligatorio
1.1	Identificador
	1.1.1 Catálogo
	1.1.2 Entrada
1.2	Título
1.3	Idioma

1.4	Descripción	Obligatorio
1.5	Palabra clave	Recomendado
1.6	Ámbito	Opcional
1.7	Estructura	Opcional
1.8	Nivel de agregación	Obligatorio

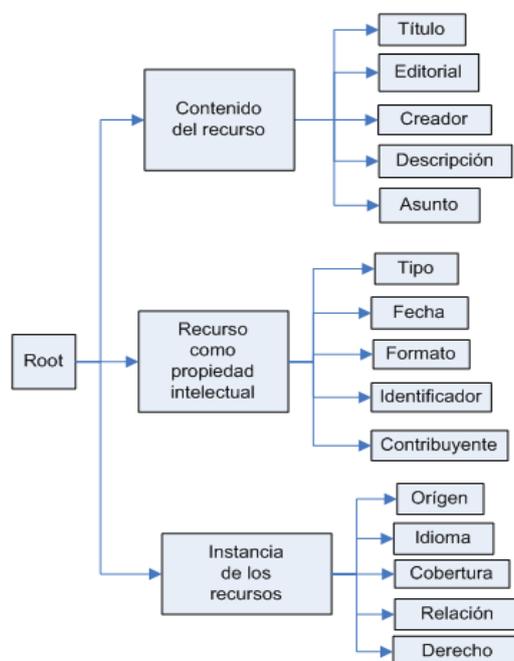
Fuente: (Temesio, 2015)

### 1.2.3.2 Estándar Dublin Core.

Este estándar pretende proporcionar un conjunto de metadatos que facilite la recuperación de recursos en un ambiente de red.

Es un estándar de metadatos simple pero eficaz compuesto por un conjunto de elementos para describir una amplia gama de recursos de red, cada elemento es opcional y puede repetirse (Castro, Castillo, Lillo, Quiroz, & Riquelme, 2009). En el anexo 2 podemos visualizar un ejemplo de cómo trabajar con Dublin Core.

En la figura 3 se presenta la estructura del modelo de metadatos del estándar Dublin Core.



**Figura 3: Representación del nivel simple Dublin Core**

Fuente: (Ward, 2002)

Al igual que LOM este estándar también tiene campos obligatorio, tal como se presentan en la tabla 6 (Ferrerias, 2008), en la tabla

**Tabla 6: Dublin Core**

1. Título	Obligatorio
2. Creador	Obligatorio
3. Tema	Obligatorio
4. Descripción	Obligatorio
5. Editor	Obligatorio
6. Contribuyente	Opcional
7. Fecha	Obligatorio
8. Tipo	Obligatorio
9. Formato	Recomendado
10. Identificador	Obligatorio
11. Fuente	Opcional
12. Idioma	Recomendado
13. Relación	Opcional
14. Cobertura	Opcional
15. Derechos	Recomendado

Fuente: (Ferrerias, 2008)

### **1.2.3.3 OBAA**

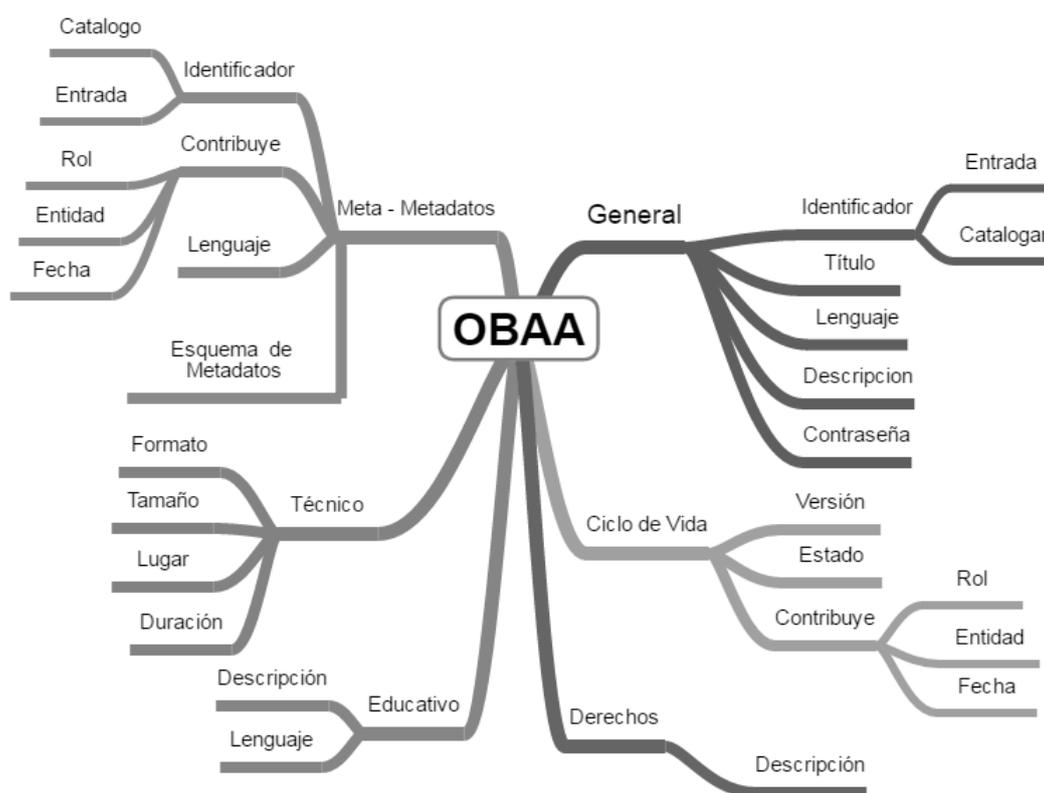
Es una iniciativa brasileña desarrollada por la Universidad Federal de Rio Grande do Sul(UFRGS),que propone un estándar de metadatos para OAs compatible con el estándar IEEE LOM.

OBAA se basa en el estándar de metadatos de objetos de aprendizaje IEEE LOM, con todas sus categorías y metadatos, complementando las categorías de técnica y educación e incluyendo dos nuevas categorías que representan los aspectos de accesibilidad y segmentación (Vicari et al., 2009).

De acuerdo (Vicari et al., 2009) las funcionalidades que nos ofrece OBAA son:

- Adaptabilidad
- Independencia Tecnológica
- Accesibilidad
- Contextualización
- Interoperabilidad

En la figura 4 se presenta los metadatos de la propuesta OBAA:



**Figura 4: Representación de los elementos de OBAA**

Fuente: (Temesio, 2015)

### 1.2.3.4 Referencias entre Dublin Core y LOM

Según (Temesio, 2015) (Vicari et al., 2010) los estándares más usados y de mayor incidencia en la descripción de metadatos de Recursos Educativos Abiertos son el Dublin Core y el LOM, por tal motivo utilizan las referencias para traducir campos comunes entre estos dos estándares. Cabe recalcar que estas referencias son condicionadas si la relación de dos campos depende de un tercer campo, dicha relación se la presenta en la tabla 7.

**Tabla 7: Referencias entre DublinCore y LOM**

Dublin Core	LOM	Condición	Diferencias
DC.Identifier	DC.Identifier	No hay condición	Dublin Core utiliza dc: alternativa para variantes del título
DC.Title	General.Title	No hay condición	
DC.Language	General.Language	No hay condición	
DC.Description	General.Description	No hay condición	Dublin core para complementarse

			usa los calificadores, Table of Contents y Abstract
DC.Subject	General.Keyword	No hay condición	No existe diferencias
DC.Coverage	General.Coverage	No hay condición	No existe diferencias
DC.Type	Educational.LearningResourceType	No hay condición	No existe diferencias
DC.Date	LifeCycle.Contribute.Date	LifeCycle.Contribute.Role tiene un valor de Editor	No existe diferencias
DC.Creator	LifeCycle.Contribute.Entity	LifeCycle.Contribute.Role tiene un valor de Autor	No existe diferencias
DC.OtherContributor	LifeCycle.Contribute.Entity	Tipo de contribución especificada en LifeCycle.Contribute.Role	No existe diferencias
DC.Publisher	LifeCycle.Contribute.Entity	LifeCycle.Contribute.Role tiene un valor de Editor	No existe diferencias
DC.Format	Technical.Format		Dublin core utiliza Extent: definido para el tamaño o duración del recurso. Medium: definido para el material o "envase" físico del recurso
DC.Rights	Rights	No hay condición	LOM utiliza para describir este campo cost, copyright and other restrictions y description mientras que dublin core utiliza unicamente accesRights
DC.Relation	Relation.Resource.Description	No hay condición	LOM emplea mientras que dublin core usa

Fuente: (Guzmán & Peñalvo, 2012)

### 1.3 Indicadores de calidad para determinar la importancia de un repositorio

Contar con indicadores de calidad se fundamentan como indispensable en el proceso de mejora continua de ahí la importancia de contar con ellos, de manera que aseguren la eficiencia y calidad de los mismos, además de detectar posibles fallos en cualquiera de los aspectos que lo forman.

No existe una sola forma para evaluar los repositorios, sin embargo sí se pueden utilizar algunos criterios generales. Se ha realizado una recopilación de trabajos sobre criterios aplicados a la evaluación y análisis de repositorios y a partir de éstos se aplicara el más adecuado.

### 1.3.1 Indicadores según Westell

Las primeras evaluaciones de repositorios fue el realizado por (Westell, 2006) que estableció indicadores de calidad que se deben de cumplir en un repositorio para considerarlo importante. Estos indicadores se los muestra en la tabla 8.

**Tabla 8: Indicadores de Calidad según Westell**

Indicador	Descripción
Políticas	Si existía un mandato institucional
Apoyo Institucional	Como se integra el repositorio con la planificación de la institución
Factores Económicos	Modelo de financiamiento
Retrospectivo	Relación con los centros de digitalización
Masa de documentos	Método de medición para determinar la cantidad de documentos que de estabilidad al repositorio
Interoperabilidad	Conectividad con otros repositorios
Marketing	Como se realiza la promoción del repositorio
Preservación	Estrategia para conservar el estado del repositorio y evitar que sufra algún tipo de daño.

Fuente: (Mary Westell, 2006)

### 1.3.2 Indicadores según Alma Swan

(Swan & Houghton, 2012) considera importante al momento de evaluar un repositorio los siguientes indicadores de calidad que se los muestra en la tabla 9.

**Tabla 9: Indicadores de Calidad según Alma Swan**

Indicadores	Descripción
Contenidos	El porcentaje anual de resultados de investigación actual introducidos en el repositorio de diferentes tipos (artículos de revistas, actas de congresos, tesis); porcentaje de investigación de años anteriores recuperados y depositados; colecciones especiales digitalizadas y almacenadas.
Conocimiento de los usuarios	Pueden medirse en el nivel de conocimiento de los autores del acceso al recurso abierto, y de los problemas de derechos de autor, así como del nivel de conocimiento de los problemas y desarrollos de la comunicación académica.
Flujo de trabajo del repositorio	Este indicador examina cinco puntos: si los procedimientos tienen la suficiente calidad, si los tiempos de rendimiento son estables o mejoran, si se anticipan y suavizan los picos y los tiempos de menos trabajo, o si el repositorio está inmerso en la institución
Financiación	Es fundamental para que un repositorio avance en los aspectos tecnológicos

Fuente: (Swan & Houghton, 2012)

### 1.3.4 Indicadores según Yong Ho King y Hyun Hee Kim

El estudio llevado a cabo con repositorios institucionales coreanos dividió los indicadores en cuatro categorías y cada una de ellas contenía hasta seis posibles indicadores, como se muestra en la tabla 10:

**Tabla 10: Indicadores de Calidad según Yong Ho King y Hyun Hee Kim**

Categoría	Indicador 1	Indicador 2	Indicador 3	Indicador 4	Indicador 5	Indicador 6
Contenido	Metadatos	Tamaño	Características del procedimiento	Características Formales		
Gestión y Política	Criterios de procedimiento o (legales y costes)	Apoyo de la Institución	Participación de profesores			
Sistema	Acceso	Integración	Confianza en el sistema	Interoperabilidad	Funcionalidad	Cooperación
Uso y usuarios	Accesibilidad	Usabilidad	Satisfacción de usuarios	Relevancia	Cumplimiento de plazos	Usuarios potenciales

Fuente: (Serrano Vicente, Melero Melero, & Abadal Falgueras, 2014)

### 1.3.5 Indicadores según María Cassella

El estudio de indicadores de evaluación del rendimiento de repositorios realizado por María Casella según (Serrano Vicente, Melero Melero, & Abadal, 2014) divide a los indicadores en internos que son aquellos que ayudan a los gestores del repositorio a alinear sus estrategias con la misión y objetivos de la institución; y los externos que ayudan a los gestores a medir el impacto a nivel nacional e internacional y determinar el valor como una herramienta de búsqueda por parte de los usuarios finales. En la tabla 11 se muestran los indicadores

**Tabla 11: Indicadores de Calidad según María Cassella**

Categoría	Indicador	Descripción
Internos	Perspectiva del usuario/depositante	Contenido, quien deposita, cantidad de documentos
	Procesos internos	Depósitos anuales, depósitos diarios, acceso a texto completo, colecciones
	Servicios del valor añadido	Cvs, estadísticas, IDs autor, guías, FAQs, vocabularios controlados, add-ons, alertas RSS, atom feeds
	Factores económicos	Coste por depósito y por descarga
	Aprendizaje y desarrollo del personal	Personal dedicado, personas implicadas, formación del personal

Externos	Interoperabilidad	Interacción con otras redes de repositorios consorciados (para unir esfuerzos)
	Financiación externa	Interacción con otras redes de repositorios consorciados (para unir esfuerzos)
	Participación en proyectos nacionales e internacionales	Nivel de internacionalización

Fuente: María Cassella según (Serrano Vicente et al., 2014)

### 1.3.1 Indicadores según RECOLECTA (Recolector de Ciencia Abierta)

En diciembre del 2010, se publicó una guía editada por REBIUN, FECYT y RECOLECTA (Cruz, Manuel, & València, 2014) en donde se obtuvo 31 indicadores de calidad que se los agrupado en las categorías mostradas en la tabla 12:

**Tabla 12: Indicadores de Calidad RECOLECTA**

Indicador	Descripción
Visibilidad:	Presencia del repositorio en directorios nacionales o internacionales, con su denominación normalizada.
Políticas	Explicitación pública por escrito de las políticas de reutilización de metadatos y contenidos, de archivo de documentos y de preservación de los contenidos.
Aspectos legales	Asesoramiento explícito en cuestiones relacionadas con la propiedad intelectual de los contenidos distribuidos en el repositorio.
Metadatos	Especificación de formatos usados, clasificaciones, indización de contenidos, uso de identificadores persistentes normalizados.
Directrices DRIVER:	Indicación sobre si los contenidos son de acceso abierto o no, uso del set driver, tratamiento de los registros borrados, formato de la respuesta a la orden Identify del protocolo OAI-PMH.
Logs y estadísticas:	Indicación y/o comprobación de almacenamiento de logs de acceso al servidor web.
Seguridad, autenticidad e integridad de los datos:	Constatación de la implementación de un procedimiento documentado de elaboración de copias de seguridad (procedimiento que puede no ser público).

Fuente: (Cruz et al., 2014)

### 1.3.2 Indicadores según Paul Vierkant

El estudio realizado por (Vierkant, 2012) a repositorios institucionales alemanes estableció los siguientes indicadores:

- Tamaño

- Donde se alojan
- Idiomas a los que se está traducido
- Servicios de valor añadido: posibilidades de exportar las citas, estadísticas de uso, RSS, marcadores sociales a nivel de ítem, posibilidades de enlazar con redes sociales a nivel de ítem.
- Software
- Formato de metadatos
- Registros en páginas de repositorios
- Firma de declaraciones de acceso abierto.

### 1.3.3 Indicadores según Javiera Atenas Rivera

(Atenas, 2014) define sus indicadores de calidad en base a la revisión literaria publicada entre los años 1996 y 2013, en la cual se obtuvo como resultado 10 indicadores de calidad para el diseño de REA, mismos que se presentan en la tabla 13.

**Tabla 13: Indicadores de Calidad según Rivera**

Indicadores	Descripción
Recursos destacados	Su relevancia radica por sus recursos novedosos o la calidad que son importantes para los usuarios.
Evaluación de recursos por parte de usuarios	Es necesario que los usuarios puedan evaluar los REA albergados en los repositorios utilizando diferentes medios ya sea por sistemas de votación, rankings o ratings.
Revisión por pares	El contar con la evaluación de expertos en la temática correspondiente al recurso asegura la calidad del mismo
Autoría de los recursos	Es de vital importancia el mencionar los autores de los recursos y que estos sean claramente identificados en los repositorios, ya que permitirá el posterior reconocimiento de las contribuciones además de ser la condición básica de las licencias abiertas.
Palabras claves	El uso de estos vocabularios es recomendable para facilitar la búsqueda y recuperación de los materiales.
Sistemas de metadatos	La literatura es enfática en recomendar la Implementación de sistemas como Dublin Core, Learning Object Metadata (IEE-LOM) u Open Archives Initiative (OAI-PMH), para facilitar la descripción estandarizada de los recursos y la interoperabilidad de los descriptores entre diferentes repositorios
Soporte multilingüe	Se recomienda el diseño de interfaces multilingües para utilizar las herramientas de búsqueda, selección y recuperación de contenidos en distintos idiomas por parte de los diferentes usuarios de los repositorios

Soporte a redes sociales	La literatura sugiere que los repositorios integren sistemas automatizados que permitan a los usuarios compartir el acceso directo a los recursos mediante redes sociales como Twitter, Facebook y Tumblr entre otras, utilizando modelos de enlaces permanentes (permalinks)
Licencias Creative Commons	Los RREA como requerimiento esencial deben explicitar claramente las licencias abiertas (Creative Commons) en todos y cada uno de los recursos que albergan, en caso de ser estos reutilizados por parte de usuarios del repositorio.
Acceso al código fuente y archivo original	Al ser un recurso abierto es necesario que los repositorios permitan y faciliten a los usuarios descargar el recurso en su formato original para poder ser adaptado, traducido o modificado

Fuente: (Atenas, 2014)

A continuación se presenta en la tabla 14, una tabla comparativa de todos los modelos referentes a los indicadores de calidad nombrados anteriormente por varios autores, partiendo su comparación de acuerdo a las diversas categorías que han sido planteadas.

**Tabla 14: Tabla comparativa de indicadores de calidad propuestos por varios autores**

Categorías/ Autores	Westell (2006)	Thibodeau (2007)	Swan (2012)	Kim (2008)	Cassella (2010)	PRG (2012)	RECOLECTA (2010)	Vierkant (2013)	Atenas, (2014)
Marketing	Medición				Estadísticas	Nº visitantes por año Nº descargas		Registros en páginas de repositorios Redes sociales	Recursos destacados
Económicos	Modelo de financiación		Planes de financiación anuales, monitorizar procedimientos, prever procesos		Capacidad del repositorio de atraer financiación	Presupuesto, ayudas específicas			
Interoperabilidad y cooperación	Conectividad con otros repositorios	Reparto de tareas con otros repositorios		Interoperabilidad	Interacción con otras redes de repositorios consorciados (para unir esfuerzos)			Idiomas en los que se puede buscar Formato de metadatos	Soporte multilingüe
Políticas	Mandato			Apoyo de la institución			Política pública sobre el depósito de documentos	Firmas de declaraciones de acceso abierto	Licencias Creative Commnos

Categorías/ Autores	Westell (2006)	Thibodeau (2007)	Swan (2012)	Kim (2008)	Cassella (2010)	PRG (2012)	RECOLECTA (2010)	Vierkant (2013)	Atenas, (2014)
Contenidos			Porcentaje anual de resultados de investigación actual introducidos en el repositorio de diferentes tipos Porcentaje de investigación de años anteriores recuperados y depositados						Evaluación de recursos por parte de usuarios Revisión por pares Palabras claves Autoría de los recursos Sistemas de metadatos
Servicios de valor añadido		Servicios de valor añadido			Personal dedicado, personas implicadas, formación del personal			Posibilidad de exportar las citas, estadísticas de uso, RSS, Marcadores sociales a nivel de ítem, posibilidad de enlazar con redes sociales por ítem.	Soporte a redes sociales

Fuente: (Serrano Vicente, Melero, & Abadal, 2014) y Autor

## **1.4 Extracción de metadatos**

La extracción de metadatos es una forma de garantizar la calidad de la información que se almacena en dichos repositorios, ya que partiendo de datos extraídos se realizará un proceso de análisis y se podrá determinar la veracidad de dicha información.

Para la extracción de los metadatos de un repositorio se debe de considerar 3 aspectos importantes como:

1. El tipo de archivos en los que se desea trabajar por ejemplo, html, txt, pdf, doc, etc.
2. Los metadatos que se extraerá
3. Las técnicas y herramientas que se utilizará para la extracción.

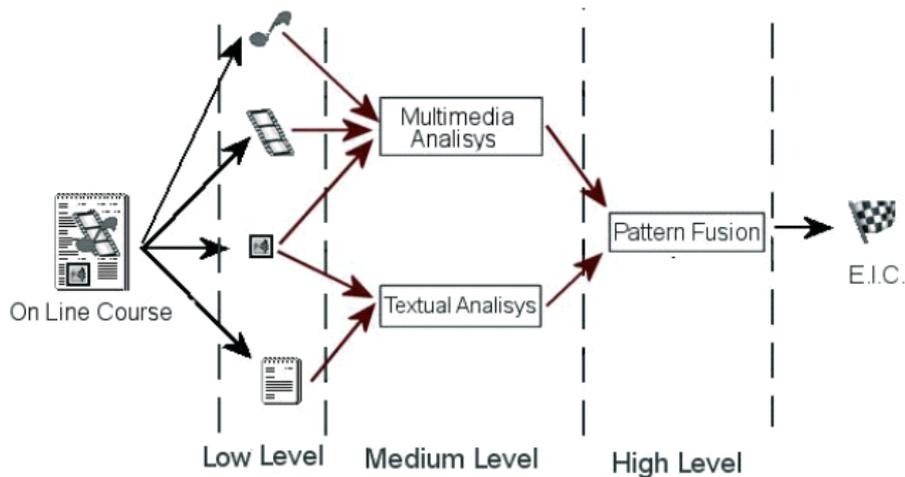
A continuación se muestran herramientas utilizadas para la extracción de metadatos.

### **1.4.1 Extracción utilizando SAXEF**

Es un Sistema creado por The Center on Communication Studies (Univ. Palermo, Italia), que automáticamente extrae indicadores didácticos de cualquier página Web. SAXEF produce una tarjeta de identificación E-learning (EIC) con el siguiente información sobre la naturaleza de un objeto (Alfano, Lenzitti, Visalli, Objects, & Indicators, 2007):

- Temas principales;
- Temas secundarios;
- Teórica o práctica;
- Sintética o analítica;
- Tipos de medios y el nivel multimedialidad;
- El nivel de complejidad;
- Enlaces a otros EIC con los mismos temas;
- Enlaces a otros EIC con temas relacionados.

En la figura 5 se muestra la arquitectura de SAXEF



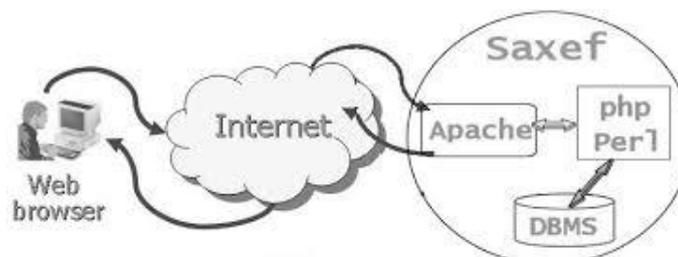
**Figura 5: Arquitectura SAXEF**

Fuente: (Alfano et al., 2007)

A continuación se detalla cómo funciona y trabaja cada uno de los niveles de la arquitectura de SAXEF:

- Nivel bajo: identifica y separa los componentes de los medios de los objetos de aprendizaje (texto, imágenes, vídeo, audio, animaciones, etc.);
- Nivel medio: extrae características de cada medio de comunicación mediante el uso de especificación algoritmos Lized (análisis de textos, análisis multimedia, ..);
- Nivel alto: fundir las características de los medios y mostrar la estructura e indicadores de los objetos de aprendizaje a través de la creación de sus EIC (Tarjeta de Identificación E-learning).

SAXEF se ha implementado como una aplicación web utilizando los lenguajes Perl y PHP y una base de datos MySQL. Perl y PHP han sido escogidos por su facilidad de uso, capacidad de cadena de manipulación, interfaz web óptima (HTML y XML) y posibilidad de insertar las consultas SQL en el código (SAFEX). El esquema de implementación se muestra en la figura 6.



**Figura 6: SAFEX Implementación**

Fuente: (Alfano et al., 2007)

#### **1.4.1.1 Análisis en el nivel bajo**

Una vez que el usuario proporciona la dirección de la página web para ser examinado SOFAX analiza su código (html, xhtml, asp, php, etc.), encuentra los objetos que lo componen y los almacena junto con la dirección de la página web.

#### **1.4.1.2 Análisis de texto en el nivel medio**

Se realiza un análisis del texto mediante los siguientes pasos:

- Artículos, preposiciones, pronombres, verbos comunes, etc. se eliminan.
- Las palabras con mayor ocurrencia individual y ocurrencia en parejas se calculan
- Se identifican las etiquetas más relevantes como son <title> y <meta>
- Se selecciona las palabras más importantes dentro del texto, este proceso se logra considerando los porcentajes de las palabras tanto individual como en parejas.
- Se le proporciona a cada palabra seleccionada un peso, este peso es un puntaje que obtiene la palabra dependiendo de donde y cuando aparezca la palabra en el texto.

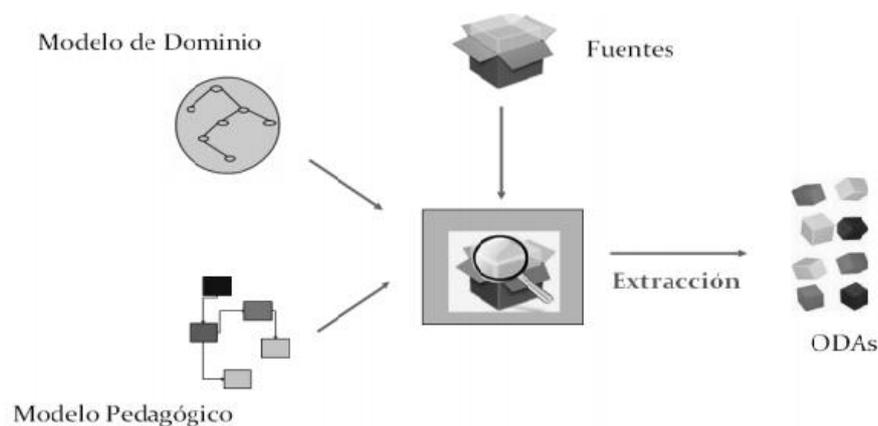
#### **1.4.1.3 Análisis en el nivel alto**

Esto se hace mediante la adopción de los resultados del análisis de texto y teniendo en cuenta los temas principales de las palabras con las dos puntuaciones más altas y los temas secundarios de las palabras con las siguientes cuatro puntuaciones más altas.

#### **1.4.2 Extracción utilizando Looking4LO**

Looking4LO es un sistema genérico y flexible, capaz de extraer OA con sus respectivos metadatos de archivos XML y HTML, documentos Word, presentaciones Power Point, archivos PDF y paquetes SCORM. (Motz et al., 2009)

En la figura 7 se representa el proceso que realiza Looking4LO para la extracción y los participantes de dicho proceso:



**Figura 7: Visión general del Sistema Looking4LO**

Fuente: (Motz et al., 2009)

El sistema Looking4LO recibe como entrada un conjunto de documentos no estructurados sobre los que realiza la extracción de Objetos Digitales de Aprendizaje (ODAs). Se basa en una representación ontológica del área temática por la que interesa anotar la información, y en la definición de un conjunto de Componentes Pedagógicos (ejercicio, definición, ejemplo, etc.) modelados a través de reglas. El resultado obtenido son ODAs extraídos de los documentos de acuerdo a la especificación de los Componentes Pedagógicos junto con un conjunto de metadatos que son generados de forma automática (Motz et al., 2014)

El prototipo funciona de la siguiente manera: primeramente en el modelo de dominio se define el tema que el usuario le interesa buscar luego de eso en el modelo pedagógico se trabaja con los datos que se necesita obtener del tema seleccionado., se selección la fuente que se utilizara para la extracción dicha extracción se realiza utilizando ontologías y por ultimo como resultado de dicha extracción se obtiene las ODAs que en dicho prototipo poseen 4 metadatos:

- Fuente(autor, fecha, etc)
- Generales (idioma)
- Específicos( Nivel de interactividad)
- Externos que son ingresados manualmente por el usuario del sistema (valor del ODA).

Según el prototipo un ODA tiene 4 conjuntos de metadatos que usa para la clasificación de la información dependiendo del origen:

- **Fuente:** información asociada al archivo tales como autor, fecha de creación, etc.
- **Generales:** esta información es generada automáticamente por el sistema referente al contenido del documento tal como el idioma.

- **Específicos:** son generados automáticamente y contiene información específica de un componente pedagógico.
- **Externos:** los encargados del sistemas son los que añaden información de forma manual, esta información pertenece a todos los ODAs generados durante el proceso de elección del sistema.

### 1.4.3 Extracción utilizando SCRAPY.

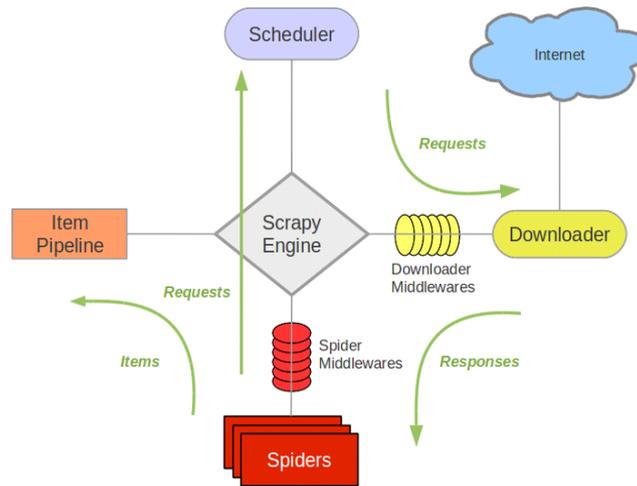
Scrapy es un marco de aplicación para el rastreo de sitios web y la extracción de datos estructurados que pueden ser utilizados para una amplia gama de aplicaciones útiles, como la minería de datos, procesamiento de la información o de archivo histórico (Scrapy, 2015)

Con scrapy se puede extraer datos utilizando muy pocas líneas de código y no importa si el archivo robots.txt (es un archivo de texto que dicta unas recomendaciones para que todos los crawlers también llamados buscadores y robots de buscadores cumplan (Hernández, Alvarado, & Chávez, 2011) está configurado para que las spiders de google no entren.

#### **Ventajas de usar Scrapy**

- Es muy simple de utilizar
- Se escribe las reglas de la extracción y scrapy se encarga de rastrear el sitio que se especifica.
- La respuesta de extracción de Scrapy es rápida y además fue diseñada para trabajar con más de 500 sitios completos diarios y todos en un mismo servidor.
- Proporciona varios mecanismos para conectar nuevas extensiones sin modificar el framework.
- Scrapy está bien documentado y cuenta con una comunidad amplia y apoyo comercial.

En la figura 8 se presenta una visión general de la arquitectura de Scrapy con sus componentes y un esquema de flujo de datos que tiene lugar dentro del sistema, lo que se muestra por la flechas de color verde.



**Figura 8: Arquitectura de Scrapy**

Fuente: (Scrapy, 2015)

A continuación se presenta una descripción de cada uno de los componentes de dicha arquitectura:

- **Scrapy Engine**

El motor es el encargado de controlar el flujo de datos entre los datos y eventos de activación al producirse ciertas acciones.

- **Scheduler**

Recibe del motor solicitudes y las encola, para posteriormente alimentar la arquitectura cuando sean solicitadas por el mismo motor.

- **Downloader**

Son los responsables de descargar las páginas web con las que se trabajara, la pasa al motor que es el encargado de asignar a los spiders para que ejecuten su trabajo.

- **Spiders**

Son clases personalizadas de acuerdo a las necesidades del usuario en las cuales se definen la forma de en qué se realiza la limpieza de un determinado sitio, incluyendo la forma de rastrear los enlaces y la forma de extraer los datos de una página estructurada

- **Item Pipeline**

Después de que un artículo es limpiado es enviado a ítem pipeline y analiza si el artículo puede seguir o le da de baja a dicho artículo.

Las acciones que realiza Item Pipeline son las siguientes:

- ✓ Comprobar si los artículos tiene todos los campos que se especificó en los spiders
  - ✓ Limpieza de datos
  - ✓ Eliminar duplicados
  - ✓ Almacenar en base de datos los elementos obtenidos.
- **Downloader middlewares**  
Son ganchos específicos que se encuentran entre el motor y el Downloader y se encargan de gestionar las solicitudes y respuestas entre estos componentes.
- **Spider middlewares**  
Son ganchos específicos que se encuentran entre el motor y los spiders y se encargan de gestionar las solicitudes y respuestas entre estos componentes.  
Flujo de datos durante el proceso de extracción:
    - ✓ El motor pide al Scheduler la URL que se rastreará.
    - ✓ El Scheduler devuelve la URL para rastrear al motor, y el motor la envía al Downloader, pasando por el Downloader Middleware (dirección de solicitud).
    - ✓ Una vez que la página termine la descarga genera una respuesta (con la página) y lo envía al motor, que pasa por el Downloader Middleware (dirección de respuesta).
    - ✓ El motor recibe la respuesta del Programa de descargas y lo envía al Spider para el procesamiento, pasando por la spider Middleware (dirección de entrada).
    - ✓ La spider procesa la respuesta y devuelve elementos y nuevas peticiones (a seguir) al motor.
    - ✓ El motor pasa los artículos limpiados y nuevas peticiones devueltas por una Spider a través de la Spider Middleware (dirección de salida) y, a continuación, envía los productos elaborados a los Item Pipeline y las solicitudes procesadas al Scheduler.

- ✓ El proceso se repite desde el paso (1) hasta que no haya más solicitudes del Scheduler.

#### 1.4.4 Comparativa de métodos de extracción

En la tabla 15 se muestra una tabla comparativa de los métodos de extracción mencionados anteriormente.

**Tabla 15: Comparativa de los métodos de extracción**

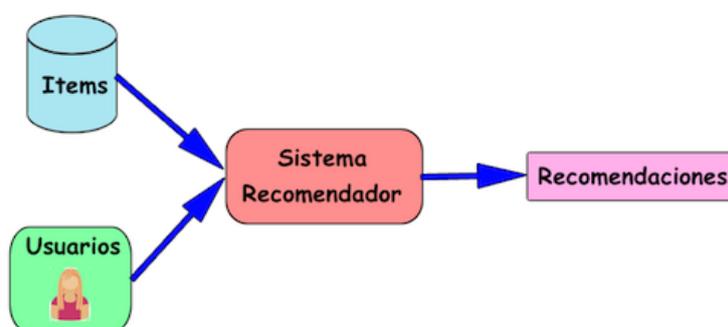
Sistema	Tipo de Archivo de entrada	Metadata extraída	Recursos para la extracción
SAFEX	<ul style="list-style-type: none"> <li>• HTML</li> <li>• XHTML</li> <li>• ASP</li> <li>• PHP</li> </ul>	SAXEF no sigue al estándar LOM, produce una tarjeta de identificación de E-learning (EIC) <ul style="list-style-type: none"> <li>• títulos principales</li> <li>• títulos secundarios</li> <li>• teórico o práctico</li> <li>• sintético o analítico</li> <li>• tipo multimedia y nivel multimedia</li> <li>• nivel de complejidad</li> <li>• enlaces a otras EICs con los mismos títulos</li> <li>• enlaces a otras tarjetas con títulos relacionados</li> </ul>	<ul style="list-style-type: none"> <li>• Stop Words</li> <li>• Direct Mapping rules</li> <li>• Heuristic Mapping</li> <li>• Rules</li> <li>• Statistical Measure</li> </ul>
Looking4LO	<ul style="list-style-type: none"> <li>• HTML</li> <li>• Archivos SCORM</li> <li>• Tipos de archivos no estructurados, tales como TXT, PDF y PPT</li> </ul>	<ul style="list-style-type: none"> <li>• Author (LOM 2.3.2)</li> <li>• Reading time</li> <li>• Has imagen</li> <li>• Interactivity level (LOM 5.3)</li> </ul>	<ul style="list-style-type: none"> <li>• Ontologías Tokenizer</li> <li>• Sentences Splitter POS Tagger</li> <li>• Gazetteer Transducer</li> <li>• Herramientas de procesamiento de lenguaje natural (GATE, General</li> <li>• Architecture for Text Engineering)</li> </ul>
Scrappy	<ul style="list-style-type: none"> <li>• HTML</li> <li>• XHTML</li> <li>• PHP</li> </ul>	Se puede extraer cualquier tipo de metadatos, depende de la estructura del usuario.	<ul style="list-style-type: none"> <li>• Item</li> <li>• Spiders</li> <li>• Pipelines</li> </ul>

Fuente: (Pinilla, Gutiérrez, & Ballejos, 2014) y el Autor

En base al estudio preliminar de cada uno de estos sistemas de extracción se pudo concluir que la mejor opción es Scrappy debido a la flexibilidad en sus tipos de archivos de entrada y en especial por qué se puede extraer información de diferentes estructuras de metadatos.

## 1.5 Sistemas Recomendadores

Un sistema de recomendación es un sistema inteligente que proporciona a los usuarios una serie de sugerencias personalizadas (recomendaciones) sobre un determinado tipo de elementos (ítems).(Moya, 2015). En la figura 9 se muestra el esquema del funcionamiento básico de un sistema recomendador.



**Figura 9: Esquema del funcionamiento básico de un sistema recomendador.**

Fuente: (Moya, 2015)

El sistema recomendador lo que hace es estudiar las características de los usuarios y mediante un análisis y procesamiento de datos brindar un conjunto de ítems que pueden resultar interesantes y de ayuda para los usuarios.

### 1.5.1 Características Generales

En el dominio de la educación existe gran cantidad y diversidad de material multimedia que puede ser utilizado en la enseñanza y que constituye una importante contribución al proceso enseñanza-aprendizaje. La mayoría de este material es viable a través de diversos repositorios de objetos de aprendizaje, donde cada objeto tiene metadatos descriptivos. Estos metadatos permiten recuperar aquellos objetos que satisfagan el tema de la consulta teniendo en cuenta sus características y preferencias. Existen características que contribuyen a identificarlos como:

- Ítems que se quiere recomendar: son los elementos que se recomendará al usuario, mismos elementos difieren mucho entre cada sistema.

- Usuarios para realizar las recomendaciones: información del perfil de los usuarios.
- Valoraciones de usuarios a los ítems: puede ser implícita (se evalúa los ítems de acuerdo al comportamiento al usuario) o explícita (el usuario da una valoración a cada ítem).
- Registro de información acerca de usuarios e ítems: dependiendo del sistema recomendador se puede tener almacenada información referente a usuarios e ítems para el proceso recomendador.

### 1.5.2 Técnicas algorítmicas utilizadas en los sistemas de recomendación

(Ricci, 2011) menciona que las técnicas algorítmicas utilizadas en sistemas de recomendación se clasifican en:

- Basadas en memoria: Las técnicas algorítmicas basadas en memoria utilizan la base de datos completa de usuarios-ítems para generar las predicciones.
- Basada en modelo: Estos algoritmos primero modelan los votos de los usuarios. Tratan el problema como una dificultad de predicción estadística y calculan el valor esperado para cada ítem en función de los votos anteriores.

### 1.5.3 Clasificación de los sistemas de recomendación basada en filtros

Según (Moya, 2015) los Sistemas Recomendadores la información que se encontraba en la red, es por esto que esta clasificación se encuentra basada en diferentes filtrados a continuación se describen cada uno de ellos:

- ✓ **Filtrado basado en contenido** (Salter & Antonopoulos, 2006): las recomendaciones que se realizan al usuario se basa en el conocimiento que se tiene de los ítems que el usuario ha valorado en el pasado.
- ✓ **Filtrado demográfico** (Krulwich, 1997): las recomendaciones que se realizan al usuario activo, se realizan comparando las valoraciones positivas de otros usuarios que comparten edad, sexo, situación geográfica, profesión, etc. con el usuario activo. En este tipo de filtrado se presupone que usuarios con características sociales similares comparten las preferencias sobre los ítems a recomendar
- ✓ **Filtrado colaborativo**: en una base de datos se encuentran almacenados datos de usuarios y de ítems, y también las valoraciones que usuarios dieron a un gran número de ítems., el filtrado colaborativo mediante los datos de usuarios identifica cuáles son similares al usuario activo, y

partiendo de ahí recomienda aquellos ítems que son valorados por el usuario activo y que tiene gran valoración por parte de usuarios similares. (Moya, 2015)

- ✓ **Métodos de filtrados híbridos:** utilizan la mezcla de algunos de los tres filtrados mencionados anteriormente por lo general siempre se combina el filtrado basado en contenidos o el filtrado demográfico. (Moya, 2015)

En la tabla 16 se presenta la comparativa entre los sistemas recomendadores mediante filtros.

**Tabla 16: Comparativa de los sistemas Recomendadores basados en filtros**

<b>Sistemas Recomendadores</b>	<b>Se basan</b>
Filtrado basado en contenido	Conocimiento de ítems mediante valoraciones pasadas
Filtrado demográfico	Valoraciones positivas de otros usuarios de similares características sociales
Filtrado colaborativo	Valoraciones realizadas por el usuario activo y que tengan mayor valoración por parte de usuarios similares.
Filtrados híbridos	Filtrado basado en contenido y el demográfico.

Fuente: El Autor

### **1.5.3 Sistemas de Recomendación basado en Latent Dirichlet Allocation (LDA).**

LDA es un modelo que genera tópicos sobre la base de la frecuencia de palabras a partir de un conjunto de documentos. LDA es particularmente útil para encontrar mezclas razonablemente exactas de los temas dentro de un conjunto determinado documento (Barber, 2014).

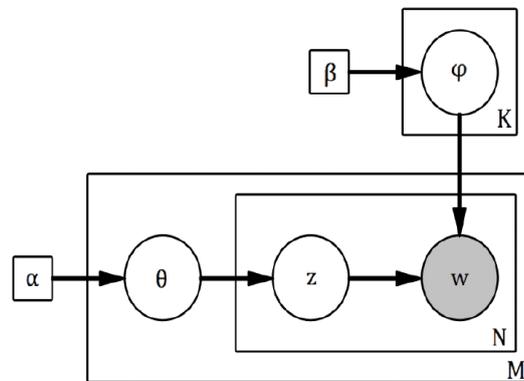
En el proceso de lenguaje natural, LDA es un modelo generativo que permite que conjuntos de observaciones puedan ser explicados por grupos no observados que explican por qué algunas partes de los datos son similares. Por ejemplo, si las observaciones son palabras en documentos, presupone que cada documento es una mezcla de un pequeño número de tópicos (también denominados como tópicos) y la aparición de cada palabra en un documento se debe a una de las categorías a las que

el documento pertenece. LDA es un ejemplo de modelo de categorías y fue presentado como un modelo en grafo para descubrir categorías por David Blei, Andrew Ng y Michael Jordan en 2002.

### 1.5.3.1 Funcionamiento y modelo de representación

En el LDA se obtienen dos matrices de probabilidad donde la primera distribución de matriz es la de que dado un tema salga una palabra y la segunda distribución de probabilidad es la de que un documento pertenezca a un tema (Moya, 2015).

En la figura 10 se muestra la representación gráfica de LDA y como se tienen en cuenta las distribuciones de probabilidad explicando su significado posteriormente:



**Figura 10: Representación gráfica de LDA**

Fuente: (Moya, 2015)

En esta imagen se representa las constantes como cuadrados y las distribuciones de probabilidad como círculos, donde:

- **K**: Número de tópicos.
- **N**: Número de palabras.
- **M**: Número de documentos
- **$\alpha$** : Parámetro de Dirichlet. Este parámetro es un vector de K componentes que describe el conocimiento a priori que se tiene sobre como los temas se distribuyen en los documentos.
- **$\beta$** : Parámetro de Dirichlet. Este parámetro es un vector de N componentes que describe el conocimiento a priori que se tiene sobre como las palabras se distribuyen en cada tema.
- **$\theta$** : Distribución de probabilidad de que un documento pertenezca a un tema.
- **Z**: Distribución de probabilidad de que una palabra pertenezca a un tema.
- **W**: Identifica todas las palabras en todos los documentos.
- **$\phi$** : Distribución de probabilidad de que dado un tema salga una palabra.

La idea central de LDA es obtener una proyección de los datos en un espacio de menor (o incluso igual) dimensión que los datos entrantes, con el fin de que la separabilidad de las clases sea la mayor posible. Es una técnica supervisada ya que para poder buscar esa proyección se debe entrenar el sistema con patrones etiquetados.

La distribución de probabilidad puede que salga de una palabra en un documento dado un tema. Dependiendo del caso la tabla de distribución varía de acuerdo a las coincidencias de la palabra, sin embargo cabe recalcar que cada uno de estas coincidencias puede o no coincidir con el tema en un 100%. A continuación se explica con un ejemplo.

### **Ejemplo 1:**

LDA fue aplicado a un conjunto de mensajes de correo electrónico de Sarah Palin, a continuación se indican algunos de los temas que el algoritmo aprendió:

- **Trig / Familia / La inspiración** : la familia, web, correo, dios, hijo, de, felicitaciones, niños, vida, niño, abajo, trig, bebé, nacimiento, amor, usted, síndrome, muy, especial, bendiga, edad, marido, años, gracias, mejor, ...
- **Vida Silvestre / BP Corrosión** : caza, pescado, alces, fauna, caza, osos, polar, oso, de subsistencia, administración, área, tablero, caza, los lobos, el control, el departamento, el año, el uso, el lobo, el hábitat, los cazadores, caribú, el programa de , Denby, pesca, ...
- **Energía / combustible / aceite / Minería**: energía, combustible, costos, aceite, Alaska, precios, costes, nome, ahora, altas, siendo, como local, público, energía, minas, crisis, los precios, los recursos, necesidad, de la comunidad, Fairbanks, rebaja, el uso, la minería, pueblos, ...
- **Gas** : gas, petróleo, tubería, Agia, proyecto, natural, al norte, los productores, las empresas, impuestos, la empresa, la energía, el desarrollo, la pendiente, la producción, los recursos, la línea, gasoducto, TransCanada, dijo, mil millones, el plan, la administración, millones, la industria, ...
- **Educación / Residuos** : escuela, residuos, educación, estudiantes, escuelas, millones, leer correo electrónico, mercado, política, estudiante, año, alto, noticias, estados, del programa, en primer lugar, informe, negocios, administración, anuncios, información, informes, 2008, trimestre, ...
- **Campaña Presidencial / Elecciones** : correo, web, de, agradezca, usted, cuadro, McCain, Sarah, muy, bueno, grande, juan, esperanza, presidente, sinceramente, wasilla, trabajo, mantener, hacer, añadir, familia, republicana, el apoyo , haciendo, por vía oral, ...

En un correo electrónico que cayó 99% en el / la categoría Trig Familia / Inspiración (palabras particularmente representativos están resaltados en azul):

Hello Governor Palin, Our **family** wanted to congratulate **you** and your **family** on the **birth** of your **son, Trig**. Our fourth **child**, Daniel, was **born** with **Down Syndrome**, and we can't imagine our **family** without him. Recently, I met a mom with a 34-year-old **daughter** with DS and she said it best: "Don't **you** feel like you've been chosen to be a member of a **very special** club?" **God** bless your **family**, what a **beautiful** example of **love** you are to all who see you! the Paul & Tricia Pietig **family**, Des Moines, Iowa

A continuación un extracto de un correo electrónico que cayó un 10% en la categoría Campaña Presidencial / Elecciones (en rojo) y el 90% en la categoría de la fauna / BP corrosión (en verde):

We understand that **you** have been discussed as a possible choice for the **Vice Presidency**.

As **people** who **support** the democratic process and care about protecting our **wildlife** for future generations, we want **you** to know that we don't believe **people** in our states would vote for **you** for any office if they knew your record on these issues.

It is troubling that **you** are **now** working to deny more than 50,000 Alaskans a vote on **aerial** killing of **wolves** and **bears** with legislation now **being** considered in the Alaska legislature.

### **1.5.3.2 Métricas de evolución de sistema recomendador en el contexto de LDA**

Obtenidos los resultados referentes a recomendaciones por parte de los usuarios, cotejar la calidad de los resultados obtenidos es de gran importancia. Con este propósito las medidas de calidad nos permiten comprobar si los mecanismos de recomendaciones empleados están funcionando de forma correcta.

De acuerdo a (Ortega Fernando, Bobadilla, & Antonio, 2012) entre las medidas de calidad más frecuentes tenemos:

- Medidas de calidad de las predicciones
  - Accuracy: mide el error cometido en las predicciones.

- Coverage: determina la capacidad de recomendación de los k vecinos.
- Medidas de calidad de las recomendaciones
  - Precisión: mide el porcentaje de predicciones acertadas.
  - Recall: representa el total de predicciones correctas que pueden realizarse.

A continuación profundizaremos en las medidas de calidad de las recomendaciones. La calidad de las recomendaciones realizadas puede medirse mediante dos medidas, que por lo general aparecen unidas que son la precisión y el recall,

### **1.5.3.2.1 Precisión**

Este concepto fue definido por Kent (Kent, Berry, Luehrs, & Perry, 1955), como factor de pertinencia. Hay otros autores que se refieren a él, como ratio de aceptación. Para Salton (Salton & McGill, 1983), la precisión es la proporción de material recuperado realmente relevante, del total de los documentos recuperados. La precisión según (Moya, 2015) busca comprobar cuantos de los ítems recomendados al usuario activo son realmente relevantes.

Siendo así la fórmula propuesta por (Salton & McGill, 1983):

$$Precision = \frac{Contenido\ Recomendado\ Correctamente\ (N_{rs})}{Contenido\ Total\ Recomendado\ (N_s)}$$

Esta medida está relacionada con dos conceptos, el de ruido y el de silencio informativo. De este modo, cuanto más se acerque el valor de la precisión a 0, mayor será el número de documentos recuperados que no le sirvan al usuario y por lo tanto el ruido que encontrará será mayor.

La salida obtenida en la recuperación es ordenada en función de la relevancia, por lo que los documentos más relevantes están al comienzo de la salida, de esta manera a medida que avanzamos en el número de documentos recuperados, la precisión decae.

### **1.5.3.2.2 Recall o exhaustividad**

Muchos autores, por influencia del término inglés la denominan "recall" o "rellamada". Es la proporción de material relevante recuperado, del total de los documentos que son relevantes en la base de datos, independientemente de que éstos, se recuperen o no. Esta medida es inversamente proporcional a la precisión. Fue formulada, al igual que la de precisión por Kent (Kent et al., 1955), con el nombre de factor de

exhaustividad. El recall de acuerdo a (Moya, 2015) se encarga de comprobar que porcentaje de los ítems relevantes para un usuario le fueron recomendados .

La fórmula propuesta por (Salton & McGill, 1983):

$$Recall = \frac{\text{Contenido Recomendado Correctamente } (N_{rs})}{\text{Contenido Relevante } (N_r)}$$

Si el resultado de este cálculo tiene como valor 1, tendremos la exhaustividad máxima, ya que hemos encontrado todo lo relevante que había en la base de datos, por lo tanto no tendremos ni ruido ni silencio informativo: la recuperación será perfecta.

Para alcanzar una exhaustividad alta, es necesario utilizar como índice términos generales de alta frecuencia, es decir, que aparezcan en muchos documentos de la colección. Para alcanzar una precisión alta, es necesario que los términos aparezcan con frecuencia alta, pero en pocos documentos y con nula en el resto

#### **1.5.4 Arquitectura de software**

En la actualidad a medida que el tamaño y la complejidad de los sistemas de software aumenta, el problema de diseño va más allá de los algoritmos y estructuras de datos de la computación: el diseño y la especificación de la estructura general del sistema surge como un nuevo tipo de problema. Los problemas estructurales incluyen la organización y la estructura bruta de control global; protocolos de comunicación, sincronización y acceso a los datos; asignación de funcionalidad a elementos de diseño; distribución física; composición de elementos de diseño; de escala y rendimiento; y la selección entre alternativas de diseño.

Cada vez es más claro que la ingeniería de software eficaz requiere instalación de software en el diseño arquitectónico, es por eso que a continuación se mencionan puntos de relevancia. En primer lugar, es importante ser capaz de reconocer los paradigmas comunes, de manera que las relaciones de alto nivel entre sistemas pueden ser comprendidas y para que los nuevos sistemas pueden ser construidos como variaciones de los sistemas antiguos.

En segundo lugar, conseguir la arquitectura adecuada es a menudo crucial para el éxito de un diseño de sistema de software; el incorrecto puede conducir a resultados desastrosos. En tercer lugar, la comprensión detallada de las arquitecturas de software permite al ingeniero para tomar decisiones basadas en principios entre las alternativas de diseño. En cuarto lugar, una representación arquitectónica del sistema es a menudo esencial para el análisis y la descripción de las propiedades de alto nivel de un sistema complejo.

#### **1.5.4.1 Cliente-Servidor**

“Es un modelo para el desarrollo de sistemas de información en el que las transacciones se dividen en procesos independientes que cooperan entre sí para intercambiar información, servicios o recursos. Se denomina cliente al proceso que inicia el diálogo o solicita los recursos y servidor al proceso que responde a las solicitudes. En este modelo las aplicaciones se dividen de forma que el servidor contiene la parte que debe ser compartida por varios usuarios, y en el cliente permanece sólo lo particular de cada usuario” (EcuRed, 2017).

#### **1.5.4.2 Basado en componentes de arquitectura**

Esta arquitectura se enfoca en la descomposición del diseño en componentes funcionales o lógicos que expongan interfaces de comunicación bien definidas. Esto provee un nivel de abstracción mayor que los principios de orientación por objetos y no se enfoca en asuntos específicos de los objetos como los protocolos de comunicación y la forma como se comparte el estado.(Peláez, 2017)

Según (Peláez, 2017) esta arquitectura tiene las siguientes características:

- Es un estilo de diseño para aplicaciones compuestas de componentes individuales.
- Pone énfasis en la descomposición del sistema en componentes lógicos o funcionales que tienen interfaces bien definidas.
- Define una aproximación de diseño que usa componentes discretos, los que se comunican a través de interfaces que contienen métodos, eventos y propiedades.

#### **1.5.4.3 Arquitectura basada en capas**

La arquitectura basada en capas se enfoca en la distribución de roles y responsabilidades de forma jerárquica proveyendo una forma muy efectiva de separación de responsabilidades. El rol indica el modo y tipo de interacción con otras capas, y la responsabilidad indica la funcionalidad que está siendo desarrollada. (Peláez, 2017).

El estilo de arquitectura basado en capas según (Peláez, 2017) se identifica por las siguientes características:

- Describe la descomposición de servicios de forma que la mayoría de la interacción ocurre solamente entre capas vecinas.
- Las capas de una aplicación pueden residir en la misma maquina física (misma capa) o puede estar distribuido sobre diferentes computadores (n-capas).
- Los componentes de cada capa se comunican con otros componentes en otras capas a través de interfaces muy bien definidas.
- Este modelo ha sido descrito como una “pirámide invertida de re-uso” donde cada capa agrega responsabilidad y abstracción a la capa directamente sobre ella.

#### **1.5.4.4 Arquitectura orientada a objetos**

Esta arquitectura incita a desarrolladores y programadores a pensar en términos de objetos, en vez de procedimientos, cuando planifican el código. Un objeto agrupa datos encapsulados y procedimientos para representar una entidad o clase.

El diseño orientado a objetos es la disciplina que define los objetos y sus interacciones para resolver un problema de negocio que fue identificado y documentado durante el análisis orientado a objetos (AOO).

#### **1.5.4.5 Arquitectura orientada a servicios**

En esta arquitectura los procesos del software se descomponen en servicios, que después se hacen disponibles y visibles en una red. Cada servicio provee funcionalidades para poder ser adecuado a las necesidades de la empresa, mientras esconde los detalles subyacentes de implementación. SOA aborda la complejidad, inflexibilidad y debilidades de los enfoques existentes en los diseños de procesos, flujos de trabajo e integración de aplicaciones.(Epicor, 2017)

#### **1.5.4.6 Arquitectura basada en micro servicios**

El estilo arquitectónico microService es un enfoque para el desarrollo de una sola aplicación como un conjunto de servicios pequeños, cada uno que se ejecuta en su propio proceso.(Lewis, 2014)

Para la implantación de una arquitectura de micro servicios hemos tener en cuenta 3 aspectos principalmente según (Sánchez, 2016):

- Un modelo de referencia en el que definir las necesidades de una arquitectura de micro servicios.
- Un modelo de implementación en el que decidiremos y concretaremos la implementación de los componentes vistos en el modelo de referencia.
- Un modelo de despliegue donde definir cómo se van a desplegar los distintos componentes de la arquitectura en los diferentes entornos.

En la tabla 17 se muestran una comparación de los estilos arquitectónicos comunes.

**Tabla 17: Comparativa entre las Arquitecturas**

<b>Estilo de Arquitectura</b>	<b>Descripción</b>
Cliente-Servidor	Segrega el sistema en dos aplicaciones, donde el cliente realiza solicitudes al servidor. En muchos casos, el servidor es una base de datos con la lógica de aplicación representado como procedimientos almacenados.
Basado en componentes de arquitectura	Se descompone el diseño de aplicaciones en componentes funcionales o lógicos reutilizables que exponen bien definidas las interfaces de comunicación.
Arquitectura en capas	Particiones las inquietudes especiales de la aplicación en grupos apilados (capas).
Orientado a objetos	Un paradigma de diseño basado en la división de responsabilidades para una aplicación o sistema en objetos reutilizables y autosuficientes individuales, cada uno que contiene los datos y el comportamiento pertinentes al objeto.
Arquitectura orientada a servicios (SOA)	Se refiere a las aplicaciones que exponen y consumen funcionalidad como un servicio mediante contratos y mensajes
Arquitectura basada en microservicios	Desarrollo de una aplicación como un conjunto de servicios pequeños.

Fuente: (Microsoft Corporation, 2009)

## 1.6 Resultados obtenidos

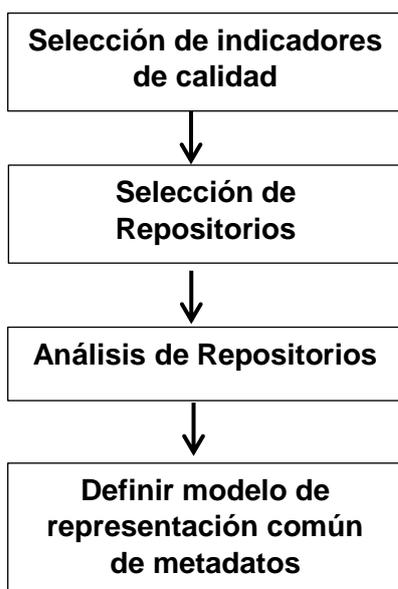
Tras el desarrollo del capítulo 1 se obtuvo los siguientes resultados:

- La situación actual de Repositorios de REA e identificar temas afines que nos apoyaran en el desarrollo del trabajo de titulación.

- Indicadores de calidad utilizados por diferentes autores para evaluar repositorios, lo que nos brinda una base para identificar indicadores que utilizaremos para la selección de repositorios para trabajar en el siguiente trabajo de titulación.
- Los estándares de metadatos utilizados por los repositorios para la representación de recursos.
- Identifico mecanismos utilizados en la extracción de información en sitios web.
- Tener una idea clara de lo que es un sistema recomendador y conocer su mecanismo de implementación.
- Obtuvo métricas que se emplean para la evaluación de sistemas recomendadores.

**CAPITULO 2:**  
**DESARROLLO DEL MODELO DE REPRESENTACIÓN**

En este capítulo se desarrolla el modelo de representación común, el cual se resume de una manera muy breve en la figura 11. En base al análisis de estudios planteados por diferentes autores se definen los diferentes indicadores de calidad, los cuales nos permiten la selección de repositorio entre varias opciones consideradas, posterior a esto se realiza un análisis de los repositorios seleccionados con el fin de identificar los metadatos con mayor incidencia dentro de cada repositorio para de esta manera definir el modelo de representación común que se desarrollará.



**Figura 11: Resumen capítulo 2**

## **2.1 Indicadores de calidad para valorar la importancia de los repositorios**

Determinar los indicadores de calidad para evaluar o medir el valor de un producto, servicio, actividad u objeto es necesario para saber si éste está funcionando de acuerdo a unos criterios preestablecidos o por el contrario se deben corregir errores. Se puede hablar de una clara tendencia o moda de evaluación en los sistemas bibliotecarios, quizás por el carácter tan crítico de los clientes/usuarios, quienes desean adquirir los mejores servicios. La mayoría de los repositorios, se han sometido alguna vez a un proceso de evaluación, con el objetivo de mostrar a sus usuarios que se encuentran ante un servicio de calidad. Además la introducción a nuevas herramientas para la gestión que puedan ser implementadas implica, la incorporación de nuevos productos, servicios electrónicos haciendo necesaria la actualización frecuente de este tipo de prácticas. En este estudio vamos a utilizar los indicadores de calidad como herramienta para evaluar repositorios institucionales en abierto. Éstos, entendidos como depósitos que contienen información científica generada por universidades o centros de investigación, a la que se puede acceder libre y

gratuitamente a través de la Web, se han convertido en un producto estrella dentro de las bibliotecas académicas. Casi todas las universidades cuentan con un archivo abierto que contiene la producción científica de sus investigadores y han sido las bibliotecas las que han tomado la iniciativa de liderar, gestionar y por lo tanto evaluar estos servicios. El proceso de evaluación ayudará, a sus responsables, a conocer la eficiencia y calidad de los mismos, además de detectar posibles fallos en cualquiera de los aspectos que lo forman. Nos fijaremos en indicadores de recursos, de procesos o de actividades, pero también en indicadores de impacto o de resultados, útiles para justificar la rentabilidad de los servicios y de percepción, que miden la opinión de los usuarios, teniendo en cuenta la dificultad que conlleva medir datos subjetivos como son las impresiones y sensaciones de las personas.

Una vez analizados y comparados los indicadores de calidad expuestos en los diferentes estudios mencionados anteriormente y de acuerdo a las necesidades que se tuvo para determinar la importancia de los repositorios de REA como son la calidad y confiabilidad de sus recursos educativos, se plasmaron los siguientes indicadores de calidad, mismos que se muestran en la tabla 18.

**Tabla 18: Criterios de calidad definidos para evaluar los repositorios**

Indicador de Calidad	Descripción
Evaluación de usuario	La valoración de los recursos albergados en los repositorios por parte de los usuarios.
Revisión por Pares	La evaluación de expertos en la temática correspondiente al recurso.
Autoría de Recursos	Los autores de los recursos alojados en los repositorios deben estar claramente identificados.
Sistemas de Metadatos	La utilización de modelos de datos o estándares de metadatos para la descripción de los recursos.
Soporte Multilingüe	Interfaces multilingües con las que cuenta el repositorio
Licencias de Derecho de Autor	Los RREA deben explicitar claramente las licencias abiertas (Creative Commons) en todos y cada uno de los recursos que albergan.

Fuente: El Autor

## **2.2 Análisis de repositorios en base a los indicadores de calidad establecidos.**

Con ayuda de los indicadores de calidad anteriormente establecidos se analizará a continuación una muestra de 9 repositorios ya previo a estudios realizados por (Santos-Hermosa, Ferrán-Ferrer, & Abadal Falgueras, 2015), (Moreno, Duque, & Pineda, 2012), (Atenas, 2014) , son considerados los más representativos dentro del

tema de REA, y de esta forma contamos con un punto de partida confiable al momento de elegir los repositorios con los que se desarrollará el siguiente trabajo de titulación.

### 2.2.1 MERLOT (Multimedia Educational Resources for Learning and Online Teaching)

Los recursos encontrados desde el sitio Web de MERLOT, están diseñados para apoyar, principalmente, a las comunidades de docentes y estudiantes de Educación Superior, quienes a su vez pueden también compartir sus materiales de aprendizaje y pedagogía. (Ministerio de Educación Nacional de Colombia, 2012)

Países participantes en este sitio son: Estados Unidos, Chile, Países de África pertenecientes a MAN (The MERLOT África Network), Etiopía, Israel, India

Merlot cuenta con 66.287 recursos libres y abiertos en línea y utiliza para el manejo de datos el estándar LOM (ROAP)

En la tabla 19 se presenta la forma que el Repositorio Merlot trabaja cada uno de los indicadores de calidad propuesta anteriormente, cabe mencionar que este análisis se lo realizó de forma visual directamente en el repositorio:

**Tabla 19: Análisis de Merlot en base a indicadores de calidad**

<b>Indicador de Calidad</b>	<b>Merlot (Multimedia Educational Resources for Learning and Online Teaching)</b>
<b>Evaluación de usuario</b>	<p>El Repositorio si cuenta con opción para evaluar al recurso por parte de los usuarios, misma que podrá acceder únicamente si se crea un usuario en el repositorio. Dicha evaluación consta de 4 partes:</p> <ol style="list-style-type: none"> <li>1. El voto (obligatorio) en el cual se puede calificar en base a 5 métricas que son: <ul style="list-style-type: none"> <li>• Excepcional (5 estrellas)</li> <li>• Recomendado ( 4 estrellas)</li> <li>• Está bien , pero no genial (3 estrellas)</li> <li>• Decepcionante (2 estrellas)</li> <li>• Pobre (1 estrella)</li> </ul> </li> <li>2. Se utiliza el curso que ofrece la opción de sí o no. (opcional)</li> <li>3. Escribir un comentario. (opcional)</li> <li>4. El tiempo dedicado a la revisión del sitio. (opcional)</li> </ol>
<b>Revisión por Pares</b>	<p>MERLOT utiliza el método de revisión por pares que se trata en que distintos evaluadores trabajen a la vez en la revisión de un material mismo que aumentará la posibilidad de encontrar y corregir errores, este procesos servirá para decidir si dicho material funcionara en el curso. Este método es realizado por catorce Editorial Bordas, que son facultades con</p>

	<p>experiencia en el uso de tecnologías en el ámbito de la enseñanza y la revisión de materiales utilizados en un sistema en línea. Esta revisión se la hace en base a 3 indicadores que son: <b>calidad del contenido, efectividad como herramienta de enseñanza y facilidad de uso</b>. Para cada uno de estos indicadores el revisor detalla los puntos fuertes y debilidades que observo durante la revisión. Una vez terminada la revisión se calificara al recurso bajo un estándar de 5 estrellas, misma que se obtiene mediante un promedio de la calificación de los 3 indicadores, siendo el intervalo de 3-5 tomado como aceptable, y menos de esto se toma como recurso no valido.</p>		
<b>Autoría de Recursos</b>	<p>Están claramente definidos los autores de todos los recursos que se presentan en este repositorio y utilizan la siguiente estructura: Nombre + Apellido o el nombre de la Institución desarrolladora del recurso. Y en el caso que sean más de dos autores se los separa mediante una coma (,).</p>		
<b>Sistema de metadatos</b>	<p>Este repositorio trabaja bajo el estándar LOM utilizando los siguientes Metadatos para describir todos sus recursos:</p>		
	Ciclo	Metadato	
		1 nivel	2 nivel
	General	Title	
		Description	
		Lenguaje (Idioma)	
	Life Cycle	Contribute	Identidad (Autor)
			Fecha( Fecha de añadido en Merlot)
	Technical	Format (Format Tecnico)	
		Requirement (Requerimientos Tecnicos)	
Educational	Learning resource type (tipo de material)		
	Context (Audiencia Primaria)		
Rights	Cost (Costo)		
	Copyright and others restrictions (copyright)		
<b>Soporte Multilingüe</b>	<p>El repositorio cuenta con una lista de 40 diferentes idiomas para que los usuarios puedan utilizar las herramientas de búsqueda y recuperación de contenidos</p>		

<b>Licencias de Derecho de Autor</b>	Para definir los derechos de autor utiliza las licencias Creative Commons
--	---

Fuente: El Autor

### 2.2.2 OER COMMONS (Open Educational Resources Commons)

OER Commons es un proyecto creado en el 2007 cuya misión es permitir a docentes de todos los niveles descubrir recursos educativos abiertos de calidad contrastada y permitirles colaborar en el uso, evaluación y mejora de los materiales (e-Learning, docs, 2013)

La mayoría de recursos son de educación superior, con gran concentración en las Áreas de ciencia y tecnología, matemáticas y estadística. (Centro de Innovación en Tecnología y Educación, 2014)

OER Commons es un proyecto de ISKME, un líder reconocido en la aplicación de investigación educativa a la práctica educativa.

Cuenta con 63245 recursos disponibles. (ISKME, 2014)

En la tabla 20 se presenta la forma que el Repositorio OER COMMONS trabaja cada uno de los indicadores de calidad propuesta anteriormente, cabe mencionar que este análisis se lo realizó de forma visual directamente en el repositorio:

**Tabla 20: Análisis de OER COMMONS en base a indicadores de calidad**

<b>Indicador de Calidad</b>	<b>OER COMMONS</b>
<b>Evaluación de usuario</b>	<p>El repositorio cuenta con una herramienta llamada Rubric que ayuda bajo los siguientes parámetros a la evaluación de los recursos por parte de los usuarios:</p> <ul style="list-style-type: none"> <li>• Calidad de la Explicación de la materia.</li> <li>• Utilidad de los materiales diseñados para apoyar la enseñanza</li> <li>• Calidad de Evaluaciones: Se aplica a los objetos diseñados para determinar lo que un estudiante sabe antes, durante, o después de que se enseña un tema. Cuando muchas evaluaciones se incluyen en un solo objeto, la rúbrica se aplica a todo el conjunto</li> <li>• Calidad de la interactividad tecnológica</li> <li>• Calidad del equipo y ejercicios de práctica</li> <li>• Las oportunidades de aprendizaje más profundo</li> </ul>
<b>Revisión por Pares</b>	Para evaluar el repositorio no cuenta con este indicador.

<b>Autoría de Recursos</b>	Están claramente definidos los autores de todos los recursos que se presentan en este repositorio y utilizan la siguiente estructura : Nombre + Apellido o el nombre de la Institución desarrolladora del recurso. Y en el caso que sean más de dos autores se los separa mediante una coma (,).	
<b>Sistema de metadatos</b>	Este repositorio trabaja bajo el estándar Dublin Core utilizando los siguientes Metadatos para describir todos sus recursos:	
	<b>Metadatos</b>	
	dc.title	Titulo
	dc.contributor.author	Autor
	dc.subject	Palabras claves
	dc.date.accessioned	Fecha de publicación
	dc.description	Descripción
	dc.identifier.uri	URI
	dc.Rights	Derechos reservados
	dc.language.iso	Idioma
dc.type	Tipo	
<b>Soporte Multilingüe</b>	El repositorio cuenta con el idioma Ingles para que los usuarios puedan utilizar las herramientas de búsqueda y recuperación de contenidos	
<b>Licencias de Derecho de Autor</b>	Todos los recursos cuentan dentro de su información un gráfico donde detalla el tipo de Licencia Creative Commons correspondiente.	

Fuente: El Autor

### 2.2.3 REPOSITORIO DIGITAL UTPL

Es un repositorio que contiene trabajos escritos de investigación de profesionales de formación de diferentes titulaciones y modalidades de estudio: Presencial y a Distancia pertenecientes a la Universidad Técnica Particular de Loja.

Cuenta con 14074 recursos disponibles y utiliza para el manejo de datos el estándar Dublin Core (utpl).

En la tabla 21 se presenta la forma que el Repositorio Digital UTPL trabaja cada uno de los indicadores de calidad propuesta anteriormente, cabe mencionar que este análisis se lo realizo de forma visual directamente en el repositorio:

**Tabla 21: Análisis de Repositorio Digital UTPL en base a indicadores de calidad**

Indicador de Calidad	Repositorio Digital UTPL
<b>Evaluación de usuario</b>	Para evaluar el repositorio no cuenta con este indicador.
<b>Revisión por Pares</b>	Para evaluar el repositorio no cuenta con este indicador.

<b>Autoría de Recursos</b>	Están claramente definidos los autores de todos los recursos que se presentan en este repositorio y utilizan la siguiente estructura: Nombre + Apellido o el nombre de la Institución desarrolladora del recurso. Y en el caso que sean más de dos autores se los separa mediante una coma (,).																				
<b>Sistema de metadatos</b>	<p>Este repositorio trabaja bajo el estándar Dublin Core utilizando los siguientes Metadatos para describir todos sus recursos:</p> <table border="1"> <thead> <tr> <th colspan="2">Metadatos</th> </tr> </thead> <tbody> <tr> <td>dc.title</td> <td>Título</td> </tr> <tr> <td>dc.contributor.author</td> <td>Autor</td> </tr> <tr> <td>dc.subject</td> <td>Palabras claves</td> </tr> <tr> <td>dc.date.accessioned</td> <td>Fecha de publicación</td> </tr> <tr> <td>dc.description</td> <td>Descripción</td> </tr> <tr> <td>dc.identifier.uri</td> <td>URI</td> </tr> <tr> <td>dc.Rights</td> <td>Derechos reservados</td> </tr> <tr> <td>dc.language.iso</td> <td>Idioma</td> </tr> <tr> <td>dc.type</td> <td>Tipo</td> </tr> </tbody> </table>	Metadatos		dc.title	Título	dc.contributor.author	Autor	dc.subject	Palabras claves	dc.date.accessioned	Fecha de publicación	dc.description	Descripción	dc.identifier.uri	URI	dc.Rights	Derechos reservados	dc.language.iso	Idioma	dc.type	Tipo
Metadatos																					
dc.title	Título																				
dc.contributor.author	Autor																				
dc.subject	Palabras claves																				
dc.date.accessioned	Fecha de publicación																				
dc.description	Descripción																				
dc.identifier.uri	URI																				
dc.Rights	Derechos reservados																				
dc.language.iso	Idioma																				
dc.type	Tipo																				
<b>Soporte Multilingüe</b>	El repositorio cuenta con los idiomas Inglés y Español para que los usuarios puedan utilizar las herramientas de búsqueda y recuperación de contenidos																				
<b>Licencias de Derecho de Autor</b>	Los ítems de DSpace están protegidos por copyright, con todos los derechos reservados.																				

Fuente: El Autor

#### 2.2.4 UNESCO

La UNESCO es el primer miembro de las Naciones Unidas en adoptar una política de acceso abierto en sus publicaciones. La nueva política, que entró en vigor el 13 de julio de 2013, permite descargar, traducir, adaptar, distribuir y compartir gratuitamente los datos y publicaciones de esta organización (Australian Research Council (ARC), 2013)

Dentro de la página principal web de la UNESCO se encuentra alojado un repositorio UNESDOC cuenta con una interfaz en 6 idiomas y contiene más de 146.000 documentos de la UNESCO en texto completo publicados desde 1945, así como los metadatos de las colecciones de la Biblioteca de la UNESCO y centros de documentación en las oficinas exteriores y los Institutos. Además UNESDOC utiliza para el manejo de datos el estándar LOM

Serán depositados en el repositorio de libre acceso:

- Inmediatamente después de la fecha oficial de publicación, todos los recursos publicados por la UNESCO desde el 31 de julio de 2013.
- Recursos publicados por la UNESCO antes del 31 de julio de 2013 y aquellos recursos que la UNESCO posee los derechos de todas las partes.
- Recursos publicados por los editores externos a partir del 31 julio de 2013.

En la tabla 22 se presenta la forma que el Repositorio UNESCO trabaja cada uno de los indicadores de calidad propuesta anteriormente, cabe mencionar que este análisis se lo realizo de forma visual directamente en el repositorio:

**Tabla 22: Análisis del Repositorio UNESDOC en base a indicadores de calidad**

Indicador de Calidad	UNESDOC	
<b>Evaluación de usuario</b>	Para evaluar el repositorio no cuenta con este indicador.	
<b>Revisión por Pares</b>	Para evaluar el repositorio no cuenta con este indicador.	
<b>Autoría de Recursos</b>	Están claramente definidos los autores de todos los recursos que se presentan en este repositorio y utilizan la siguiente estructura: Nombre + Apellido o el nombre de la Institución desarrolladora del recurso. Y en el caso que sean más de dos autores se los separa mediante una coma (,).	
<b>Sistema de metadatos</b>	Este repositorio trabaja bajo el estándar Dublin Core utilizando los siguientes Metadatos para describir todos sus recursos:	
	Metadatos	
	dc.title	Titulo
	dc.contributor.author	Autor
	dc.subject	Palabras claves
	dc.date.accessioned	Fecha de publicación
	dc.description	Descripción
	dc.identifier.uri	URI
<b>Soporte Multilingüe</b>	El repositorio cuenta con una lista de 12 diferentes idiomas para que los usuarios puedan utilizar las herramientas de búsqueda y recuperación de contenidos	
<b>Licencias de Derecho de Autor</b>	Para evaluar el repositorio no cuenta con este indicador.	

Fuente: El Autor

## 2.2.5 Repositorio de objetos de aprendizaje (ROAP)

En la tabla 23 se presenta la forma que el Repositorio ROAP trabaja cada uno de los indicadores de calidad propuesta anteriormente, cabe mencionar que este análisis se lo realizó de forma visual directamente en el repositorio:

**Tabla 23: Análisis de Repositorio ROAP en base a indicadores de calidad**

Indicador de Calidad	Roap: Repositorio de objetos de aprendizaje	
<b>Evaluación de usuario</b>	Para evaluar el repositorio no cuenta con este indicador.	
<b>Revisión por Pares</b>	Para evaluar el repositorio no cuenta con este indicador.	
<b>Autoría de Recursos</b>	Están claramente definidos los autores de todos los recursos que se presentan en este repositorio y utilizan la siguiente estructura: Nombre + Apellido o el nombre de la Institución desarrolladora del recurso. Y en el caso que sean más de dos autores se los separa mediante una coma (,).	
<b>Sistema de metadatos</b>	Este repositorio trabaja bajo el estándar Dublin Core utilizando los siguientes Metadatos para describir todos sus recursos:	
	Metadatos	
	dc.title	Titulo
	dc.contributor.author	Autor
	dc.subject	Palabras claves
	dc.date.accessioned	Fecha de publicación
	dc.description	Descripción
	dc.identifier.uri	URI
	dc.Rights	Derechos reservados
	dc.language.iso	Idioma
	dc.type	Tipo
	dc.format	Formato
<b>Soporte Multilingüe</b>	El repositorio cuenta con el idioma Español para que los usuarios puedan utilizar las herramientas de búsqueda y recuperación de contenidos	
<b>Licencias de Derecho de Autor</b>	Para evaluar el repositorio no cuenta con este indicador.	

Fuente: El Autor

## 2.2.6 LA FLOR (Federación Latinoamericana de Repositorios de Objetos de Aprendizaje)

El proyecto Repositorio Latinoamericano de Objetos de Aprendizaje fue creado en el 2010 cuyo objetivo es brindar acceso a profesores y estudiantes de América Latina a una mayor cantidad de objetos de aprendizaje generados en la región y fuera de ella a través de una federación abierta de repositorios latinoamericanos de objetos de aprendizaje implantada con base en estándares y software libre.

Según la Federación Latinoamericana de Repositorios de Objetos de Aprendizaje (FLOR, 2014) existen 56.858 recursos de aprendizaje cuyos principales proveedores corresponden a: SIDWeb (46684) sidweb\_agro (726)

Este repositorio utiliza el estándar Dublin Core para el manejo de metadatos.

En la tabla 24 se presenta la forma que el Repositorio LA FLOR trabaja cada uno de los indicadores de calidad propuesta anteriormente, cabe mencionar que este análisis se lo realizó de forma visual directamente en el repositorio:

**Tabla 24: Análisis de Repositorio La FLOR en base a indicadores de calidad**

Indicador de Calidad	La FLOR	
<b>Evaluación de usuario</b>	Para evaluar el repositorio no cuenta con este indicador.	
<b>Revisión por Pares</b>	Para evaluar el repositorio no cuenta con este indicador.	
<b>Autoría de Recursos</b>	Están claramente definidos los autores de todos los recursos que se presentan en este repositorio y utilizan la siguiente estructura: Nombre + Apellido o el nombre de la Institución desarrolladora del recurso. Y en el caso que sean más de dos autores se los separa mediante una coma (,).	
<b>Sistema de metadatos</b>	Este repositorio trabaja bajo el estándar Dublin Core utilizando los siguientes Metadatos para describir todos sus recursos:	
	Metadatos	
	dc.title	Título
	dc.contributor.author	Autor
	dc.subject	Palabras claves
	dc.date.accessioned	Fecha de publicación
	dc.description	Descripción
	dc.identifier.uri	URI
	dc.Rights	Derechos reservados
	dc.language.iso	Idioma
	dc.type	Tipo
	dc.format	Formato

<b>Soporte Multilingüe</b>	El repositorio cuenta con los idiomas Inglés y Español para que los usuarios puedan utilizar las herramientas de búsqueda y recuperación de contenidos
<b>Licencias de Derecho de Autor</b>	Para evaluar el repositorio no cuenta con este indicador.

Fuente: El Autor

### **2.2.7 FEB (Federación de Repositorios Educa Brasil)**

FEB es un servicio ofrecido por la RNE (Rede Nacional de En-sino e Pesquisa) de Brasil con el fin de tener disponible una infra-estructura para la confederación de repositorios de OAs a partir de un único punto de búsqueda, pregonando como principio guía la reutilización de estos recursos (FEB, 2010).

El sistema soporta principalmente el estándar OBAA, pero también permite el uso de cualquier otro como por ejemplo LOM o DublinCore siempre que exista previamente un esquema de mapeo con OBAA (Búsquedas desde una Federación).

FEB cuenta con un total de 16 repositorios:

- UFRGS - Universidad Federal de Río Grande do Sul
- LUME - Repositorio Digital de la Universidad Federal de Río Grande do Sul
- CESTA - Entidades de Apoyo Colección para el uso de la tecnología en el aprendizaje
- Obaa - Repositorio de Objetos de Aprendizaje Basado Agentes
- MEC - Ministerio de Educación
- BNDIGITAL - Digital de la Biblioteca Nacional de Brasil Fiocruz
- ARCA - Repositorio Institucional de la Fiocruz
- UFMA - Universidad Federal de Maranhão Tesis y Disertaciones Biblioteca Diario Portal Publicaciones Científicas
- NIIF - Instituto Federal de Río Grande do Sul Campus Repositorio Canoas
- IFSUL - Instituto Federal de Río Grande del Sur
- ENGEO - Repositorio de Objetos Educativos de Ingeniería Geotécnica
- UFSC - Universidad Federal de Santa Catarina Repositorio Institucional
- UFSCar - Universidad Federal de São Carlos Conocimiento Libre - Digital Repository Materiales de Enseñanza
- LACLO - Conferencia Latinoamericana de Objetos y Tecnologías de Aprendizaje Federación Laflor - Federación Latinoamericana de Repositorios de Objetos de Aprendizaje
- UNA-SUS - SUS Universidad Abierta
- ARES - Recogida y Recursos Educativos en Salud

- UFMT - Universidad Federal de Mato Grosso Tesis y Disertaciones Repositorio

En la tabla 25 se presenta la forma que el Repositorio FEB trabaja cada uno de los indicadores de calidad propuesta anteriormente, cabe mencionar que este análisis se lo realizó de forma visual directamente en el repositorio:

**Tabla 25: Análisis de Repositorio FEB en base a indicadores de calidad**

<b>Indicador de Calidad</b>	<b>FEB: Federación Educa Brazil</b>	
<b>Evaluación de usuario</b>	Para evaluar el repositorio no cuenta con este indicador.	
<b>Revisión por Pares</b>	Para evaluar el repositorio no cuenta con este indicador.	
<b>Autoría de Recursos</b>	Están claramente definidos los autores de todos los recursos que se presentan en este repositorio y utilizan la siguiente estructura: Nombre + Apellido o el nombre de la Institución desarrolladora del recurso. Y en el caso que sean más de dos autores se los separa mediante una coma (,).	
<b>Sistema de metadatos</b>	Este repositorio trabaja bajo el estándar Dublin Core utilizando los siguientes Metadatos para describir todos sus recursos:	
	Metadatos	
	dc.title	Titulo
	dc.contributor.author	Autor
	dc.subject	Palabras claves
	dc.date.accessioned	Fecha de publicación
	dc.description	Descripción
	dc.identifier.uri	URI
	dc.Rights	Derechos reservados
	dc.language.iso	Idioma
	dc.type	Tipo
	dc.format	Formato
<b>Soporte Multilingüe</b>	El repositorio cuenta con el idioma Portugués para que los usuarios puedan utilizar las herramientas de búsqueda y recuperación de contenidos	
<b>Licencias de Derecho de Autor</b>	Para evaluar el repositorio no cuenta con este indicador.	

Fuente: El Autor

### **2.2.8 PROCOMON**

Es un espacio destinado a usos educativos y de aprendizaje, fundamentalmente por parte de la comunidad docente y el alumnado, pero también por el público en general. Todos podrán buscar, consultar y descargar objetos de aprendizaje en distintos

formatos. Es una iniciativa del Ministerio de Educación, Cultura y Deporte (MECD) de España que se sustenta sobre el Proyecto Agrega, enriqueciéndolo con nuevos elementos. Se concreta en un espacio web que Integra objetos de aprendizaje y experiencias docentes constituyendo una Red inteligente, social y distribuida. Esta iniciativa cuenta con el repositorio de Recursos Educativos Abiertos (REA) creado por el MECD y las Comunidades Autónomas mismo en el que la comunidad educativa puede encontrar 84307 recursos de aprendizaje y crear material didáctico estructurado, clasificado de forma estandarizada (LOM-ES), preparado para su descarga y uso directo por el profesorado y el alumnado (Procomún, 2016) .

En la tabla 26 se presenta la forma que el Repositorio PROCOMUN trabaja cada uno de los indicadores de calidad propuesta anteriormente, cabe mencionar que este análisis se lo realizo de forma visual directamente en el repositorio:

**Tabla 26: Análisis de Repositorio PROCOMUN en base a indicadores de calidad**

Indicador de Calidad	PROCOMUN		
<b>Evaluación de usuario</b>	El Repositorio si cuenta con opción para evaluar al recurso por parte de los usuarios, misma que podrá acceder únicamente si se crea un usuario en el repositorio		
<b>Revisión por Pares</b>	Para evaluar el repositorio no cuenta con este indicador.		
<b>Autoría de Recursos</b>	Para evaluar el repositorio no cuenta con este indicador.		
<b>Sistema de metadatos</b>	Este repositorio trabaja bajo el estándar LOM utilizando los siguientes Metadatos para describir todos sus recursos:		
	Ciclo	Metadato	
		1 nivel	2 nivel
	General	Title	
		Description	
		Language (Idioma)	
	Life Cycle	Contribute	Identidad (Autor)
		Fecha( Fecha de añadido en Merlot)	
Technical	Format (Formato Tecnico)		
Educational	Learning resource type (tipo de material)		
	Context (Destinatario)		
<b>Soporte Multilingüe</b>	El repositorio cuenta con una lista de 7 diferentes idiomas para que los usuarios puedan utilizar las herramientas de búsqueda y recuperación de contenidos		
<b>Licencias de Derecho de Autor</b>	Para evaluar el repositorio no cuenta con este indicador.		

Fuente: El Autor

## 2.2.9 Comparación y selección de Repositorios

Para determinar que un repositorio sea de calidad es necesario un análisis minucioso en donde se clasifican aquellos que cumplan con la mayoría de estos indicadores de calidad establecidos previamente. Adicionalmente se procedió a establecer pesos con sus respectivas interpretaciones para cada indicador, con una escala que comprende de 3 a 1 como se observa en la siguiente tabla. En la interpretación de cada uno de los indicadores de calidad se establecen de acuerdo al porcentaje total y de acuerdo al número de escalas de manera que sea proporcional a la magnitud de las mismas y tenga coherencia entre ellas.

En la tabla 27 se presenta una escala de valoración que se implementará posteriormente en el análisis de los repositorios.

**Tabla 27: Escala de Valoración**

Indicador de Calidad	Escala		
	Importante	Aceptable	Considerable
	3	2	1
Evaluación de usuario	Posee una importancia del 100% contar con la opinión y referencias de otros usuarios para evaluar la veracidad del recurso, y es completamente confiable debido a la forma como el repositorio trabaja este método de evaluación.	Es aceptable pero depende mucho de la forma con la que manejan cada repositorio. Referente a usuarios, o miembros del repositorio.	De acuerdo a su poca incidencia en los repositorios no es considerado importante.
Autoría de Recursos	La autoría de los recursos sea confiable y se encuentre descrita completamente	La autoría de los recursos no este descrita correctamente	No se encuentra identificada la autoría del recurso.
Sistema de Metadatos	Utiliza un estándar para la representación de metadatos de los recursos educativos abiertos.	Utiliza un estándar y metadatos no estructurados bajo ningún estándar conjuntamente para la representación de metadatos de los recursos educativos abiertos.	No utiliza un estándar para la representación de metadatos de los recursos educativos abiertos.
Soporte Multilingüe	El repositorio sea adaptable a varios idiomas	El repositorio sea adaptable a por lo menos 2 idiomas al de origen y al inglés como idioma global.	El repositorio sea adaptable a un solo idioma
Licencias de Derecho de Autor	Consta de licencias abiertas creative commons	Utilizan otro tipo de licencias diferentes a creative commons	No consta de licencias abiertas

Indicador de Calidad	Escala	
	Si	No
	3	1
Revisión por Pares	Por qué es indispensable contar con la opinión y referencias de personas expertas en el tema para evaluar la veracidad e importancia del recurso.	De acuerdo a su poca incidencia en los repositorios no es considerado importante.

Fuente: El autor

En la siguiente tabla 28 se puede observar los pesos que se establecieron para cada uno de los repositorios en base a escala de valoración

**Tabla 28: Incidencias de indicadores de calidad en repositorios**

Indicadores	Repositorios								
	MERLOT	OER COMMONS	UTPL	UNESCO	TEMOA	LA FLOR	ROAP	REB	PROCOMUN
Evaluación de usuario	3	3	1	1	1	1	1	1	2
Revisión por Pares	3	1	1	1	1	1	1	1	1
Autoría de Recursos	3	3	3	2	1	2	2	1	1
Sistema de Metadatos	3	3	3	3	2	2	2	2	2
Soporte Multilingüe	3	1	1	2	1	1	1	1	2
Licencias de Derecho de Autor	3	3	1	1	1	1	1	1	1

Fuente: Autor

De acuerdo al estudio realizado y la comparación de los repositorios con los indicadores de calidad antes establecidos, se procedió a la selección de los repositorios de recursos educativos abiertos con los que se trabajara. Esta selección se la realizó tomando en cuenta a los repositorios que más indicadores cumplen y además de contar con el mayor peso, a continuación en la tabla 29 se presenta la lista de repositorios:

**Tabla 29: Selección de los repositorios de acuerdo al peso**

Repositorio	Peso
MERLOT	18
OER COMMONS	14
Repositorio Digital UTPL	10
Repositorio de la UNESCO	10
TEMOA	7

LA FLOR	8
ROAP	8
REB	7
PROCOMUN	9

Fuente: El autor

## 2.3 Análisis de Repositorios seleccionados

Se definirá un modelo de datos basándose en los metadatos más representativos de cada repositorio, por tal razón se procede primeramente en analizar la estructura de los metadatos de los REA de cada repositorio.

### 2.3.1 Metadatos MERLOT

En la figura 12 se presenta la estructura del repositorio MERLOT con sus metadatos bien definidos y localizados, recordando que este repositorio utiliza el estándar LOM.

The image shows a screenshot of a MERLOT resource page for 'ADN DESDE EL PRINCIPIO'. On the left, LOM metadata tags are listed with arrows pointing to their corresponding fields on the page:

- `<lom:title>...</lom:title>` points to the title 'ADN DESDE EL PRINCIPIO'.
- `<lom:description >...</lom:description>` points to the main description text.
- `<lom:format>...</lom:format>` points to 'Tipo de material: Simulación'.
- `<lom:date>...</lom:date>` points to 'Fecha de agregación a MERLOT: 11 de abril de, el año 2000'.
- `<lom:contribute>...</lom:contribute>` points to 'Autor: Laboratorio Cold Spring Harbor'.
- `<lom:keywords> ...</lom:keywords>` points to 'Palabras clave: la genética, el ADN, la historia sobre el ADN, genes, modelos de ADN'.
- `<lom:language>...</lom:language>` points to 'Idioma: Inglés'.

The resource page itself is divided into several sections:

- ADN DESDE EL PRINCIPIO** (Title)
- Calidad** (Quality): Includes 'Peer Reviews: ★★★★★', 'Calificación del usuario: ★★★★★', 'Discusión (53 comentarios)', 'Ejercicios de aprendizaje (11)', 'Bookmark Colecciones (422)', 'ePortfolios de oaiif (26)', and 'Accesibilidad Información (ninguno)'.
- Descripción**: 'ADN desde el principio es un tutorial animado de ADN, los genes y la herencia. La ciencia detrás de cada concepto se explica el uso de animaciones relacionadas con temas de ADN, una galería de imágenes, entrevistas en video, problemas, biografías y enlaces relacionados con el ADN. Hay tres secciones, Classical Genetics, las moléculas de Genética y organización del material genético. Sus características fundamentales son las explicaciones claras de los experimentos clásicos y las excelentes fotografías de los investigadores y su información labs.For y ... Más'.
- Acciones**: 'Ir a los materiales', 'Hablar de este material'.
- Metadatos**: 'Tipo de material: Simulación', 'Formato técnico: Exce', 'Fecha de agregación a MERLOT: 11 de abril de, el año 2000', 'Fecha de modificación en MERLOT: 08 de julio de, el año 2016', 'Autor: Laboratorio Cold Spring Harbor', 'Remitente: Jeff Campana', 'Palabras clave: la genética, el ADN, la historia sobre el ADN, genes, modelos de ADN'.
- Acciones**: 'Enlace roto', 'Reportar como inapropiado'.
- ACERCA DE** (About): 'Audiencia principal: Colegio de Educación General', 'Compatibilidad móvil: no hay información en este momento', 'Requerimientos técnicos: Flash 3 y RealAudio', 'Idioma: Inglés', 'Coste de cada una: no', 'Código fuente disponible: no', 'Información de acceso disponible: no', 'Copyright: Sí', 'Creative Commons: no hay', 'Explorar en categorías: Ciencia y Tecnología / Agricultura y Ciencias del Ambiente / Ciencia Animal, Ciencias Sociales / Psicología / Biológica, Ciencias Sociales / Psicología / Biológica / Evolución y Comportamiento, Ciencia y Tecnología / Ciencias de la Comunicación y enfermedades / Fundamentos y Afines, Ciencia y Tecnología / Biología / Genética, Ciencia y Tecnología / Biología / Biología Molecular, Ciencia y Tecnología / Agricultura y Ciencias del Ambiente / Ciencias de las Plantas'.
- CONEXIONES** (Connections): Social media sharing options (Like, Tweet, G+, PinIt, etc.) and a QR code.

Figura 12: Estructura de los metadatos delo recursos en el Repositorio Merlot

Fuente: El Autor

### 2.3.2 Metadatos OER COMMONS

En el figura 13 se presenta la estructura del repositorio OER COMMONS con sus metadatos bien definidos y localizados, recordando que este repositorio utiliza el estándar DUBLIN CORE.

The screenshot shows a resource page with the following elements:

- Title:** ".0001 The Value of A Unit With Four Cyphers Going Before It" (indicated by a red arrow and the tag `<dc:title>...</dc:title>`)
- Identifier:** ".00001" (indicated by a brown arrow and the tag `<dc:identifier>...</dc:identifier>`)
- Creator:** "Library of Congress" (indicated by a green arrow and the tag `<dc:creator>...</dc:creator>`)
- Language:** "English" (indicated by a blue arrow and the tag `<dc:language>...</dc:language>`)
- Format:** "Downloadable docs, Graphics/Photos, Text/HTML" (indicated by a purple arrow and the tag `<dc:format>...</dc:format>`)
- Description:** "Abstract. A satire on dissension and political intrigue within Andrew Jackson's administration..." (indicated by a yellow arrow and the tag `<dc:description>...</dc:description>`)
- Subject:** "1831 Political Cartoons" (indicated by an orange arrow and the tag `<dc:subject>...</dc:subject>`)

Figura 13: Estructura de los metadatos de los recursos en el Repositorio OER COMMONS

Fuente: El Autor

### 2.3.3 Metadatos UNESCO

En el figura 14 se presenta la estructura del repositorio UNESCO con sus metadatos bien definidos y localizados, recordando que este repositorio utiliza el estándar DUBLIN CORE.



UNESDOC

1/14927

<b>Title:</b>	Ensuring adequate, efficient & equitable financing in schools: school finance in the Asia-Pacific	<dc:title>...</dc:title>
<b>Series:</b>	UNESCO Asia-Pacific education thematic brief	<dc:identifier>...</dc:identifier>
<b>Corporate author:</b>	UNESCO Office Bangkok and Regional Bureau for Education in Asia and the Pacific	<dc:creator>...</dc:creator>
<b>Publ Year:</b>	2017	
<b>Original Language:</b>	English	<dc:language>...</dc:language>
<b>Document code:</b>	THA/DOC/IQE/17/01-E	
<b>General notes:</b>	(Electronic version only)	
<b>On line:</b>	Yes	
<b>Licence Type:</b>	CC BY SA	
<b>Document Type:</b>	UNESCO document	
<b>Catalog Number:</b>	246922	

Figura 14: Estructura de los metadatos de los recursos en el Repositorio UNESDOC

Fuente: El Autor

### 2.3.4 Metadatos REPOSITORIO DIGITAL UTPL

En el figura 15 se presenta la estructura del repositorio REPOSITORIO DIGITAL UTPL con sus metadatos bien definidos y localizados, recordando que este repositorio utiliza el estándar DUBLIN CORE.

<dc:title>...</dc:title>	→	Title:	A case study about strategies used to teach a student with Asperger syndrome
<dc:creator>...</dc:creator>	→	Authors:	Yaguachi Torres, Mariela Elizabeth
<dc:subject>...</dc:subject>	→	metadatas.dc.contributor.advisor:	Camacho Minucha, Gina Karina
<dc:date>...</dc:date>	→	Keywords:	Educación especial Escuelas - Síndrome de asperger Dificultades de aprendizaje Licenciado en ciencias de la educación mención inglés - Tesis y disertaciones académicas
<dc:description >...</dc.description>	→	Issue Date:	2016
<dc:identifier>...</dc:identifier>	→	Citation:	guachi Torres, Mariela Elizabeth. (2016). A case study about strategies used to teach a student with Asperger syndrome. (Trabajo de Titulación de Licenciado en Ciencias de la Educación Mención Inglés). UTPL, Loja
<dc:language>...</dc:language>	→	Abstract:	This study was carried out to identify the strategies used to teach a student with Asperger syndrome. In order to explore the phenomenon before mentioned, the English teacher and the student with Asperger syndrome were taken into account as sample at Taggeschule y San Gerardo High School in Loja city. The intention of this case study was precisely to identify the strategies that the English teacher uses to teach the student with Asperger syndrome and the involvement of this student when using them. The methods used for collecting data were qualitative and quantitative. The qualitative method allowed identifying the appropriate strategies to teach the student with Asperger syndrome and the quantitative method allowed analyzing the numeric data satisfactorily. At the same way, the data collection was done satisfactorily through semi-structured interviews and observation sheets that allowed identifying the student's involvement in the English classes. Hence, the present work points out details and difficulties that not only the student with Asperger syndrome presents but also the teacher faces in the teaching-learning process
		Description:	Este estudio fue llevado a cabo para identificar las estrategias a enseñar a un estudiante con síndrome de Asperger. Con el fin de explorar el fenómeno antes mencionado, el docente de inglés y el estudiante con síndrome de Asperger fueron tomados como muestra en el colegio Taggeschule y San Gerardo de la ciudad de Loja. La intención de este caso de estudio fue precisamente para identificar las estrategias que el docente de inglés usa para enseñar al estudiante con síndrome de Asperger y el involucramiento de este estudiante cuando usan ellas. Los métodos usados para recolectar datos fueron cualitativos y cuantitativos. El método cualitativo permitió analizar satisfactoriamente las estrategias apropiadas para enseñar al estudiante con síndrome de Asperger y el método cuantitativo permitió analizar satisfactoriamente los datos numéricos. La recolección de datos fue realizada a través de entrevistas semi-estructuradas y láminas de observación. Así, el presente trabajo señala detalles y dificultades que no solamente el estudiante con síndrome de Asperger presenta sino también el docente enfrenta en el proceso de enseñanza aprendizaje
		metadatas.dc.identifier.other:	1173524
		URI:	http://dspace.utpl.edu.ec/handle/123456789/14567
		metadatas.dc.language:	spa
		metadatas.dc.type:	bachelorThesis
		Appears in Collections:	Licenciado en Ciencias de la Educación mención inglés

File	Description	Size	Format
Yaguachi Torres Mariela.pdf		2.03 MB	Adobe PDF

<lom:format>...</lom:format>

Figura 15: Estructura de los metadatos de los recursos en el Repositorio Digital UTPL

Fuente: El Autor

## 2.4 Análisis de metadatos en los repositorios

Una vez realizado el análisis y selección de los repositorios con los que se va a trabajar se procedió a determinar los metadatos con los que trabaja cada repositorio para la representación de sus recursos y se obtuvo los siguientes resultados:

### 2.4.1 Repositorio MERLOT

El repositorio de la MERLOT cuenta con 66.287 recursos mismos que trabajan con los siguientes metadatos basados en el estándar LOM. En la tabla 30 se muestran dichos valores

**Tabla 30: Incidencia de metadatos en Repositorio Merlot**

Metadato	Número de recursos que usan el metadato
Tema	66.287
Url	66.287
Descripción	66.287
Tipo de Material	66.287
Fecha de ingreso	66.287
Fecha de modificación	66.287
Author	66.287
Palabras Claves	15000
Lenguaje	66.287

Fuente: El Autor

MERLOT cuenta con la Calificación del Usuario para cada uno de los recursos, y debido a la importancia se lo usaria como un metadato adicional para el presente estudio. Son 66.287 recursos que cuentan con dicho metadato.

### 2.4.2 Repositorio UNESDOC

El repositorio de la UNESCO cuenta con 13147 recursos mismos que trabajan con los siguientes metadatos basados en el estándar Dublín Core. En la tabla 31 se muestran dichos valores

**Tabla 31: Incidencia de metadatos en Repositorio UNESDOC**

Metadato	Numero de recursos que usan el metadato
Tema	13147
url	9739
Document Type	13147
Año de Publicación	13147
Author	6266
Palabras claves	11121
Original Language	13147

Fuente: El Autor

### 2.4.3 Repositorio OER COMMONS

El repositorio Oer Commons cuenta con 63911 recursos mismos que trabajan con los siguientes metadatos basados en el estándar Dublin Core. En la tabla 32 se muestran dichos valores

**Tabla 32: Incidencia de metadatos en Repositorio OER COMMONS**

OER Commons	Número de recursos que usan el metadato
Titulo	63911
url	63911
Abstract	63911
Material Type	63911
Author	24286
Gags	63911
Lenguaje	63911
Condiciones de uso	63911
Media Format	63911
Rating	63911

Fuente: El Autor

OER COMMONS cuenta con Rating para cada uno de los recursos, y debido a la importancia se lo usaria como un metadato adicional para el presente estudio. Son 63.911 recursos que cuentan con dicho metadato.

#### 2.4.4 Repositorio Digital UTPL

El repositorio UTPL cuenta con 14074 recursos mismos que trabajan con los siguientes metadatos basados en el estándar Dublin Core. En la tabla 33 se muestran dichos valores

**Tabla 33: Incidencia de metadatos en Repositorio Digital UTPL**

UTPL	Número de recursos que usan el metadato
Titulo	14074
url	14074
Resumen	14074
Fecha de publicación	14074
Autor	14074
Palabras Claves	14074
Formato	14074
Idioma	4504

Fuente: El Autor

#### 2.5 Descripción del modelo común de metadatos de representación de recursos.

A través del estudio previo de las diferentes incidencias realizadas en los diferentes metadatos para cada uno de los repositorios analizados, en la figura 16 se representa los porcentajes globales de los metadatos que se utilizan en todos los repositorios para la representación.

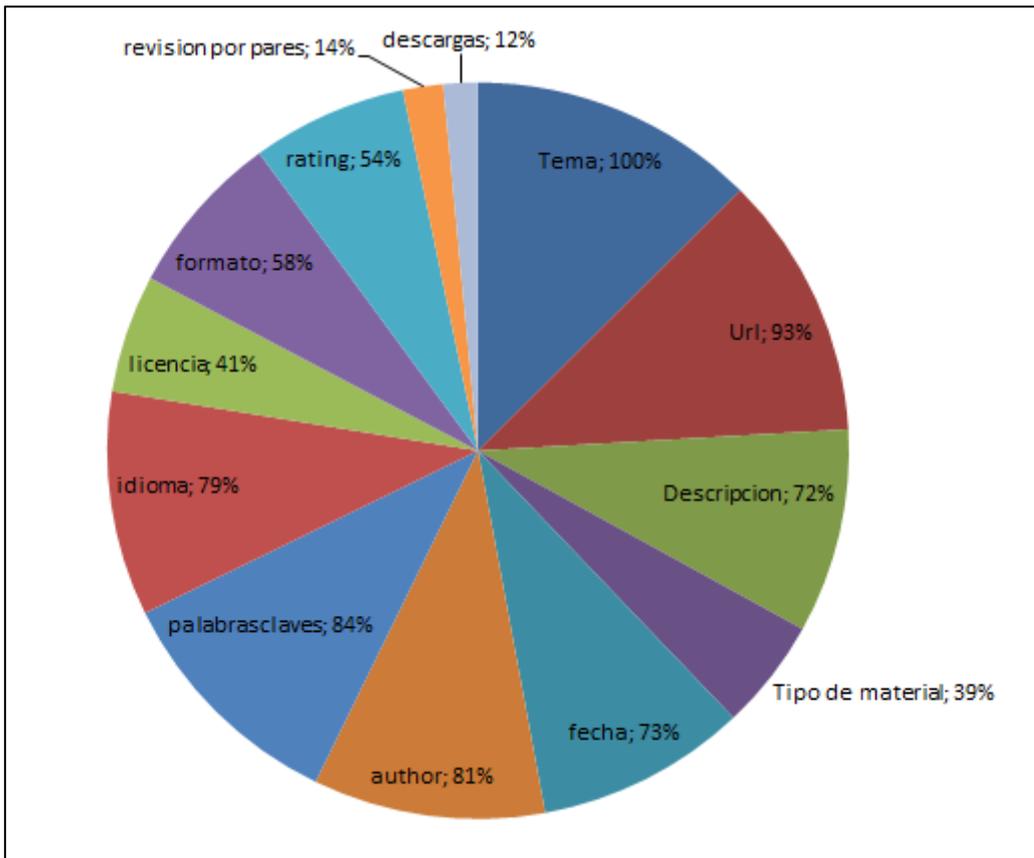


Figura 16: Porcentajes de Metadatos en Repositorios

Fuente: El Autor

En la figura 17 se presenta un gráfico explicativo del proceso que se siguió para obtener el modelo de datos.

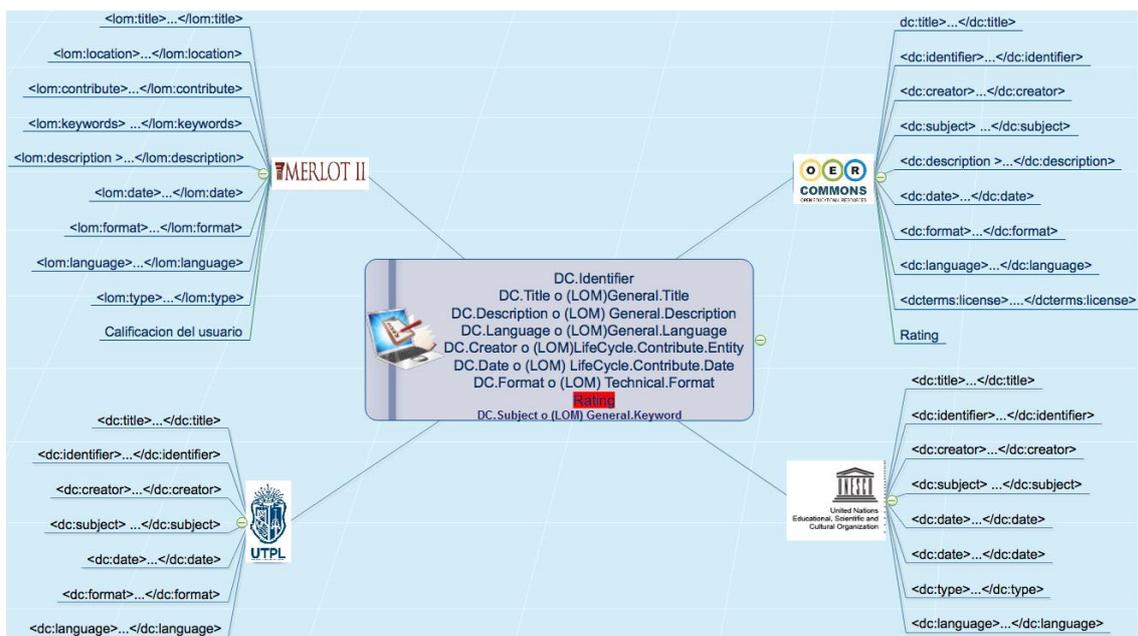


Figura 17: Modelo metadatos para la representación común de los Recursos Educativos Abiertos

Fuente: El Autor

Como se observa en la figura se trabajó con 4 repositorios mismos que empleaban para la representación de sus recursos diferentes metadatos de distintos estándares., los metadatos con mayor incidencia en los repositorios fueron utilizados como base para el modelo de datos para la representación común de recursos educativos abiertos de los distintos repositorios.

En la tabla 34 se define la estructura del modelo de datos y la descripción de cada metadato a utilizar:

**Tabla 34: Modelo de metadatos para la representación común de los Recursos Educativos Abiertos**

Estándar	Metadato	Tipo de Dato	Descripción	Observación
<b>DC.Identifier</b>	URL	String	Es una referencia inequívoca para el recurso dentro un contexto dado. Por ejemplo, los URI (URL y URN),	Existen recursos que no tiene valor en este metadato, por lo que el campo se llenara con "Recurso no disponible en Línea"
<b>DC.Title o (LOM)General.Title</b>	Título	String	Título del recurso	
<b>DC.Description o (LOM)General.Description</b>	Descripción	String	Es un breve resumen que describe el recursos	El recurso que no contenga una descripción se llenara este campo con la palabra NULL
<b>DC.Language o (LOM)General.Language</b>	Idioma	String	Idioma o idiomas que se encuentra disponible dicho recurso	Si existen más de 2 idiomas se los separara con ;
<b>DC.Creator o (LOM)LifeCycle.Contribute.Entity</b>	Autor	String	Introduce información acerca de un contribuyente a la producción del material	Se colocara el apellido seguido con el nombre del autor., Si existen más de 2 autores se los separara con
<b>DC.Date o (LOM)LifeCycle.Contribute.Date</b>	Fecha	String	Es el año de publicación del recurso	Solo se colocara el año de publicación
<b>DC.Format o (LOM)Technical.Format</b>	Formato	String	Tipo de dato del recurso, y debe de ser utilizado para conocer el software necesario para acceder a este recurso	Si existen más de 2 formatos se las separara con ,

Estándar	Metadato	Tipo de Dato	Descripción	Observación
<b>DC.Subject o (LOM) General.Keyword</b>	Palabras Claves	String	Una palabra clave o frase que describe el tema principal del objeto educativo.	Si existen más de 2 palabras claves se las separara con ;
<b>Rating</b>	Calificacion de Usuario	Int	Valoración de parte del usuario al recurso.	Si existen número decimales se procederá redondear. Solo acepta números enteros.

Fuente: El Autor

**CAPITULO 3:**  
**IMPLEMENTACION DEL MODELO DE REPRESENTACION COMÚN**

Este capítulo se centra en la implementación de la herramienta basada en el modelo de representación previamente propuesto, en donde se realiza la extracción de la metadatos de los REA de los repositorios seleccionados, esta información se almacenara en una base de datos que está vinculada a una aplicativo web, en donde se visualizará dicha información. En la figura 18 se presenta la estructura correspondiente al presente capitulo.

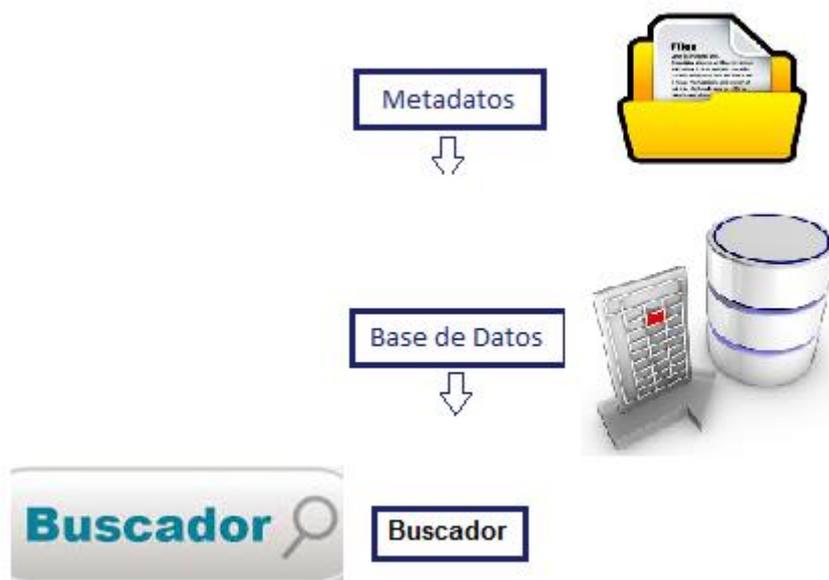
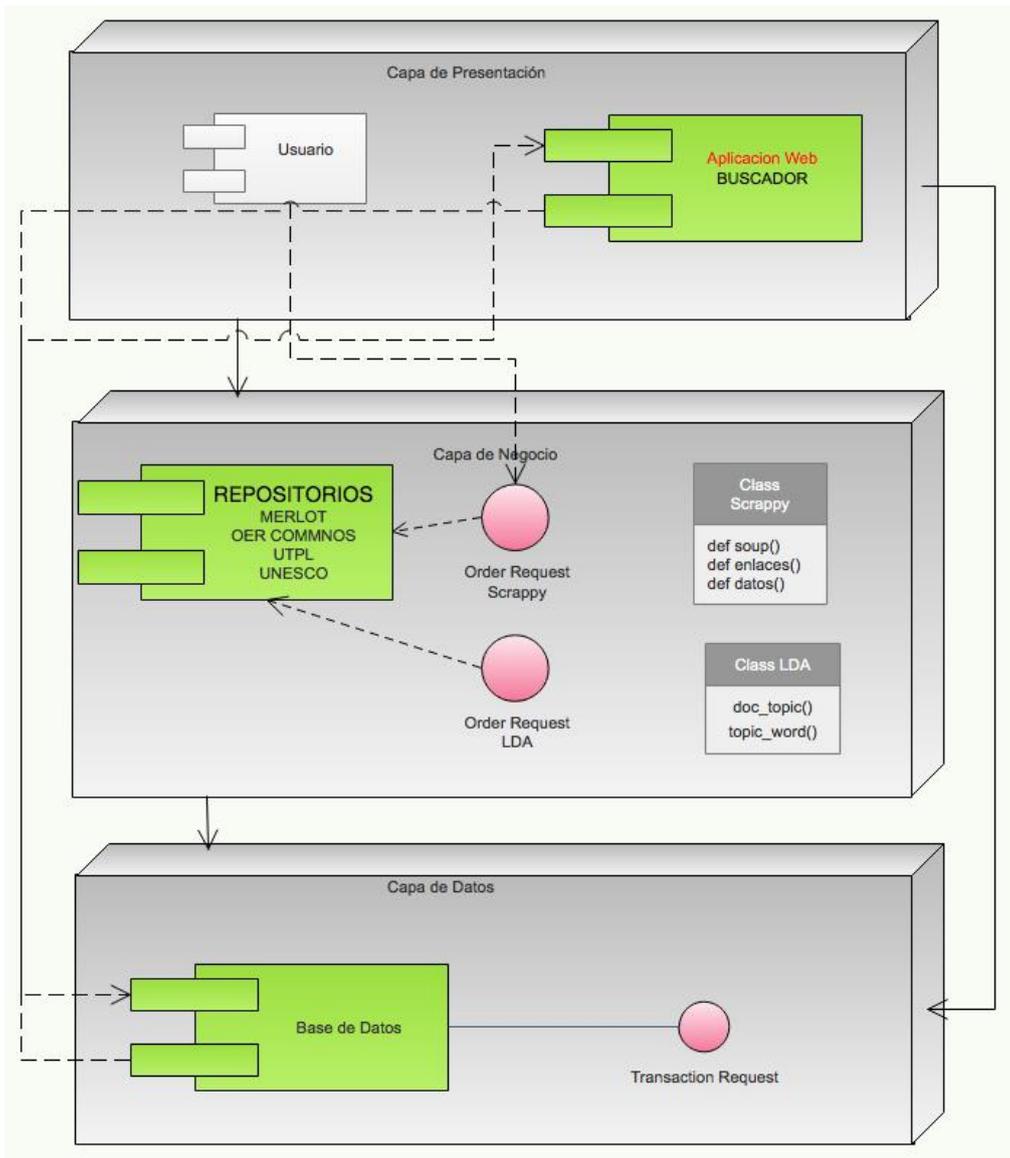


Figura 18: Resumen capítulo tres

### 3.1 Diseño

#### 3.1.1 Arquitectura tres capas

La arquitectura de tres capas encierra los factores estéticos y dinámicos más importantes de software que se desea crear, es decir la descripción de los subsistemas y los componentes que se utilizan en el sistema y las relaciones existentes entre ellos. De igual forma, la arquitectura de software a tres capas incluye todos estos aspectos antes mencionados y además ofrece mejores opciones para proyectos informáticos de gran alcance y complejidad. En la figura 19 se presenta la arquitectura tres capas del aplicativo correspondiente al presente trabajo de titulación.



**Figura 19: Arquitectura tres capas**

Fuente: Autor

- **CAPA DE PRESENTACION**  
Para la capa de presentación en nuestro caso es sencilla en donde su interfaz es un buscador, el cual es amigable con el usuario.
- **CAPA DE NEGOCIO**  
En esta capa esta representadas los requerimientos que están directamente involucradas con el negocio en forma de funcionalidades. . Es aquí donde se establecen todas las reglas que deben cumplirse. Esta capa se comunica con la de presentación, para recibir las solicitudes y presentar los resultados, y con la capa de acceso a datos, para solicitar al gestor de base de datos almacenar o recuperar datos de él.

Esta capa consta de dos clases que son: la de Scrapy que está conformada por tres métodos el soup, enlaces y datos, que se emplean para la extracción de los metadatos de los repositorios que utilizaremos en el desarrollo de esta tesis. Y la clase de LDA que es en donde analizamos los metadatos obtenidos de la clase Scrapy mediante el algoritmo LDA.

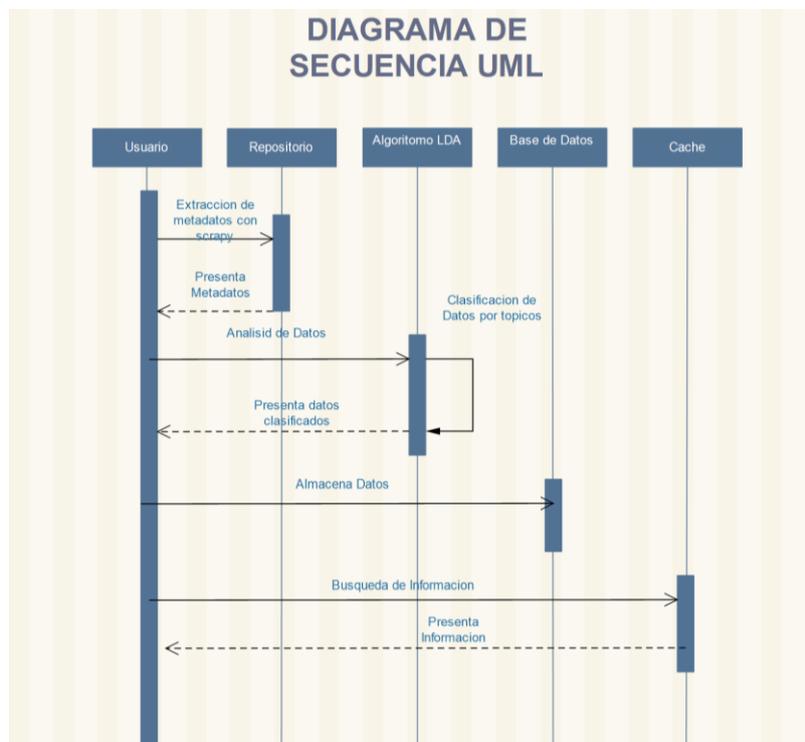
- **CAPA DE DATOS**

En esta capa se almacenan la información y devuelven datos a la capa de negocio, aunque para esto también es necesario en algunos casos, que tengan procedimientos almacenados y funciones dentro de la capa. En una arquitectura de tres capas, esta capa es la única que puede acceder a los mismos.

Para el siguiente proyecto se creó una base de datos (Prueba) que contiene una tabla Materias que contiene los metadatos extraídos de los repositorios y los resultados arrojados del análisis mediante el algoritmo Lda.

### 3.1.2 Diagrama de secuencias de la herramienta

En la figura 20 se muestra el diagrama de secuencias correspondiente al aplicativo desarrollado en el presente trabajo de titulación.



**Figura 20: Diagrama de Secuencias de la aplicación**

Fuente: El Autor

### 3.1.3 Interfaz del Buscador de REA

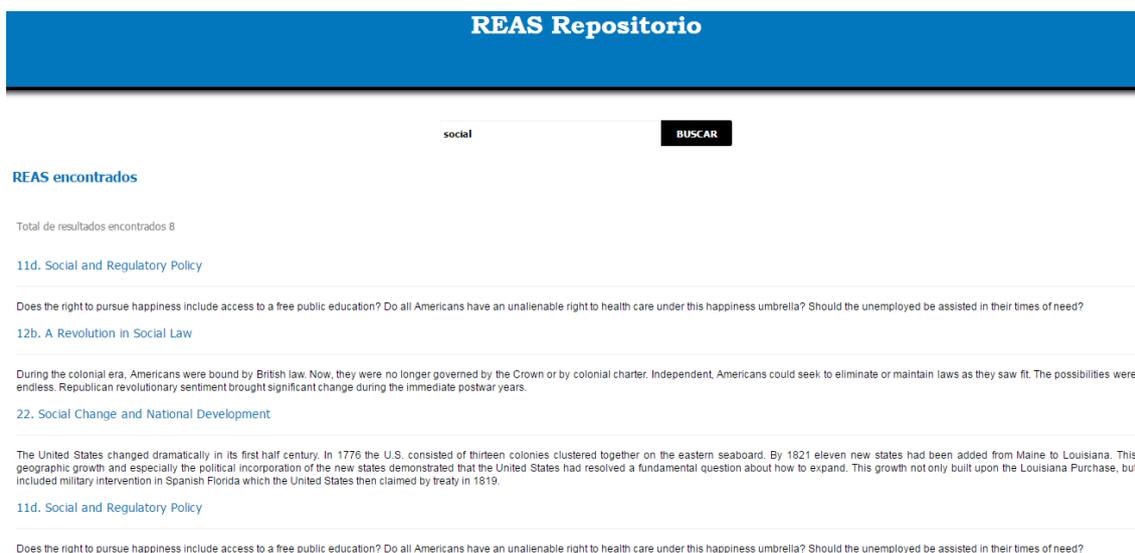
A continuación se presenta las diferentes interfaces con las que se trabajara el buscador para la representación del modelo antes elaborado.



**Figura 21: Página principal del buscador**

Fuente: El Autor

En la figura 22 se muestra la forma de cómo se presentas los resultados de la búsqueda.



**Figura 22: Página de Resultados de la búsqueda**

Fuente: El Autor

Posteriormente en la figura 23 se visualizará los metadatos del recurso seleccionado además de los recursos que son relacionados al tema que pertenece los recursos que se seleccionó.

11d. Social and Regulatory Policy

**Resumen:**

Does the right to pursue happiness include access to a free public education? Do all Americans have an unalienable right to health care under this happiness umbrella? Should the unemployed be assisted in their times of need?

**Localización:**

[Click aqui para direccionar](#)

**Tipo de Recurso:**

Images and Illustrations, Readings

**Formato del Recurso:**

Graphics/Photos, Text/HTML

**Fecha en la que se ingreso el recurso:**

NULL

**Fecha en la que se modifico el recurso:**

NULL

**Idioma del Recurso:**

English

**Autores:**

**Copyright:**

NULL

**Recursos relacionados**

#00001 The Value of A Unit With Four Cypheers Going Before It
#01 Java Tutorial: User: Hello World Programm
02.10 Linear Programming Activity/Graphs algebra II
#02 Java Tutorial: Methoden / Funktionen
#03 Java Tutorial: Variablen
#04 Java Tutorial: Schleifen / Loops
#05 Java Tutorial: Fallunterscheidung / if
#06 Java Tutorial: Klassen
#07 Java Tutorial: Vererbung
#08 Java Tutorial: Dynamische Datenstrukturen

**Figura 23: Metadatos del recurso educativo seleccionado y recursos relacionados**

Fuente: El Autor

### 3.2 Codificación

#### 3.2.1 Extracción de metadatos

La primera fase de la implementación cuenta con la extracción de los metadatos de los repositorios, para ello se utilizó scrapy y a continuación se detalla la forma como se trabajó.

**Librerías utilizadas:**

- BeautifulSoup: es una biblioteca de Python que nos permite la extracción de datos HTML y XML.
- urllib2: este módulo de Python se lo utilizo para buscar las paginas urls de los recursos que se encuentran dentro del repositorio.
- Pymysql: paquete utiliza para la interacción posterior con una base de datos
- Sys: se utiliza para transformar el código ASCII por defecto a otras codificaciones, como UTF-8

**Variables Globales:**

- base\_url: esta variable es creada para introducir la página del repositorio que contiene todos los recursos con los que se trabajara.
- pag\_url: en esta variable se le asigna la dirección de la página principal del repositorio y con la que se trabaja los enlaces de los recursos para extraer sus metadatos.

### 3.2.2 Extracción de metadatos de Repositorio Merlot

A continuación se presenta el código de extracción del repositorio Merlot y en la siguiente tabla se detalla los métodos que se utilizó para realizar la extracción de metadatos de los recursos educativos abiertos del repositorio. (Véase en el [ANEXO 3: Codificación de la extracción de Repositorios](#) )

#### Repositorio Merlot

**Def soup:** Este método permite la extracción de datos html. Su dato de entrada es la **url** y como salida los datos almacenados en la variable **html**

```
def soup(url):
    response = urllib2.urlopen(url)
    html = response.read()
    return BeautifulSoup(html)
```

**Def Enlaces:** Elabora los enlaces url para buscar los recursos en cada repositorio. Su entrada es el **url del repositorio** y como salida url de recursos en la variable **enlaces**

```
enlaces(num):
enlaces = []
for pag in range(num):
    curso_url = ''.join([base_url, "?page="+str(pag+1)+"&keywords=&sort.property=relevance&nosearchlanguage=&pageSize="])
    html=soup(curso_url)
    div = html.find("div", { "class" : "hitlistDetailContent" })
    for div_1 in div.find_all("div", { "class" : "hitlistmaterial_tableless" }):
        menu_enlace= div_1.h3.span.a.get("href")
        enla = menu_enlace.split("?id=")
        en = enla[1].split("&")
        enlace_pag = pag_url + "/merlot/viewMaterial.htm?id="+en[0]
        enlaces.append(enlace_pag)

return enlaces
```

**DEF datos:** Extracción de los metadatos de todos los recursos encontrados en el repositorio. Su entrada **url de recursos** y como salida **lista de metadatos de los recursos**.

```
def datos(num):
    data = []
    data1 = []
    a = []
    link = enlaces(num)
    cont = 0
    for links in range(len(link)):
        html = soup(link[links])
        if cont >= cont:
            data.append([])
            data1.append([])

    tit=html.title.getText().replace("'", '').replace('"', '').replace("?", '
').replace(":", '').replace("/", '').replace("(", '').replace(")", '').rep
lace("$", '').replace("#", '')
    descrip = html.find("div", { "id" : "descriptionsshort" })
    description=descrip.getText().replace(')', '')
```

```

        data[cont].append(tit)
        data[cont].append(link[links])
        data[cont].append("
".join(description.replace("'", '').split()))
        data1[cont].append("
".join(description.replace("'", '').split()))

        materal = html.find("div", { "id" : "materialtypediv" })
        mattype = materal.a.text
        language = html.find("div", { "id" : "languages_div" })
        leng = language.text.split(":")
        data[cont].append(" ".join(leng[1].split()))
        authors = html.find("div", { "id" : "author_info_div" })

        for author in authors.find_all("div"):
            autor =
author.text.replace(' ', '').replace('(', '').replace(')', '')
            a.append(" ".join(autor.split()))

            aut = ", ".join(str(j) for j in a)
            data[cont].append(aut)
            a = []
            dateadd = html.find("div", { "id" : "dateaddeddiv" })
            datead = dateadd.text.split(":")
            data[cont].append("
".join(datead[1].replace(' ', '').split()))

            formato = html.find("div", { "id" : "formatdiv" })
            if not(formato is None):
                format = formato.text.split(":")
                data[cont].append(" ".join(format[1].split()))
            else:
                data[cont].append("NULL")

            datemodifie = html.find("div", { "id" : "datemodifieddiv"
        })

            datemod = datemodifie.text.split(":")
            copyright = html.find("div", { "id" : "copyright_div" })
            copy = copyright.text.split(":")
            creativeCommos = html.find("div", { "id" : "commons_div"
        })

            commons =creativeCommos.text.split(":")
            tags = html.find("div", { "id" : "keywords_div" })
            if not(tags is None):
                keys = tags.text.split(":")
                data[cont].append(" ".join(keys[1].split()))
            else:
                data[cont].append("NULL")

            rating2 = html.find("meta", attrs={"itemprop":
"ratingValue"})
            description2 = rating2.get('content')
            print description2
            data[cont].append(description2)
            guardar(data1, tit)
            insertar(data);

        cont = cont +1

return data

```

**Def guardar:** Almacenar metadatos de los recursos en un documento. Su entrada **nombre del recurso y descripción** y como salida **archivo .txt**

```
def guardar(data, nom):
    f = open("archivos/"+nom+'.txt', 'w')
    lista = range(len(data))
    for i in lista:
        datos = "\n".join(str(j) for j in data[i])
    print (datos)
    f.write(datos+"\n")
    f.close()

    return "Datos Guardados"
```

**Def insertar:** Elaboración de una consulta sql. Su entrada es **data** y como salida **sentencia sql**

```
def insertar(data):
    for i in range(len(data)):
        datos = "\n".join(str(j) for j in data[i])
        dat = datos.split("\n")
        lista = range(len(dat))
        datIng = []
        for i in lista:
            datIng.append(''+dat[i]+'')
        dato = ",".join(str(j) for j in datIng)

        query = "INSERT INTO materias (`id`, `nombre_materia`,
`url_materia`, `descripcion_materia`, `idioma_materia`,
`autor_materia`, `fecha_materia`, `formato_materia`,
`tags_materia`, `rating_materias`) VALUES (NULL,%s)" % dato
        print query
        run_query(query)
    return "Datos Insertados"
```

Una vez obtenido los resultados se los procedió a almacenar en una base de datos con la estructura que se la presenta en la figura 24, para en lo posterior trabajar con el aplicativo.

#	Nombre	Tipo	Cotejamiento	Atributos	Nulo	Predeterminado	Extra
1	id 	int(11)			No	Ninguna	AUTO_INCREMENT
2	url_materia	varchar(100)			No	Ninguna	
3	nombre_materia	varchar(100)			No	Ninguna	
4	descripcion_materia	varchar(100)			No	Ninguna	
5	idioma_materia	varchar(10)			No	Ninguna	
6	autor_materia	varchar(50)			No	Ninguna	
7	fecha_materia	text			No	Ninguna	
8	formato_materia	text			No	Ninguna	
9	tags_materia	varchar(50)			No	Ninguna	
10	rating_materias	int(10)			No	Ninguna	
11	Ida_materias	int(10)			No	Ninguna	

**Figura 24: Estructura de la Base de datos**

Fuente: El Autor

### 3.3 Resultados Obtenidos

Tras el desarrollo del capítulo 3 se obtuvo los siguientes resultados:

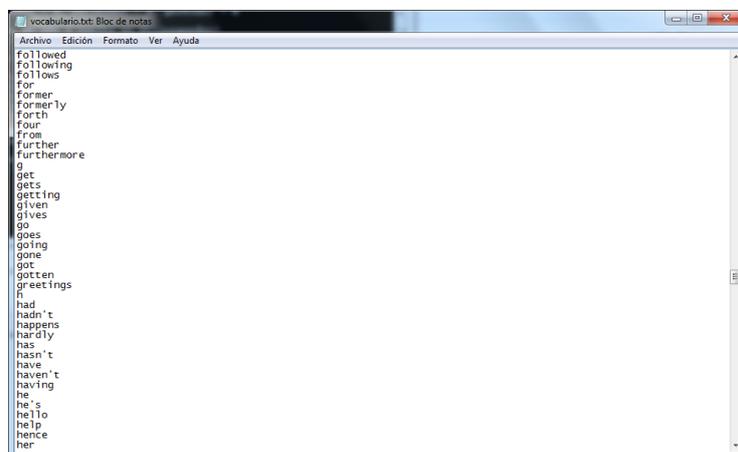
- Se pudo plasmar el modelo elaborado en el capítulo 2 en un aplicativo web, en donde se ve reflejado de forma organizada la información obtenida producto de la extracción.
- Almacenar en una base de datos estructura toda la información obtenida de los REA de los repositorios seleccionados.

**CAPÍTULO 4:**  
**VALIDACION DEL MODELO DE REPRESENTACION IMPLEMENTADO A TRAVÉS**  
**DE UN SISTEMA RECOMENDADOR**

En este capítulo se validará el modelo común de representación implementado en un aplicativo, para lo que se desarrollará un sistema de recomendación basado en LDA, además se utilizará métricas de evaluación como son la precisión y recall para evaluar el sistema recomendador.

#### 4.1 Valoración de recomendaciones con algoritmo LDA

Primeramente se define un diccionario en el cual se colocaran las palabras vacías que son aquellas palabras sin significado como artículos, pronombres, preposiciones, etc. que son filtradas antes o después del procesamiento de datos en lenguaje natural. En el Anexo 4 se muestra todas las palabras que conforman el diccionario. En la figura 25 se presenta la estructura correspondiente al diccionario.

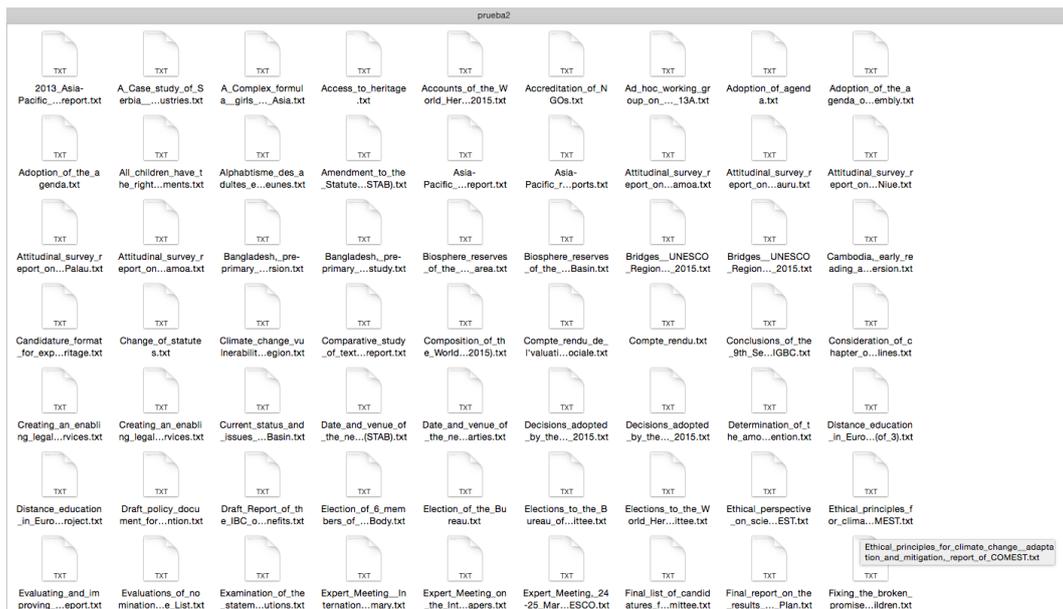


**Figura 25: Diccionario LDA**

Fuente: El Autor

Al haber obtenido con el Scrapy los metadatos de los recursos educativos, se procede a guardar en un archivo *.txt* las descripciones de estos recursos. Sobre estos archivos se emplea Lda (*topics.py*) para definir los temas (llamados topics) al que pertenecen los recursos y para definir las palabras de dichos temas.

En la figura 26 se observa la forma como se encuentra agrupado las descripciones de los recursos con los que se trabajará.



**Figura 26: Descripciones de los recursos educativos abiertos**

Fuente: El Autor

A continuación en el siguiente apartado se puede observar el algoritmo utilizado.

```
import lda
import sys
import os
import scipy
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer

archivos=os.listdir("/home/rbgordillo/prueba2")
nombre_archivos=[]
for a in archivos:
    nombre_archivo=os.path.join("/home/rbgordillo/prueba2",a)
    nombre_archivos.append(nombre_archivo)

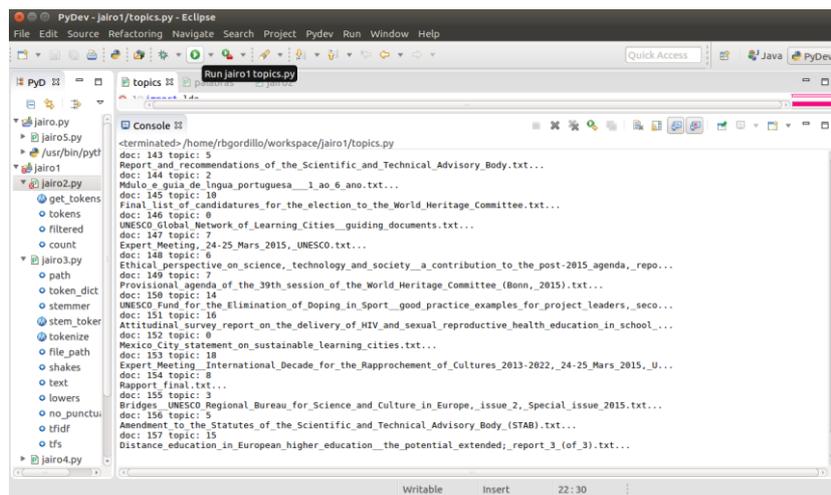
with open("/home/rbgordillo/vocabulario/vocabulario.txt","r") as stopword:
    stopword=stopword.read().split()

Vector = CountVectorizer(input='filename', stop_words=stopword,
decode_error='replace')
MatrizDatos = Vector.fit_transform(nombre_archivos)
Vocabulario = Vector.get_feature_names()
Titulos = os.listdir("/home/rbgordillo/prueba2")

Modelo = lda.LDA(n_topics=20, n_iter=100, alpha=1, eta =0.05,
random_state=1)
Modelo.fit(MatrizDatos)

doc_topic = Modelo.doc_topic_
for n in range(158):
    topic_most_pr = doc_topic[n].argmax()
    print("doc: {} topic:
    {{}\n{}}...".format(n,topic_most_pr,Titulos[n][:100]))
```

En la figura 27 podemos observar el resultado obtenido aplicando LDA



**Figura 27: Resultado de tópicos**

Fuente: Autor

El siguiente algoritmo es un modelo de la que se usó para determinar las palabras que pertenecen a un tópico, posteriormente se la almacena en una base de datos para su utilización. A continuación se puede observar el algoritmo completo se utilizó.

```
import lda
import sys
import os
import scipy
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer

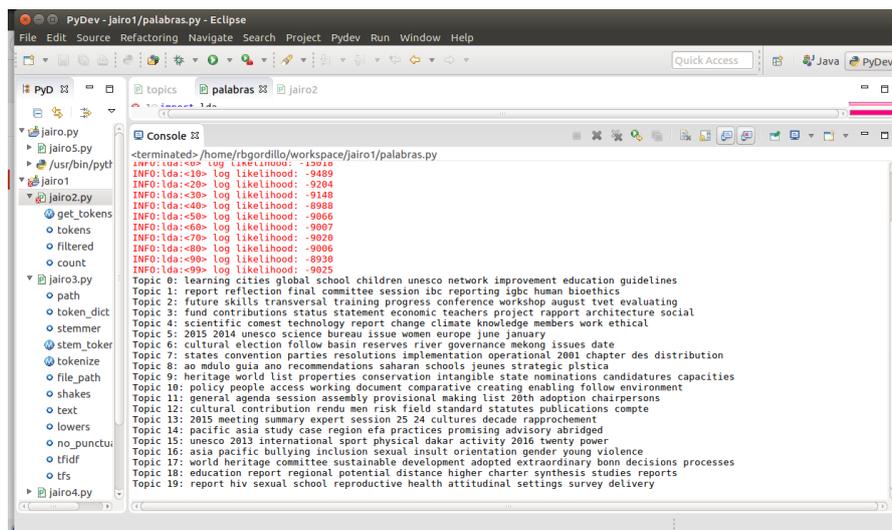
archivos=os.listdir("/home/rbgordillo/prueba2")
nombre_archivos=[]
for a in archivos:
    nombre_archivo=os.path.join("/home/rbgordillo/prueba2",a)
    nombre_archivos.append(nombre_archivo)
with open("/home/rbgordillo/vocabulario/vocabulario.txt","r") as
stopword:
    stopword=stopword.read().split()

Vector = CountVectorizer(input='filename', stop_words=stopword,
decode_error='replace')
MatrizDatos = Vector.fit_transform(nombre_archivos)
Vocabulario = Vector.get_feature_names()
Titulos = os.listdir("/home/rbgordillo/prueba2")

Modelo = lda.LDA(n_topics=20, n_iter=100, random_state=1)
Modelo.fit(MatrizDatos)

topic_word = Modelo.topic_word_
n_top_words = 10
for i, topic_dist in enumerate(topic_word):
    topic_words = np.array(Vocabulario)[np.argsort(topic_dist)][:-
(n_top_words+1):-1]
    print('Topic {}: {}'.format(i, ' '.join(topic_words)))
```

En la figura 28 se puede observar el resultado aplicando LDA



**Figura 28: Resultado de tópicos\_word**

Fuente: Autor

El proceso de validación mediante el algoritmo LDA se lo aplico únicamente a los REA en inglés, debido a que el algoritmo no reconoce caracteres especiales como que se encontraban en las descripciones de los recursos en español.

A continuación se muestran las varias pruebas realizadas modificando los valores de los parámetros, estas pruebas permitirán obtener valores que nos devuelvan información coherente como resultados de las búsquedas.

Se presentan a continuación valores referentes a, las palabras correspondientes a los Tópicos con diferentes parámetros tanto en el número de tópicos como en el número de interacciones.

**Tabla 35: Topic\_World = n\_topics= 10, n\_iter= 100**

Topic 0 :clay, follow, personas, making, 1831, print, especiales, secretary, decision, appears
Topic 1 :magazines, anglicisms, ecuadorian, research, analysis, general, interest, cultural, scholarly, order
Topic 2 :education, create, report, health, hiv, survey, people, policy, reproductive, equation
Topic 3 :science, agenda, reading, including, test, learn, human, bone, list, excellent
Topic 4 :ich, proceso, ocw, presente, accesibilidad, anfang, estudio, lcke, nachdem, einige
Topic 5 :language, spanish, english, variables, newspapers, newspaper, social, sections, national, pages
Topic 6 :world, heritage, unesco, 2015, committee, session, report, meeting, states, parties
Topic 7 :asia, pacific, education, learning, report, school, region, cities, efa, skills
Topic 8 :student, study, work, writing, case, program, teacher, reflection, teaching, aspects
Topic 9 :site, students, interactive, learning, tutorial, project, video, information, topics, resources

Fuente: El Autor

**Tabla 36: Topic\_World = n\_topics= 10, n\_iter= 200**

Topic 0:proceso,ocw,presente,estudiante,personas,accesibilidad,estudio,asperger,ingls,ensear
Topic 1:pk,pp,pu,zn,ba,ox,uo,kb,2r,1w
Topic 2:education,asia,pacific,report,school,ich,sexual,region,study,efa
Topic 3:xmlpk,xml,zl,ww,qq,tx,l9,rx,meta,cr
Topic 4:newspapers,social,newspaper,sections,national,pages,related,clay,reading,reports
Topic 5:world,heritage,unesco,2015,report,session,committee,list,learning,meeting
Topic 6:configurations2,lb,qw,qe,gb,lf,fu,cm,vy,manifest
Topic 7:magazines,anglicisms,language,research,ecuadorian,study,analysis,general,english,spanish
Topic 8:students,student,learning,writing,skills,program,learn,case,project,teacher
Topic 9:site,interactive,resources,information,concepts,simulations,includes,topics,materials,find

Fuente: El Autor

**Tabla 37: Topic\_World = n\_topics= 10, n\_iter= 5**

Topic 0:information,based,approach,collection,students,determine,award,proceso,quito,ms
Topic 1:pk,national,standard,authentic,videos,internet,project,xml,einige,distance
Topic 2:video,asia,ich,school,sexual,practices,includes,related,skills,topics
Topic 3:report,2013,education,clay,teaching,pacific,writing,heritage,languages,level
Topic 4:research,descriptive,ecuador,words,investigation,newspapers,states,number,social,newspaper
Topic 5:world,interest,unesco,heritage,students,committee,learning,science,interactive,simulations
Topic 6:configurations2,learning,policy,pp,final,xmlpk,study,cities,body,assignments
Topic 7:magazines,anglicisms,language,ecuadorian,analysis,english,spanish,general,research,qualitative
Topic 8:education,report,program,learn,concepts,pacific,content,teacher,unesco,presente
Topic 9:site,study,scholarly,cultural,found,variable,order,variables,case,selected

Fuente: El Autor

**Tabla 38: Topic\_World = n\_topics= 20, n\_iter= 100**

Topic 1:student,writing,skills,english,learning,case,information,teacher,teaching,document
Topic 2:magazines,ecuadorian,analysis,language,general,scholarly,cultural,interest,order,descriptive
Topic 3:iz,pk,lq,uw,mn,yk,mm,oz,2x,qp
Topic 4:ocw,proceso,personas,accesibilidad,derechos,desarrollo,appears,ms,calidad,especiales
Topic 5:presente,estudio,estudiante,ingls,ensear,asperger,sndrome,implcita,utpl,explcita
Topic 6:report,learning,school,sexual,education,cities,health,global,reproductive,hiv
Topic 7:configurations2,hz,meta,manifest,lc,bj,w0,js,ii,images
Topic 8:2015,unesco,world,session,report,committee,heritage,meeting,general,agenda
Topic 9:site,interactive,students,resources,concepts,simulations,includes,topics,materials,including
Topic 10:follow,xml,status,d6,graph,39,steps,l3,ing,kc

Topic 11:xmlpk,je,wa,qb,uf,bm,6b,bk,basin,i2
Topic 12:video,music,documents,tutorial,ist,molecules,10,und,das,konzept
Topic 13:study,language,work,program,foreign,grammar,physical,efl,making,students
Topic 14:science,issue,inclusion,bullying,bureau,6j,technology,zy,mj,l9
Topic 15:social,newspapers,newspaper,sections,people,pages,national,change,section,2009
Topic 16:ich,create,anfang,videos,nachdem,guia,die,ao,mdulo,nun
Topic 17:heritage,world,states,list,convention,parties,clay,cultural,1831,print
Topic 18:site,region,study,practices,case,award,promising,paper,plan,version
Topic
19:anglicisms,research,spanish,variable,found,words,variables,quantitative,ecuador,investigation

Fuente: El Autor

**Tabla 39:Topic\_World = n\_topics= 20, n\_iter= 200**

Topic 1:bone,kb,gr,n9,wq,gn,yn,ew,bb,1u
Topic 2:pk,configurations2,pp,xml,xmlpk,zn,ev,ox,1w,meta
Topic 3:site,tutorial,project,concepts,topics,materials,interactive,video,award,collection
Topic 4:zl,2r,lf,ne,yb,kg,uu,pj,wu,jg
Topic 5:site,learn,information,create,includes,assessment,part,authentic,student,find
Topic 6:language,study,english,writing,student,work,program,foreign,research,teaching
Topic 7:states,ich,meeting,parties,anfang,change,summary,die,nun,gemacht
Topic 8:pu,ba,manifest,tx,eb,uo,vo,dc,ed,w7
Topic 9:magazines,anglicisms,ecuadorian,general,spanish,interest,scholarly,order,research,found
Topic 10:world,heritage,learning,list,convention,unesco,cities,global,cultural
Topic 11:nr,video,und,ww,ist,cr,konzept,um,ko,rbgordillo
Topic 12:education,school,report,pacific,asia,skills,learning,sexual,people,approach
Topic 13:clay,print,1831,library,secretary,appears,national,men,war,sketch
Topic
14:analysis,anglicisms,cultural,language,descriptive,variables,words,research,investigation,social
Topic 15:2015,report,unesco,session,2014,2013,science,agenda,general,committee
Topic 16:lb,cm,fu,hh,gb,qe,qw,yu,af,cs
Topic 17:students,simulations,based,interactive,including,teacher,aspects,test,selected,merlot
Topic 18:proceso,ocw,accesibilidad,personas,derechos,calidad,desarrollo,especiales,sitios,poder
Topic 19:asia,pacific,study,case,region,practices,efa,promising,abridged,version

Fuente: El Autor

**Tabla 40:Topic\_World = n\_topics= 20, n\_iter= 5**

Topic 1:education,reading,practice,xml,make,regional,bone,interest,distance,implementation
Topic 2:pk,video,pu,tu,cm,statement,zy,lists,lf,science
Topic 3:cultural,research,order,investigation,technology,project,related,topics,document,number

Topic 4:site,data,results,proceso,human,1831,ensear,enseanza,important,ncate
Topic 5:information,create,tutorial,body,languages,time,standards,advisory,presents,eb
Topic 6:research,list,study,pages,teaching,pp,content,zl,method,articles
Topic 7:ich,session,meeting,including,anfang,24,2009,network,global,einige
Topic 8:learn,student,teacher,resources,teachers,english,developed,part,xm1pk,lesson
Topic 9:anglicisms,magazines,analysis,scholarly,qualitative,found,quantitative,methods,site,program
Topic 10:learning,world,unesco,heritage,concepts,follow,cities,authentic,equation,committee
Topic 11:configurations2,settings,videos,music,lcke,zn,skill,habe,exercises,national
Topic 12:report,education,sexual,health,skills,approach,school,reflection,hiv,work
Topic 13:magazines,ecuadorian,language,spanish,words,clay,ocw,general,written,section
Topic
14:anglicisms,general,interest,variables,study,descriptive,language,newspapers,ecuador,ecuadorian
Topic 15:2015,unesco,report,2013,aspects,future,session,meta,status,knowledge
Topic 16:asia,pacific,committee,convention,student,school,children,nr,assignments,activities
Topic 17:students,simulations,language,access,estudio,readers,report,evolution,online,change
Topic 18:study,case,learning,conducted,asia,presente,derechos,abridged,print,higher
Topic 19:world,heritage,variable,pacific,region,practices,cultural,parties,order,paper

Fuente: El Autor

Como se puede observar en las tablas anteriores entre mayor número de interacciones y mayor número de tópicos existe datos sin concordancia. Como resultados de las pruebas realizadas se definen los siguientes valores considerados idóneos para trabajar dentro del sistema de recomendación implementado en el presente trabajo de titulación.

**N\_topics= 10**

**N\_iter=5**

Se presentan a continuación valores referentes a, el número de tópico con diferentes parámetros como son el número de tópicos, el número de interacciones, valores de Alpha y Eta

**Tabla 41: Doc\_Topic**

N_topics: 20 N_iter=100 Alpha:0.05 Eta:0.05	N_topics: 10 N_iter=100 Alpha:0.05 Eta:0.05	N_topics: 10 N_iter=5 Alpha:0.05 Eta:0.05	N_topics: 20 N_iter=5 Alpha:0.05 Eta:0.05
doc: 0 topic: 13	doc: 0 topic: 1	doc: 0 topic: 4	doc: 0 topic: 15
doc: 1 topic: 4	doc: 1 topic: 8	doc: 1 topic: 5	doc: 1 topic: 15
doc: 2 topic: 3	doc: 2 topic: 6	doc: 2 topic: 2	doc: 2 topic: 2
doc: 3 topic: 3	doc: 3 topic: 5	doc: 3 topic: 0	doc: 3 topic: 5
doc: 4 topic: 6	doc: 4 topic: 1	doc: 4 topic: 8	doc: 4 topic: 10

doc: 5 topic: 17	doc: 5 topic: 9	doc: 5 topic: 6	doc: 5 topic: 16
doc: 6 topic: 8	doc: 6 topic: 9	doc: 6 topic: 9	doc: 6 topic: 12
doc: 7 topic: 1	doc: 7 topic: 6	doc: 7 topic: 2	doc: 7 topic: 8
doc: 8 topic: 8	doc: 8 topic: 1	doc: 8 topic: 6	doc: 8 topic: 1
doc: 9 topic: 17	doc: 9 topic: 9	doc: 9 topic: 9	doc: 9 topic: 18
doc: 10 topic: 1	doc: 10 topic: 8	doc: 10 topic: 0	doc: 10 topic: 10
doc: 11 topic: 16	doc: 11 topic: 6	doc: 11 topic: 2	doc: 11 topic: 1
doc: 12 topic: 11	doc: 12 topic: 9	doc: 12 topic: 0	doc: 12 topic: 9
doc: 13 topic: 3	doc: 13 topic: 6	doc: 13 topic: 5	doc: 13 topic: 6
doc: 14 topic: 16	doc: 14 topic: 8	doc: 14 topic: 1	doc: 14 topic: 6
doc: 15 topic: 5	doc: 15 topic: 6	doc: 15 topic: 2	doc: 15 topic: 11
doc: 16 topic: 16	doc: 16 topic: 6	doc: 16 topic: 6	doc: 16 topic: 7
doc: 17 topic: 11	doc: 17 topic: 5	doc: 17 topic: 8	doc: 17 topic: 12
doc: 18 topic: 18	doc: 18 topic: 5	doc: 18 topic: 1	doc: 18 topic: 1
doc: 19 topic: 16	doc: 19 topic: 6	doc: 19 topic: 9	doc: 19 topic: 17
doc: 20 topic: 16	doc: 20 topic: 6	doc: 20 topic: 0	doc: 20 topic: 17
doc: 21 topic: 19	doc: 21 topic: 7	doc: 21 topic: 9	doc: 21 topic: 2
doc: 22 topic: 8	doc: 22 topic: 9	doc: 22 topic: 8	doc: 22 topic: 7
doc: 23 topic: 11	doc: 23 topic: 8	doc: 23 topic: 8	doc: 23 topic: 8
doc: 24 topic: 0	doc: 24 topic: 4	doc: 24 topic: 0	doc: 24 topic: 0
doc: 25 topic: 9	doc: 25 topic: 4	doc: 25 topic: 6	doc: 25 topic: 11

Fuente: El Autor

Como resultados de las pruebas realizadas se definen los siguientes valores considerados idóneos para trabajar dentro del sistema de recomendación implementado en el presente trabajo de titulación.

**N\_topics= 10**

**N\_inter=5**

**Alpha= 0.05**

**Etha= 0.05**

#### **4.2 Precisión y Recall**

Una vez aplicado el modelo es necesario determinar la calidad de las recomendaciones, para el efecto se aplicara las medidas de calidad de las recomendaciones (Precisión, Recall).

Para calcular la precisión y exhaustividad, se ha definido la tabla de contingencia en donde se encontraran los ítems que fueron clasificados primeramente como relevantes e irrelevantes y luego como recomendados y no-recomendados a los usuarios en base a los valores obtenidos de la extracción de datos de REA de los repositorios Merlot y OERCOMMNOS ya que estos repositorios son los únicos que cuentan con la opción de rating.

- **Base de Datos:** recursos educativos abiertos del repositorio MERLOT y OERCOMMNOS
- **Encontrados:** REA de MERLOT y OER COMMNOS en idioma inglés y que tengan rating.
- **Relevantes:** REA de MERLOT y OER COMMNOS en idioma inglés y que tengan rating mayor a 3.
- **Relevantes no Recuperados:** REA de MERLOT y OER COMMNOS que no estén en idioma inglés y que no tengan rating mayor a 3.
- **Irrelevantes:** REA de MERLOT y OER COMMNOS en idioma inglés y que tengan rating menor o igual 3.

La Tabla 42, muestra los datos necesarios que se tienen que obtener y poder aplicar las medidas de precisión y recall de acuerdo a las fórmulas antes indicadas.

**Tabla 42: Tabla de Contingencia**

Base de Datos	810
Datos encontrados	668
Relevantes	456
Relevantes no recuperados	354
Irrelevantes recuperados	212

Fuente: Autor

Con los datos obtenidos de las recomendaciones se realizan los cálculos y como resultado tenemos que:

Precisión: 0,682

Recall: 0,562

Esto significa que los valores obtenidos a partir del sistema de recomendación implementado, cuenta con un 68% de precisión y con un 56% de recall, que representan más de la mitad del total de valores logrados , por lo que se considera al sistema como una sistema preciso y confiable.

## CONCLUSIONES

- El indicador de calidad más relevante para la selección de los repositorios de REA es que pertenezca a una Institución de Educación Superior o a Organismos destacados dentro del campo Educativo, que proteja y garantice la calidad y propiedad de los recurso educativos.
- En base al análisis realizado a los repositorios de REA seleccionados, se pudo identificar información referente a la estructura de la metadatos de los repositorios, la cual nos permitió extraer datos que nos sirvió para la elaboración del modelo de común de representación.
- Siendo LOM y Dublin Core estándares de metadatos utilizados para la representación de REA dentro de los repositorios, Dublin Core resulta el estándar correcto debido a que cuenta con una estructura de metadatos más simple y concreta para la representación de REA.
- Debido a su flexibilidad y sencillez de adaptación en cualquier escenario, Scrappy resulta el mecanismo más óptimo para la extracción de contenidos en sitios web sobre todo al tratarse con estructuras diferentes.
- Un modelo de representación común para repositorios de REA resulta importante, ya que de esta forma toda la información se encontraría estructurada uniformemente lo que facilitaría la búsqueda y recuperación de dicha información.
- Las métricas de precisión y recall permiten la evaluación del sistema de recomendación y de esta manera garantizar que la información obtenida es confiable y eficaz.

## RECOMENDACIONES

Posteriormente de las conclusiones establecidas como producto de la investigación realizada, se presentan las siguientes recomendaciones:

- Para la publicación de una REA, definir licencias de uso es necesario para facilitar y mejorar la usabilidad de dicho recurso.
- Para la representación de REA dentro de repositorios, el estándar Dublin Core es el aconsejable ya que a diferencia de LOM, cuenta con los metadatos concretos y necesarios para la representación.
- Definir e implementar indicadores de calidad para la selección de RREA, ya que de esta manera se cuenta con la seguridad y confiabilidad de la información con la que se trabaja.
- Para la extracción de contenidos en sitios web con estructuras diferentes se recomienda utilizar Scrappy ya que cuenta con una arquitectura muy flexible que puede acoplarse a cualquier entorno.

## REFERENCIAS

- Alfano, M., Lenzitti, B., Visalli, N., Objects, L., & Indicators, D. (2007). SAXEF : A System for Automatic eXtraction of E-learning object Features. *Je-LKS: Journal of E-Learning and Knowledge Society*, 3(2), 83–92.
- Australian Research Council (ARC). (2013). Open access policy concerning UNESCO publications, (January), 1–4. Retrieved from [http://www.arc.gov.au/applicants/open\\_access.htm](http://www.arc.gov.au/applicants/open_access.htm)
- Azorín, C., Cívico Martín, R., Combarro Felpeto, P., García Gil, M. Á., Losada Yáñez, M., Moreno Cañizares, A., ... Rocío, S. V. (2014). GUÍA PARA LA EVALUACIÓN DE REPOSITARIOS INSTITUCIONALES DE INVESTIGACIÓN, 1–34. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Gu+a+para+la+evaluaci+n+de+repositorios+institucionales+de+investigaci+n+1#0>  
<http://digital.csic.es/handle/10261/35735>
- Barber, J. (2014). Latent Dirichlet Allocation (LDA) with Python. Retrieved from [https://rstudio-pubs-static.s3.amazonaws.com/79360\\_850b2a69980c4488b1db95987a24867a.html](https://rstudio-pubs-static.s3.amazonaws.com/79360_850b2a69980c4488b1db95987a24867a.html)
- Barker, P. (2005). What is IEEE Learning Object Metadata / IMS Learning Resource Metadata?
- Blake, L., & Olaya, V. (2016). Metadatos. Retrieved from <http://volaya.github.io/libro-sig/chapters/Metadatos.html>
- Bui, Y., & Park, J.-R. (2006). An assessment of metadata quality: A case study of the National Science Digital Library Metadata Repository. *Proceedings of CAISACSI 2006 Information Science Revisited Approaches to Innovation*, 13. Retrieved from <http://idea.library.drexel.edu/handle/1860/1600>
- Castro, M., Castillo, E., Lillo, P., Quiroz, A., & Riquelme, P. (2009). Guía para la creación de metadatos usando dublin core.
- Cristobal, C. (2010). Repositorios de objetos de aprendizaje. Retrieved February 19, 2017, from <http://es.slideshare.net/ServidorCelia/repositorios-de-objetos-de-aprendizaje-4541145>
- Cruz, B., Manuel, J., & València, U. De. (2014). Guía para la evaluación de repositorios institucionales de investigación. Madrid: FECYT, RECOLECTA y CRUE, 1–29.
- EcuRed. (2017). Cliente-Servidor - EcuRed. Retrieved March 6, 2017, from <https://www.ecured.cu/Cliente-Servidor>
- Epicor. (2017). Arquitectura Orientada a Servicios | SOA | Epicor. Retrieved March 7, 2017, from <http://www.epicor.com/lac/solutions/soa.aspx>
- Ferreras, T. (2008). Dublin core cualificado.
- Flores, M. A. C. (2009). Ontología para la recomendación de recursos educativos almacenados en el Repositorio de Objetos de Aprendizaje (ROA) DSpace.
- Hernández, G., Alvarado, A., & Chávez, C. (2011). Una mirada a los estándares de calidad relacionados con e-Learning, (37).
- Hughes, B. (2004). *Metadata Quality Evaluation: Experience from the Open Language Archives Community*. (E. L. Zhaoneng Chen, Hsinchun Chen, Qihao Miao, Yuxi Fu, Edward Fox, Ed.) (ilustrada). Shanghai: Springer.
- Kent, A., Berry, M. M., Luehrs, F. U., & Perry, J. W. (1955). Machine literature searching VIII. Operational criteria for designing information retrieval systems. *American*

- Documentation*, 6(2), 93–101. <http://doi.org/10.1002/asi.5090060209>
- Krulwich, B. (1997). LIFESTYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data. *AI Magazine*, 18(2), 37. <http://doi.org/10.1609/aimag.v18i2.1292>
- Lewis, J. (2014). microservicios. Retrieved March 7, 2017, from <https://martinfowler.com/articles/microservices.html>
- Manso, M. Á. (2009). El uso de los metadatos para el desarrollo de un modelo de interoperabilidad para las Infraestructuras de Datos Espaciales, 172. Retrieved from [https://www.academia.edu/2962720/El\\_uso\\_de\\_los\\_metadatos\\_para\\_el\\_desarrollo\\_de\\_un\\_modelo\\_de\\_interoperabilidad\\_para\\_las\\_Infraestructuras\\_de\\_Datos\\_The\\_use\\_of\\_metadata\\_for\\_the\\_development\\_of\\_a\\_model](https://www.academia.edu/2962720/El_uso_de_los_metadatos_para_el_desarrollo_de_un_modelo_de_interoperabilidad_para_las_Infraestructuras_de_Datos_The_use_of_metadata_for_the_development_of_a_model)
- Microsoft Corporation. (2009). *Microsoft Application Architecture Guide*. Retrieved from <http://www.microsoft.com/en-us/download/details.aspx?id=16236>
- Ministerio de Educación Nacional de Colombia. (2012). *Recursos Educativos Digitales Abiertos*.
- Morales, V. T., & Vicari, R. M. (2013a). en repositorios digitales de objetos de aprendizaje, 36, 183–195.
- Morales, V. T., & Vicari, R. M. (2013b). Evaluación de la calidad de metadatos en repositorios digitales de objetos de aprendizaje, 36, 183–195.
- Moreno, J., Duque, N. D., & Pineda, A. F. (2012). ROAp: Una Propuesta de Repositorio de Objetos de Aprendizaje basada en el estándar LOM. *Séptima Conferencia Latinoamericana de Objetos Y Tecnologías de Aprendizaje - LACLO 2012*, 216 – 224.
- Motz, R., Badell, C., Barrosa, M., Sum, R., Díaz, G., & Castro, M. (2009). LookIng4LO Sistema Informático para la extracción automática de objetos de Aprendizaje Caso de Estudio. *IEEE-RITA. Revista Iberoamericana de Tecnologías Del Aprendizaje*, 4(3), 223–229. Retrieved from <http://rita.det.uvigo.es>
- Motz, R., Badell, C., Barrosa, M., Sum, R., Díaz, G., & Castro, M. (2014). LookIng4LO.
- Moya, R. (2015). Mapas gráficos para la visualización de relaciones en sistemas de recomendación.
- Ochoa, X., & Duval, E. (2006). Quality metrics for learning object metadata, 1004–1011.
- Ortega Fernando, Bobadilla, J., & Antonio, H. (2012). Incorporating Group Recommendations to Recommender Systems: Alternatives and Performance.
- Peláez, J. (2017). Arquitectura basada en Componentes – Blog de Juan Peláez en Geeks.ms. Retrieved March 7, 2017, from <http://geeks.ms/jkpelaez/2009/04/18/arquitectura-basada-en-componentes/>
- Pinilla, A., Gutiérrez, M., & Ballejos, L. (2014). Extracción Automática de Metadatos a partir de Objetos de Aprendizaje en un Repositorio Institucional : Estado del Arte. *Simposio Argentino de Tecnología Y Sociedad*, 67–82.
- Ricci, F. (2011). *Recommender systems handbook*. Springer.
- Rodríguez, P., Isaza, G., & Duque, N. D. (2011). Búsqueda personalizada en Repositorios de Objetos de Aprendizaje a partir del perfil del estudiante. *CAVA*, 11.
- Salter, J., & Antonopoulos, N. (2006). CinemaScreen recommender agent: Collaborative and content-based filtering. *Intelligent Systems*, 21(February), 35–41.

- <http://doi.org/http://doi.ieeecomputersociety.org/10.1109/MIS.2006.4>
- Salton, G., & McGill, M. (1983). *An Introduction to Information Retrieval*. (McGraw Hill, Ed.). New York, NY, USA.
- Sánchez, D. (2016). Arquitectura de microservicios - Parte 1: Introducción. Retrieved March 7, 2017, from <http://enmilocalfunciona.io/arquitectura-microservicios-1/>
- Santos-Hermosa, G., Ferrán-Ferrer, N., & Abadal Falgueras, E. (2015). Evaluación de Repositorios Educativos Contexto de la investigación. *XIV Workshop Rebiun de Poyectos Digitales / VI Jornadas OS-Repositorios (Córdoba, 11-13 de Marzo de 2015)*, 11–13. Retrieved from [http://helvia.uco.es/xmlui/bitstream/handle/10396/12635/Santos\\_Hermosa.pdf?sequence=2](http://helvia.uco.es/xmlui/bitstream/handle/10396/12635/Santos_Hermosa.pdf?sequence=2)
- Scrapy. (2015). Architecture overview — Scrapy 1.1.0 documentation. Retrieved May 14, 2016, from <http://doc.scrapy.org/en/latest/topics/architecture.html#components>
- Serrano Vicente, R., Melero Melero, R., & Abadal Falgueras, E. (2014). Indicadores para la evaluación de repositorios institucionales de acceso abierto. *Anales de Documentación*, 17(2), 1–12. <http://doi.org/http://dx.doi.org/10.6018/analesdoc.17.2.190821>
- Serrano Vicente, R., Melero Melero, R., & Abadal, E. (2014). Indicadores para la evaluación de repositorios institucionales de acceso abierto. *Anales de Documentación*, 17(2). Retrieved from <http://revistas.um.es/analesdoc/article/view/analesdoc.17.2.190821/165851>
- Sheldon, T. (2001). Metadata (Linktionary term). Retrieved from <http://www.linktionary.com/m/metadata.html>
- Steinacker, A., Ghavam, A., & Steinmetz, R. (2001). Metadata Standards for Web-Based Resources. Retrieved from [http://ivizlab.sfu.ca/arya/Papers/IEEE/Multimedia/2001/Jan/Metadata Standards for Web-based Resources.pdf](http://ivizlab.sfu.ca/arya/Papers/IEEE/Multimedia/2001/Jan/Metadata%20Standards%20for%20Web-based%20Resources.pdf)
- Stopwords. (2016). Retrieved January 12, 2017, from <http://www.ranks.nl/stopwords>
- Swan, A. ;John, & Houghton. (2012). Going for Gold? The costs and benefits of Gold Open Access for UK research institutions: further economic modelling. Retrieved April 6, 2016, from [http://repository.jisc.ac.uk/610/2/Modelling\\_Gold\\_Open\\_Access\\_for\\_institutions\\_-\\_final\\_draft3.pdf](http://repository.jisc.ac.uk/610/2/Modelling_Gold_Open_Access_for_institutions_-_final_draft3.pdf)
- Swick, R. (2002). Metadata Activity. Retrieved from <https://www.w3.org/Metadata/Activity.html>
- Temesio, S. (2012). Metadatos para recursos educativos. Retrieved from <http://www.palabraclave.fahce.unlp.edu.ar/article/download/PCv5n1a03/6907>
- Temesio, S. (2015). Metadatos para recursos educativos, 5(Cc), 1–18.
- Testa, P., & Degiorgi, E. (2013). institucionales de las universidades nacionales Facultad de Educación Elemental y Especial.
- Tobares, V., Duque, N., Cadavid, J., Ovalle, D., & Vicari, R. (2013, March 14). Evaluación de la calidad de metadatos en repositorios digitales de objetos de aprendizaje. Retrieved from Ordenadores Electrónicos
- UNESCO. (2011). *A basic guide to open educational resources (OER)*.
- Vicari, R. M., Ribeiro, A., Marques, J., Santos, R., Primo, T., & Bez, M. (2009). Brazilian Proposal

for Agent-Based Learning Objects Metadata Standard - OBAA.

Vierkant, P. (2012). 2012 Census of Open Access Repositories in Germany: Turning Perceived Knowledge Into Sound Understanding. Retrieved April 6, 2016, from <http://www.dlib.org/dlib/november13/vierkant/11vierkant.html>

Westell, M. (2006). Institutional repositories: proposed indicators of success. *Library Hi Tech*, 24(2), 211–226. <http://doi.org/10.1108/07378830610669583>

## **ANEXOS**

## ANEXO 1: Resultados de la revisión de ROAs para cálculo de peso de metadatos

ROA Metadatos LOM	EN BÚSQUEDAS AVANZADAS												SUB TOTAL	EN RESULTADOS												SUB TOTAL	TOTAL	TOTAL ^2	PESOS
	1	2	3	4	5	6	7	8	9	10	11	12		1	2	3	4	5	6	7	8	9	10	11	12				
General.Identifier.Catalog												1	1												0	1	1	0,09%	
General.Identifier.Entry							1					1	2													0	2	4	0,35%
General.Title	1	1				1	1					1	5	1	1	1	1	1	1	1	1	1	1	1	11	16	256	22,18%	
General.Language	1		1	1				1	1			1	6												0	6	36	3,12%	
General.Description	1	1					1					1	4	1	1	1	1	1	1			1	1	1	8	12	144	12,48%	
General.Keyword	1	1	1	1	1	1	1	1	1	1	1	1	12						1			1	1	1	3	15	225	19,50%	
General.Coverage												1	1												0	1	1	0,09%	
General.Structure												1	1												0	1	1	0,09%	
General.Agregation Level							1					1	2							1				1	3	9	0,78%		
LifeCycle.Version												1	1												0	1	1	0,09%	
LifeCycle.Status								1				1	2												0	2	4	0,35%	
LifeCycle.Contribute.Role												1	1												0	1	1	0,09%	
LifeCycle.Contribute.Entity	1		1	1	1	1	1	1	1	1	1	1	9	1					1	1				3	12	144	12,48%		
LifeCycle.Contribute.Date												1	1												0	1	1	0,09%	
Meta-Metadata.Identifier.Catalog												1	1												0	1	1	0,09%	
Meta-Metadata.Identifier.Entry												1	1												0	1	1	0,09%	
Meta-Metadata.Contribute.Role												1	1												0	1	1	0,09%	
Meta-Metadata.Contribute.Entity												1	1												0	1	1	0,09%	
Meta-Metadata.Contribute.Date												1	1												0	1	1	0,09%	
Meta-Metadata.Metadata Schema												1	1												0	1	1	0,09%	
Meta-Metadata.Language												1	1												0	1	1	0,09%	
Technical.Format	1	1		1	1			1		1		1	7							1				1	2	9	81	7,02%	
Technical.Size												1	1												0	1	1	0,09%	
Technical.Location	1											1	2					1						1	3	9	0,78%		
Technical.RequirementCompositeType												1	1												0	1	1	0,09%	
Technical.RequirementCompositeNam												1	1												0	1	1	0,09%	
Technical.Requirement.OrComposite.MinimumVersion												1	1												0	1	1	0,09%	
Technical.Requirement.OrComposite.MaximumVersion												1	1												0	1	1	0,09%	
Technical.Installation Remarks												1	1												0	1	1	0,09%	
Technical.OtherPlatformRequirements												1	1												0	1	1	0,09%	
Technical.Duration												1	1												0	1	1	0,09%	
Educational.Interactivity Type												1	1	1	1					1				3	4	16	1,39%		
Educational.Learning Resource Type	1	1		1	1		1	1	1	1	1	1	10												0	10	100	8,67%	
Educational.Interactivity Level												1	1												0	1	1	0,09%	
Educational.Semantic Density												1	1												0	1	1	0,09%	
Educational.Intended End User Role												1	1												0	1	1	0,09%	
Educational.Context	1	1		1	1		1				1	1	7												0	7	49	4,25%	
Educational.Typical Age Range		1					1					1	3			1								1	4	16	1,39%		
Educational.Difficulty												1	1												0	1	1	0,09%	
Educational.Typical Learning Time												1	1												0	1	1	0,09%	
Educational.Description		1										1	2												0	2	4	0,35%	
Educational.Language												1	1												0	1	1	0,09%	
Rights.Cost	1	1										1	3												0	3	9	0,78%	
Rights.CopyrightandOtherRestrictions	1											1	2												0	2	4	0,35%	
Rights.Description												1	1												0	1	1	0,09%	
Relation.Kind												1	1												0	1	1	0,09%	
Relation.Resource.Identifier.Catalog												1	1												0	1	1	0,09%	
Relation.Resource.Identifier.Entry												1	1												0	1	1	0,09%	
Relation.Resource.Description												1	1												0	1	1	0,09%	
Annotation.Entity												1	1												0	1	1	0,09%	
Annotation.Date												1	1												0	1	1	0,09%	
Annotation.Description												1	1												0	1	1	0,09%	
Classification.Purpose												1	1												0	1	1	0,09%	
Classification.Taxon Path.Source							1					1	2												0	2	4	0,35%	
Classification.Taxon Path.Taxon.Id												1	1												0	1	1	0,09%	
Classification.Taxon Path.Taxon.Entry												1	1												0	1	1	0,09%	
Classification.Description												1	1												0	1	1	0,09%	
Classification.Keyword												1	1												0	1	1	0,09%	

## ANEXO 2: Ejemplo Dublin Core

```
<Meta name "DC.title" content = "Blog de la Biblioteca Universitaria">
<Meta name "dc.Creador" content = "Tapia Flores, Héctor">
<Meta name contenido "DC.Subject" = "004.Ch383 1972" (EJEMPLO DE UN Dewey)
<Meta name contenido "DC.Description" = "Pagina de un Blog Preparada PARA UN taller de TIC">
<Meta name "dc.publisher" content = "Biblioteca Universidad Arturo Prat">
<Meta name "dc.Contribuidor" content = "Unidad de Medios Digitales">
<Meta name contenido "DC.Date" = "2007-07-04">
<Meta name "dc.type" content = "text">
<Meta name "dc.Format" content = "jpg">
```

```

<Meta name contenido "dc.Identifier" = "84-999-8888-1 (ISBN)">
<Meta name contenido "DC.Source" = "2007">
<Meta name = esquema "RFC1766" contenido "DC.Language" = "es">
<Meta name contenido "DC.Relation.isPartOf" = "https://bibliopress.wordpress.com" & gt;
<Meta name contenido "DC.Coverage" = "2007-07-04">
<Meta name "DC.Rights" content = "http://www.ejemplo.cl/derechos_autor.html" & gt;

```

### ANEXO 3: Codificación de la extracción de Repositorios

- Repositorio Oer Commnos

Def soup:

```

def soup(url):
    req = urllib2.Request(url, headers=hdr)
    try:
        response = urllib2.urlopen(req)
        html = response.read()
        return BeautifulSoup(html)
    except urllib2.HTTPError, e:
        print e.fp.read()

```

Def Enlaces

```

def enlaces():
    pag=0
    cont1=0
    enlaces = []

    for p in xrange(0,1):
        html = soup(base_url+"&batch_start="+str(pag))
        for article in html.find_all("article", {"class": "index-item"}):
            menu_enlace = article.div.a.get("href")
            enlaces.append(menu_enlace)
            print menu_enlace
            cont1=cont1+1
            print cont1
        pag=pag+101

    return enlaces

```

Def datos

```

def datos():
    data = []
    data1 = []
    a = []
    link = enlaces()
    cont = 0
    for links in range(len(link)):
        html = soup(link[links])
        if cont >= cont:
            data.append([])
            data1.append([])
            tit = html.title.getText().replace("'", '')
            titulo = tit.split("|")
            div = html.find("div", {"class": "material-details"})
            if div==None:
                continue
            data[cont].append(''+titulo[0]+'')
            data[cont].append(''+link[links]+'')
            asbtract = ''
            tipo = ''
            format = ''
            idioma = ''
            autor = ''
            for subjet in div.findAll("dl"):
                dt = subjet.findAll("dt")
                dd = subjet.findAll("dd")
                for i in range(len(dd)):
                    variable = dt[i].getText()
                    if variable.find('Abstract') >= 0:
                        asbtract = dd[i].getText().replace("'", '').replace('(', '').replace(')', '')
                        asbtract = " ".join(asbtract.split())
                        data1[cont].append(''+asbtract+'')

                    elif variable.find('Material Type') >= 0:
                        tipo = dd[i].getText()
                        tipo = " ".join(tipo.split())
                    elif variable.find('Media Format') >= 0:
                        format = dd[i].getText()
                        format = " ".join(format.split())
                    elif variable.find("Language") >= 0:
                        idioma = dd[i].getText()
                        idioma = " ".join(idioma.split())
                    elif variable.find("Author") >= 0:
                        autor = dd[i].getText()
                        autor = " ".join(autor.split())
                    elif variable.find("Author") == 0:
                        autor = "NULL"

            rating = html.find("div", {"class" : "item-rating" })
            description= rating.div.get('data-rating-value')
            print description
            data[cont].append(''+asbtract+'')
            data[cont].append(''+tipo+'')
            data[cont].append(''+format+'')
            data[cont].append(''NULL'')
            data[cont].append(''NULL'')
            data[cont].append(''+idioma+'')
            data[cont].append(''+autor+'')
            data[cont].append(''NULL'')
            data[cont].append(''+description+'')
            guardar(data1, titulo[0])

        for uso in html.findAll("dd", {"class": "text"}):
            usol = uso.find("div", {"class" : "license-wrapper" })
            txto = ""
            if not(usol is None):
                txto = usol.getText().strip().replace("\n", "")
                data[cont].append(''+txto+'')

        for ul in html.findAll("section", {"class": "item-tags"}):
            tipo = ul.find("ul", {"class" : "list-unstyled" })
            txt9 = ""
            if not(tipo is None):
                for ull in tipo.findAll("li", {"class": "tag-instance keyword"}):
                    txt9 = ull.getText().strip().replace("\n", "")
                    a.append(" ".join(txt9.split()))
                # "Tipo: "+txt9
                aut = " ".join(str(j) for j in a)
                data[cont].append(''+aut+'')
                a = []
            else:
                data[cont].append("NULL")

```

```

cont = cont + 1

```

## Def nombre

```
def nombre(txt_nombre):
    chars = ('/', ':', '*', '?', '"', '<', '>', '\\', '&', '$', '.', ' ')
    for i in chars:
        txt_nombre = txt_nombre.replace(i, '_')
    return txt_nombre
```

## Def guardar

```
def guardar(data, nom):
    f = open("archivos/" + nombre(nom) + '.txt', 'w')
    lista = range(len(data))
    for i in lista:
        datos = "\n".join(str(j.encode('ascii', 'ignore')) for j in data[i])
    print (datos)
    f.write(datos + "\n")
    f.close()
    return "Datos Guardados"
```

## Def Sql

```
def sql(data):
    lista = range(len(data))
    print len(data)
    for i in lista:
        datos = ",".join(str(j) for j in data[i])
        print datos
        if datos==None:
            query = "INSERT INTO materia (`id`, `nombre_materia`,
`url_materia`, `descripcion`, `tipo`, `formato`, `fecha_ingreso`,
`fecha_modificacion`, `idioma`, `autores`, `copyright`,
`creative_commons`, `rubricas`, `tags`) VALUES
(NULL, NULL, NULL, NULL, NULL, NULL, NULL, NULL, NULL, NULL, NULL) "
        else:
            query = "INSERT INTO materia (`id`, `nombre_materia`,
`url_materia`, `descripción`, `tipo`, `formato`, `fecha_ingreso`,
`fecha_modificacion`, `idioma`, `autores`, `copyright`,
`creative_commons`, `rubricas`, `tags`) VALUES (NULL,%s)" % datos
        print query
        run_query(query)
        print "Insertado"

    return "Ingresar Datos"
```

- **Repositorio utpl**

## Def soup

```

def soup(url):
    response = urllib2.urlopen(url)
    html = response.read()
    return BeautifulSoup(html)

```

## Def Enlaces

```

def enlaces(num):
    enlaces = []
    pag = 0
    while pag <= num:
        url = ''.join([pag_url, "browse?type=title&sort_by=1&order=ASC&rpp=100&etal=4&null=%offset="+str(pag)])
        html = soup(url)
        for article in html.find_all("td", {"headers": "t3"}):
            menu_enlace = article.a.get("href")
            curso_url = ''.join([pag_url, menu_enlace])
            enlaces.append(curso_url)
        pag = pag+1
    return enlaces

```

## Def datos

```

def datos(num):
    data = []
    data1 = []
    link = enlaces(num)
    cont = 0
    for links in range(len(link)):
        html = soup(link[links])
        titulo = ""
        if cont >= cont:
            data.append([])
            data1.append([])
            tabla = html.find("table", {"class": "table
itemDisplayTable"})
            titulo = ""
            descripción = ""
            tipo = ""
            fecha_i = ""
            idioma = ""
            autor = ""
            keys = ""
            for tr in tabla.findAll("tr"):
                td = tr.findAll("td", {"class": "metadataFieldValue"})
                tdl = tr.findAll("td", {"class":
"metadataFieldLabel"})
                for i in range(len(tdl)):

```

```

        variable = td1[i].getText()
        if variable.find('Title') >= 0:
            titulo = td[i].getText()
            print titulo.encode('ascii', 'ignore')
            print link[links]
        if variable.find('Description') >= 0:
            descripcion = td[i].getText()
            print          "Description:          "+"
".join(descripcion.split()).encode('utf-8')
            if variable.find('Tipo') >= 0:
                tipo = td[i].getText()
                print tipo
            if variable.find('Issue Date') >= 0:
                fecha_i = td[i].getText()
            if variable.find('metadata.dc.language') >= 0:
                idioma = td[i].getText()
            if variable.find('Authors') >= 0:
                autor = td[i].getText()
            if variable.find('Keywords') >= 0:
                keys = str(td[i]).replace('<br/>', ', ' )
                keys = BeautifulSoup(keys).getText()
                print          "Keys:          "+"          "+"          "+"          "+"
BeautifulSoup(keys).getText()

        data[cont].append('"' + titulo.replace('"', '') + '"')
        data[cont].append('"' + link[links] + '"')
        data[cont].append('"' + "
".join(descripcion.split()).replace('"', '') + '"')
        data1[cont].append('"' + "
".join(descripcion.split()).replace('"', '') + '"')
        data[cont].append('"' + idioma + '"')
        data[cont].append('"' + autor + '"')
        data[cont].append('"' + fecha_i + '"')
        data[cont].append('"Adobe PDF"')
        data[cont].append('"' + keys + '"')
        guardar(data1, titulo)

        cont = cont + 1

    sql(data)
return data

```

## Def nombre

```
def nombre(txt_nombre):
    chars = ('/', ':', '*', '?', '"', '<', '>', '\\', '&', '$', '.', ' ')
    for i in chars:
        txt_nombre = txt_nombre.replace(i, '_')
    return txt_nombre
```

## Def guardar

```
def guardar(data, nom):
    f = open("archivos/" + nombre(nom) + '.txt', 'w')
    lista = range(len(data))
    for i in lista:
        datos = "\n".join(str(j.encode('ascii', 'ignore')) for j in data[i])
    print(datos)
    f.write(datos + "\n")
    f.close()
    return "Datos Guardados"
```

## Def sql

```
def sql(data):
    lista = range(len(data))
    print len(data)
    for i in lista:
        datos = ",".join(str(j) for j in data[i])
        query = "INSERT INTO materias (`id`, `nombre_materia`,
`url_materia`, `descripcion_materia`, `idioma_materia`,
`autor_materia`, `fecha_materia`, `formato_materia`, `tags_materia`)
VALUES (NULL,%s)" % datos
        print query
        run_query(query)
        print "Insertado"

    return "Ingresar Datos"
```

## • Repositorio Unesdoc

### Def soup

```
def soup(url):
    response = urllib2.urlopen(url)
    html = response.read()
    return BeautifulSoup(html)
```

## Def datos

```
def datos():
    data = []
    data1 = []
    cont = 0
    html = soup(base_url)
    for tabla in html.findAll("table", {"class": "bib_table"}):
        if cont >= cont:
            data.append([])
            data1.append([])
            for div in tabla.findAll("td", { "class" : "bib_data
titl_title" }):
                texto1= div.text.encode('ascii', 'ignore')
                data[cont].append("
".join(texto1.replace('"', '').replace('(', '').replace("'", '').replace(
'"', '').replace("?", '').replace(":", '').replace("/", '').replace("(", ''
).replace(")", '').replace("$', '').replace("#", '').replace(')', '').spli
t()))
                data1[cont].append("
".join(texto1.replace('"', '').replace('(', '').replace("'", '').replace(
'"', '').replace("?", '').replace(":", '').replace("/", '').replace("(", ''
).replace(")", '').replace("$', '').replace("#", '').replace(')', '').spli
t()))

                if not (div.a is None):
                    link= div.a.get("href")
                    data[cont].append(link)
                else:
                    data[cont].append("NULL")
                print "link:"+link
                print "titulo:"+texto1
                print data[cont].append("NULL")
                for div2 in tabla.findAll("td", { "class" :
"bib_data orig_uage" }):
                    texto4= div2.text.encode('ascii', 'ignore')
                    data[cont].append(texto4)
                    print "Idioma:"+texto4
                autor = tabla.find("td", { "class" : "bib_data
auth_thor" })
                autor1 =""
                if not(autor is None):
```

```

        autor1 = autor.getText().encode('ascii',
'ignore')

        data[cont].append(autor1)
        print "Autor= "+autor1
    else:
        data[cont].append("NULL")

    for div1 in tabla.findAll("td", { "class" :
"bib_data publ_year" }):
        texto2= div1.text.encode('ascii', 'ignore')
        data[cont].append(texto2)
        print "Ano de Publicacion:"+texto2
    tipo = tabla.find("td", { "class" : "bib_data
docu_type" })

    txt9 = ""
    if not(tipo is None):
        txt9 = tipo.getText().encode('ascii',
'ignore').replace("\n", "")
        data[cont].append(txt9)
        print "Tipo: "+txt9
    else:
        data[cont].append("NULL")
    claves = tabla.find("td", { "class" : "bib_data
main_tors" })

    txt3 = ""
    if not(claves is None):
        txt3 = claves.getText().encode('ascii',
'ignore').replace("\n", "")
        data[cont].append(txt3)
        print "Palabras Claves: "+txt3
    else:
        data[cont].append("NULL")
    guardar(data1, texto1)

    cont = cont + 1
    insertar(data)

print cont
print data
return data

```

## Def insertar

```
def insertar(data):
    for i in range(len(data)):
        datos = "\n".join(str(j.encode('ascii', 'ignore')) for j in
data[i])
        dat = datos.split("\n")
        lista = range(len(dat))
        datIng = []
        print len(dat)
        for i in lista:
            #print dat[i]
            datIng.append(''+dat[i]+'')
        datos = ", ".join(str(j.encode('ascii', 'ignore')) for j in
datIng)

        query = "INSERT INTO materias (`id`, `nombre_materia`,
`url_materia`, `descripcion_materia`, `idioma_materia`,
`autor_materia`, `fecha_materia`, `formato_materia`, `tags_materia`)
VALUES (NULL,%s)" % datos
        print query
        run_query(query)
        return "Datos Insertados"
```

## ANEXO 4: Diccionario LDA ( Stop Word en Ingles)

After	considering	he's	moreover	S	those	within
afterwards	contain	hello	most	Said	though	without
Again	containing	help	mostly	Same	three	won't
Against	contains	hence	much	Saw	through	wonder
ain't	corresponding	Her	must	Say	throughout	would
All	could	here	My	Saying	thru	would
Allow	couldn't	here's	myself	Says	thus	wouldn't
Allows	course	hereafter	N	Second	to	x
Almost	currently	hereby	name	Secondly	together	y
Alone	d	herein	namely	See	too	yes
Along	definitely	hereupon	Nd	Seeing	took	yet
Already	described	hers	near	Seem	toward	you
Also	despite	herself	nearly	Seemed	towards	you'd
although	did	Hi	necessary	Seeming	tried	you'll

Always	didn't	Him	need	Seems	tries	you're
Am	different	himself	needs	Seen	truly	you've
Among	do	His	neither	Self	try	your
amongst	does	hither	never	Selves	trying	yours
An	doesn't	hopefully	nevertheless	Sensible	twice	yourself
And	doing	how	New	Sent	two	yourselves
Another	don't	howbeit	next	Serious	u	z
Any	done	however	nine	Seriously	un	zero
anybody	down	I	No	Seven	under	
Anyhow	downwards	i'd	nobody	Several	unfortunately	
Anyone	during	i'll	Non	Shall	unless	
anything	e	i'm	none	She	unlikely	
Anyway	each	i've	noone	Should	until	
anyways	edu	le	Nor	shouldn't	unto	
anywhere	eg	If	normally	Since	up	
Apart	eight	ignored	Not	Six	upon	
Appear	either	immediate	nothing	So	us	
appreciate	else	In	novel	Some	use	
appropriate	elsewhere	inasmuch	Now	Somebody	used	
Are	enough	Inc	nowhere	Somehow	useful	
aren't	entirely	indeed	O	Someone	uses	
Around	especially	indicate	obviously	Something	using	
As	et	indicated	Of	Sometime	usually	
Aside	etc	indicates	Off	Sometimes	uucp	
Ask	even	inner	often	Somewhat	v	
Asking	ever	insofar	Oh	Somewhere	value	
associated	every	instead	Ok	Son	various	
At	everybody	Into	okayl	Sorry	very	
available	everyone	inward	Old	Specified	via	
Away	everything	Is	On	Specify	viz	
Awfully	everywhere	isn't	Once	Specifying	vs	
B	ex	It	One	Still	w	
Be	exactly	it'd	Ones	Sub	want	
Became	example	it'll	Only	Such	wants	
because	except	it'sl	Onto	Sup	was	
Become	f	its	Or	Sure	wasn't	
becomes	far	itself	Other	T	way	
becoming	few	j	Others	t's	we	
Been	fifth	just	Otherwise	Take	we'd	
Before	first	k	Ought	Taken	we'll	
beforehand	five	keep	Our	Tell	we're	
Behind	followed	keeps	Ours	Tends	we've	
Being	following	kept	Ourselves	Th	welcome	
Believe	follows	know	Out	Tan	well	
Below	for	knows	Outside	Thank	went	

Beside	former	known	Over	Thanks	were	
Besides	formerly	I	Overall	Thanx	weren't	
Best	forth	last	Own	That	what	
Better	four	lately	P	that's	what's	
between	from	later	Particular	Thats	whatever	
Beyond	further	latter	Particularly	The	when	
Both	furthermore	latterly	Per	Their	whence	
Brief	g	least	Perhaps	Theirs	whenever	
But	get	less	Placed	Them	where	
By	gets	lest	Please	Themselves	where's	
C	getting	let	Plus	Then	whereafter	
c'mon	given	let's	Posible	Thence	whereA	
c's	gives	like	Presumably	There	whereby	
Came	go	liked	Probably	there's	wherein	
Can	goes	likely	Provides	Thereafter	whereupon	
can't	going	little	Q	Thereby	wherever	
Cannot	gone	look	Que	Therefore	whether	
Cant	got	looking	Quite	Therein	which	
Cause	gotten	looks	Qv	Teres	while	
Causes	greetings	ltd	R	Thereupon	whither	
Certain	h	m	Rather	These	who	
Certainly	had	mainly	Rd	They	who's	
changes	hadn't	many	Re	they'd	whoever	
Clearly	happens	may	Really	they'll	whole	
Co	hardly	maybe	REAonably	they're	whom	
Com	has	me	Regarding	they've	whose	
Come	hasn't	mean	Regardless	Think	why	
Comes	have	meanwhile	Regards	Third	will	
concerning	haven't	merely	Relatively	This	willing	
consequently	having	might	Respectively	Thorough	wish	
consider	he	more	Right	Thoroughly	with	

Fuente: ("Stopwords," 2016)

### Anexo 5. Instrucciones para la ejecución del aplicativo web.

1. Instalar el Xampp
2. Ingresar a la consola de Xampp y levantar los servicios Apache y Mysql
3. Copiar la carpeta Buscador en el siguiente directorio C:\xampp\htdocs
4. Desde phyadmin de Xampp crear una base de datos con el nombre "prueba"
5. Exportar *en la base de datos prueba* la tabla REA.sql que encontraremos en el archivo REA.
6. Para acceder a la *aplicación* copiar la siguiente *dirección* en el navegador se su preferencia : <http://localhost/Buscador/>