



# **UNIVERSIDAD NACIONAL DE ROSARIO**

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Licenciatura en Ciencias de la Computación

Tesina de Grado

## **Análisis de calidad y confianza en los metadatos de objetos de aprendizaje recuperados**

Damián Alvarez Machicote

Directoras de Tesina:  
Cristina Bender      Claudia Deco

**Abril 2011**

## Resumen

El importante crecimiento que ha tenido Internet y en particular la World Wide Web, se debe a su habilidad para facilitar la posibilidad de compartir e intercambiar información y recursos. Con la necesidad de tener acceso a esta gran cantidad de información han surgido diversos problemas en su recuperación.

Este problema también sucede en el escenario de los repositorios educacionales. Sin embargo aquí no sólo se busca la documentación relevante sino que también se exige una cuidada calidad de contenido y que la misma presente ciertas características de interés o de necesidad para el usuario. Los elementos a recuperar se llaman objetos de aprendizaje, los cuales son elementos para la instrucción, aprendizaje o enseñanza basada en computadora. Los mismos se encuentran almacenados en contenedores conocidos como Repositorios de Objetos de Aprendizaje donde el objetivo principal es organizarlos y mantenerlos disponibles para diferentes usos.

En este trabajo se propone medir la calidad y confiabilidad de objetos de aprendizaje con el fin de que los recursos electrónicos de mayor calidad y confianza escalen las primeras posiciones en los resultados de búsqueda y que además se personalicen al perfil del usuario que realiza la búsqueda.

La medición de la calidad de los objetos de aprendizaje es llevada a cabo por un fuerte y completo análisis de los metadatos siguiendo el estándar de metadatos LOM (Learning Object Metadata) y la medición de la confiabilidad se obtiene mediante el análisis de las calificaciones otorgadas por los usuarios.

Para la implementación de este trabajo se ha desarrollado una aplicación web prototipo que permite realizar búsquedas en un repositorio mediante palabras claves obteniendo de esta forma objetos de aprendizaje con una cuidada calidad y que se adapten de una manera más precisa al perfil del usuario que realiza la búsqueda. El objetivo principal será la verificación de la mejora de calidad y confiabilidad de los objetos de aprendizaje recuperados.

## **Agradecimientos**

Quiero agradecer principalmente a mi madre, Graciela, que gracias a ella y a su gran esfuerzo tuve la oportunidad de venir a vivir a Rosario, estudiar y terminar esta carrera.

A mis hermanas y abuela por su apoyo brindado en cada año de esta carrera universitaria, ayudándome en cada caída y festejando cada logro obtenido.

A Florencia, una gran y excelente persona que me acompaña desde hace ya varios años. Que gracias a su apoyo, consejos y paciencia supo soportar mis autoexigencias con el objetivo de obtener este título universitario.

A mis amigos de la carrera, Mariano Salvetti, Marcos Belén, Fernando Nicola (Nano) y Hugo Gercek (Cuchi), por toda su ayuda, charlas, consejos, almuerzos, por todos los días de estudio compartidos y por su participación fundamental en la experimentación de este trabajo.

A mis dos directoras de Tesina, Claudia Deco y Cristina Bender, por su ayuda, dedicación y orientación en este trabajo. Por siempre estar atentas y disponibles a dudas o consultas. Por su empeño y énfasis en la revisión de redacción de este trabajo.

A todos mis amigos, compañeros que me acompañan día a día y que son parte de mi vida.

A todos ustedes quiero dedicarles este trabajo.

# Índice General

Introducción .....	8
<b>1. Conceptos Preliminares .....</b>	<b>9</b>
1.1 Objetos de Aprendizaje .....	9
1.1.1 Definición.....	9
1.1.2 Granularidad.....	10
1.1.3 Reutilización de los OA.....	11
1.2 Metadatos .....	11
1.2.1 Definición.....	11
1.2.2 Estándar LOM.....	12
1.3 Repositorios de Objetos de Aprendizajes .....	14
1.3.1 Definición.....	14
1.3.1 Repositorio Ariadne .....	15
1.4 Recursos Lingüísticos .....	17
1.4.1 Wordnet.....	17
1.5 Otros Recursos .....	18
1.5.1 StopWords.....	18
1.5.2 N-gramas .....	18
<b>2. Calidad en los objetos de aprendizajes .....</b>	<b>19</b>
<b>3. Propuesta .....</b>	<b>26</b>
3.1 Proceso Metadatos.....	31
3.1.1 Sistema de Puntuación.....	31
3.1.2 Ponderación de las Categorías .....	34
3.2 Proceso Keywords.....	38
3.3 Proceso Idiomas .....	45
3.4 Proceso Confiabilidad .....	51
<b>4. Implementación de la propuesta .....</b>	<b>54</b>
4.1 Tecnologías utilizadas .....	54
4.2 Repositorios utilizados .....	54
4.3 Prototipo.....	54
4.3.1 Arquitectura.....	54
<b>5. Experimentación .....</b>	<b>56</b>
5.1 ¿Los objetos de aprendizaje de mayor calidad tienden a ubicarse en las primeras posiciones?.....	56
5.2 ¿Los objetos de aprendizaje se adaptan y se personalizan al perfil de usuario que realiza la consulta? .....	58

Se realizaron las siguientes búsquedas .....	58
<b>6. Discusiones</b> .....	62
6.1 Repositorios .....	62
6.2 Prueba del Prototipo .....	62
6.3 Trabajo Futuro.....	63
<b>7. Bibliografía</b> .....	64
<b>8. Apéndice</b> .....	67
8.1 Experimentaciones .....	67
8.1.1 Consultas: Calificaciones promedios .....	67
8.1.2 Distancias de resultados entre perfiles .....	69
8.1.3 Distancias y análisis de ranking de los usuarios .....	70

# Índice de Tablas

2.1. Formulario para la determinación de la calidad de un OA Repositorios de Objetos de Aprendizajes .....	22
2.2. Metadatos LOM y nuevos metadatos agregados .....	23
2.3. Metadatos LOM y nuevos metadatos agregados6. Conclusiones y Trabajos futuros ...	24
3.1. Criterio de búsqueda y metadatos asociados .....	35
3. 2. Metadatos más solicitados por los profesores (Teacher).....	35
3. 3. Ejemplos de entrenamientos y total y tamaño de N-gramas .....	47
3. 4. Cálculo de distancia utilizando Rank-Order.....	47
5.1. Calificaciones promedios obtenidas en las 10 búsquedas .....	57
5.2. Usuarios con diferentes perfiles y objetivos.....	58
5.3. Distancia de <i>Manhattan</i> entre los resultados de búsquedas entre perfiles .....	60
5.4. Distancias de los usuarios con perfil <i>Teacher</i> en el caso A y B .....	61
5.5. Distancias de los usuarios con perfil <i>Student</i> en el caso A y B .....	61

# Índice de Figuras

1.1. Composición conceptual de un OA.....	10
1.2. OAs de distinta granularidad.....	10
1.3. Creación de un nuevo OA a partir de la utilización de otros.....	11
1.4. Categorías de metadatos LOM.....	14
1.5. Interfaz de búsqueda de <b>Ariadne</b> .....	16
1.6. Número de palabras y conceptos para la división de sustantivos en WordNet 3.0.....	17
2.1. Elementos que determinan la calidad de un OA.....	20
2.2. Elementos pedagógicos de un OA.....	21
2.3. Elementos de contenido de un OA.....	21
2.4. Elementos estéticos de un OA.....	21
3.1. Estrategia para determinar la calidad de los OA.....	26
3.2. Interfaz de usuario para realizar búsqueda mediante palabras claves.....	26
3.3. Formulario de Log In.....	27
3.4. Formulario de Nuevo Usuario.....	27
3.5. Ejemplo de un resultado de búsqueda.....	29
3.6. Estadísticas de los resultados de búsqueda.....	30
3.7. Proceso de análisis de los metadatos.....	31
3.8. Etapas del proceso keywords.....	38
3.9. Conjuntos de palabras a analizar.....	42
3.10. Etapas del proceso Idiomas.....	45
3.11. Interfaz gráfica para realizar la votación.....	52
4.1. Arquitectura.....	55
5.1. Gráfico de las calificaciones obtenidas para los diferentes casos.....	57

# Introducción

Como consecuencia de la evolución de la Web Semántica, la gestión de la información en sistemas e-learning está cambiando. Un ejemplo de esto son los Objetos de Aprendizaje (OA), concepto que corresponde a un conjunto de recursos que podrían ser utilizados como unidades independientes y reutilizables en contextos y plataformas diferentes. Lo cual despierta un gran atractivo para las instituciones que se dedican a la entrega de una educación de calidad en un mercado cada vez más competitivo. Cada objeto de aprendizaje tiene metadatos (datos sobre datos) para su descripción y la administración del mismo. Los metadatos permiten potenciar a los OAs como recursos educativos recuperables, localizables, intercambiables y reutilizables. Su importancia radica en que a través de ellos se puede llevar a cabo el primer acercamiento con el objeto y conocer rápidamente sus principales características y de esta manera saber con qué tipo de OA estamos tratando, pero esto no significa que estamos tratando con un contenido de calidad sino solamente con un contenido normalizado y estandarizado.

La evaluación de los OAs es un elemento importante que incide, tanto en el diseño, como en la comprensión de sus valores utilitarios en los procesos de instrucción y de ayuda al aprendizaje. Existe una gran cantidad de criterios de calidad para valorar las fuentes digitales, pero sólo hay unas pocas sugerencias acerca de cómo evaluar la estructura de los OAs a fin de determinar la calidad. La calidad y confiabilidad de un OA puede abordarse desde distintas perspectivas y en diversos momentos de su propio ciclo de vida desde la calidad de sus metadatos hasta la calidad del aprendizaje obtenido. Nuestra propuesta consiste en un sistema para evaluar los OAs considerando la calidad de sus metadatos como descriptores reutilizables y la confiabilidad del mismo mediante la participación de usuarios finales (profesores y estudiantes).

En la Sección 1 se detallan definiciones de conceptos que son utilizados en las demás secciones. De acuerdo con esto en la Sección 2 se describen y definen criterios relacionados a la calidad de los OAs y se explican las diferentes propuestas (estado del arte) que tratan este tema y en qué se diferencia nuestra propuesta de las existentes. En la Sección 3 se detalla nuestra propuesta dónde se sugiere la evaluación de los OAs teniendo en cuenta diferentes procesos o módulos que consideren los criterios de calidad de diferentes puntos de vista y una metodología de evaluación para cada uno de ellos. Para facilitar esta evaluación nos basamos en el estándar LOM de metadatos y en la participación de usuarios finales. En la Sección 4 se describen las tecnologías utilizadas para llevar a cabo la implementación. En la Sección 5 se describen las distintas experimentaciones y pruebas realizadas. Por último en la Sección 6 se detallan las conclusiones obtenidas y trabajos futuros.



# 1. Conceptos Preliminares

En esta Sección se presentan definiciones, conceptos y otros aspectos de interés a esta tesina que serán utilizados en las demás secciones.

## 1.1 Objetos de Aprendizaje

### 1.1.1 Definición

Según [Wiley, 2002] un Objeto de Aprendizaje (OA) es todo recurso digital que apoya a la educación y que puede ser reutilizado. Por otro lado el Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) define un Objeto de Aprendizaje como una entidad, digital o no digital, que puede ser usada para aprendizaje, educación o entrenamiento [IEEE, 2002]. El concepto de Objeto de Aprendizaje (en inglés Learning Object) abarca principalmente a un conjunto de materiales digitales los que como unidad o agrupación permiten o facilitan alcanzar un objetivo educacional. Ejemplos de los recursos digitales más pequeños incluyen a imágenes o fotos, cortos de video o audio, pequeñas porciones de texto, ecuaciones, definiciones, animaciones, pequeñas aplicaciones web, entre otros. Recursos digitales de mayor tamaño son páginas web completas que combinen texto, imágenes y otros medios de comunicación. Es decir, los objetos pueden adquirir formas muy diversas y presentarse en diferentes formatos y soportes.

Estos objetos han de contar con las siguientes características destacadas, según [García, 2005] para alcanzar su máxima calidad en sus amplios sentidos:

- **Reutilización**, objeto con capacidad para ser usado en contextos y propósitos educativos diferentes y para adaptarse y combinarse dentro de nuevas secuencias formativas.
- **Educatividad**, objeto con capacidad para generar aprendizaje.
- **Interoperabilidad**, objeto con capacidad para integrarse en estructuras y sistemas (plataformas) diferentes.
- **Accesibilidad**, facilidad para ser identificados, buscados y encontrados gracias al correspondiente etiquetado a través de diversos descriptores (metadatos) que permiten la catalogación y almacenamiento en los correspondientes repositorios.
- **Durabilidad**, vigencia de la información, sin necesidad de nuevos diseños.
- **Independencia y autonomía** de los objetos con respecto de los sistemas desde los que fueron creados y con sentido propio.
- **Generatividad**, capacidad para construir contenidos, objetos nuevos derivados de él. Capacidad para ser actualizados o modificados aumentando sus potencialidades a través de la colaboración.
- **Flexibilidad, versatilidad y funcionalidad**, con elasticidad para combinarse en muy diversas propuestas de áreas del saber diferentes.

Un elemento fundamental para la explotación de un OA es el metadato. Aunque las primeras definiciones que se dieron al término OA se enfocaban a recursos en general y a su reutilización, definiciones más actuales apuntan a la importante relación entre el objeto y sus metadatos, al afirmar que “un OA es cualquier recurso que puede ser utilizado para facilitar la enseñanza y el aprendizaje y que ha sido descrito utilizando metadatos” (véase la Figura 1.1). De esta manera, nace una relación intrínseca entre un objeto educativo y sus metadatos para formar, en conjunto, un OA. Es decir, si un recurso digital no tiene metadatos no puede considerarse OA ya que, en la práctica, los sistemas no pueden interpretarlo (o reconocerlo) y hacerlo realmente reutilizable. En la Sección 1.2 se profundiza este tema.



Figura 1.1: Composición conceptual de un OA

### 1.1.2 Granularidad

Aunque se menciona que un OA es “una pieza pequeña” o un recurso “modular” no se puede especificar una dimensión precisa. El tamaño de un OA es variable y esto se conoce como granularidad.

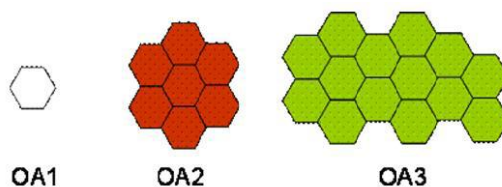


Figura 1.2: OAs de distinta granularidad

En la Figura 1.2 se han representado, a través de hexágonos, lo que pueden ser unidades de contenido o elementos que componen al OA. Por ejemplo, el OA1 podría ser una imagen y el OA2 podría ser una página web que incluye texto e imágenes. El OA3 puede ser un recurso multimedia en el que se incluyen más unidades de contenido que en los objetos anteriores.

No es posible definir la cantidad de información o elementos que un OA debe contener, esto dependerá de las necesidades y habilidades del autor para trabajar y conceptualizar trozos de contenidos que irán formando un curso. Sin embargo, según [Duncan, 2003], el mejor criterio para definir la granularidad de un objeto es por sus propósitos u objetivos. Se considera una buena práctica que los OAs cubran un único objetivo de aprendizaje y para lograrlo deben mantener independencia del contexto y no requerir de otros recursos, es decir, que sean autosuficientes.

### 1.1.3 Reutilización de los OA

La característica más notable en las diferentes definiciones de los OAs es la reutilización. Lograr la reutilización requiere tener un diseño, un desarrollo y una documentación que aseguren un alto nivel en la calidad del producto y pueda éste trabajar de forma sencilla con otros. La reutilización de un contenido aumenta su valor y produce ahorro, en diferentes sentidos, a nivel institucional o individual. El gran potencial de la reutilización de los OA es poder aprovechar los contenidos que han desarrollado otros para formar nuevos recursos. En la Figura 1.3 se ejemplifica cómo a partir de tres OA independientes se genera otro nuevo.

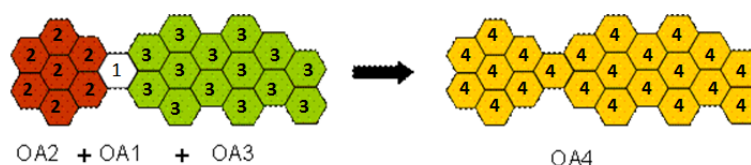


Figura 1.3: Creación de un nuevo OA a partir de la utilización de otros

Para la reutilización es necesario que el OA cuente con los metadatos de identificación, organizado y recuperado, pero lo más importante es que esos metadatos estén basados en un estándar, a fin de asegurar su compatibilidad e interoperabilidad con los sistemas que hacen uso de ellos.

## 1.2 Metadatos

### 1.2.1 Definición

Aunque el término *metadato* no es reciente, en los últimos años su popularidad se ha disparado debido a la creciente necesidad de identificar, organizar y almacenar grandes cantidades de información digital. [Hillman, 2003] señala que los metadatos han estado presentes desde que los primeros bibliotecarios crearon listas de recursos de información, y destaca que el término “meta” proviene del griego que significa “al lado de, siguiente, después, con”. Dentro del campo informático, [Caplan, 2003] documenta el nacimiento del término en las ciencias computacionales, en donde el prefijo “meta” significa “acerca de”; así, por ejemplo, un *metalenguaje* es un lenguaje utilizado para describir otro lenguaje. En el comienzo de los años noventa el término *metadato* se utilizaba como “datos acerca de los datos” para identificar archivos digitales de conjuntos de datos científicos, sociales y geoespaciales. La aparición de la web ha dado a los metadatos nuevas áreas de aplicación en donde juegan un papel primordial, principalmente para localizar recursos en Internet. Es común que los metadatos sean vistos como descriptores. Sin embargo, éstos no sólo cumplen funciones de identificación, pues también pueden contener información con fines administrativos y estructurales. En el campo de objetos de aprendizaje los metadatos son un conjunto de atributos necesarios para describir las principales características de un recurso. A través

de los metadatos se tiene un primer acercamiento con el objeto, conociendo rápidamente sus principales características. Se utilizan con el propósito de poder encontrar, gestionar, reusar y almacenar objetos de aprendizaje de manera efectiva. Son especialmente útiles en los recursos que no son textuales, por ejemplo los multimedia, y en los que su contenido no puede ser indexado por sistemas automáticos.

[Caplan, 2003] clasifica los tipos de metadatos como:

- **Metadatos descriptivos:** Su propósito es identificar cómo un recurso puede distinguirse de otro, describir cómo se encuentra un recurso, y seleccionar recursos que cubran necesidades particulares.
- **Metadatos administrativos:** Su fin es anotar información que facilite la gestión de los recursos, lo que incluye información sobre cuándo y cómo fue creado, quién es el responsable del acceso o de la actualización del contenido, e información técnica, como la versión de software o hardware necesarios para ejecutar dicho recurso.
- **Metadatos estructurales:** Su objetivo es identificar cada una de las partes que componen al recurso, definiendo la estructura que le da forma.

## 1.2.2 Estándar LOM

Learning Object Metadata (LOM) es un estándar que especifica Metadatos para Objetos Educativos [IEEE, 2002]. Especifica un esquema conceptual de datos que define la estructura de una instancia de metadatos para un objeto educativo. Para este estándar, un objeto educativo se define como *cualquier entidad, digital o no, susceptible de ser usada en aprendizaje, educación o formación.*

En lo que respecta a este estándar, una instancia de metadatos para un objeto educativo describe las características relevantes del objeto educativo al que se aplica. Dichas características se pueden agrupar en categorías como se puede observar en la Figura. 1.4.

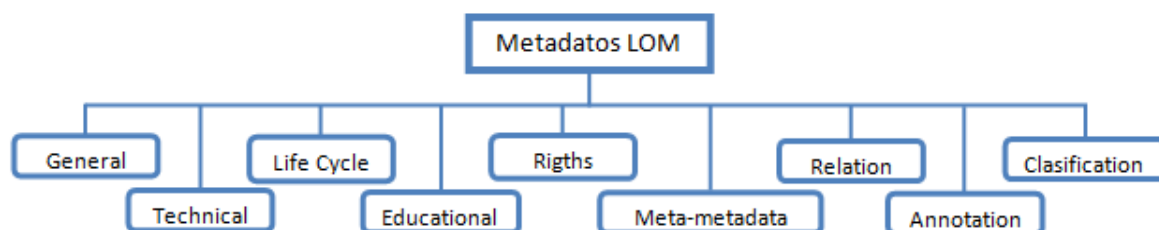


Figura 1.4: Categorías de metadatos LOM

1. **General (General):** Agrupa la información general que describe un objeto educativo de manera global.

2. **Life Cycle (Ciclo de vida):** Agrupa las características relacionadas con la historia y el estado actual del objeto, y aquellas que le han afectado durante su evolución.
3. **Meta-metadata (Meta-metadatos):** Agrupa la información sobre la propia instancia de Metadatos (en lugar del objeto educativo descrito por la instancia de metadatos).
4. **Technical (Técnica):** Agrupa los requerimientos y características técnicas del objeto educativo.
5. **Educational (Educación):** Agrupa las características educativas y pedagógicas del objeto educativo.
6. **Rights (Derechos):** Agrupa los derechos de propiedad intelectual y las condiciones para el uso del objeto educativo.
7. **Relation (Relación):** Agrupa las características que definen la relación entre este objeto educativo y otros objetos educativos relacionados.
8. **Annotation (Anotación):** Permite incluir comentarios sobre el uso educativo del objeto e información sobre cuándo y por quién fueron creados dichos comentarios.
9. **Classification (Clasificación):** Describe este objeto educativo en relación a un determinado sistema de clasificación.

Las categorías agrupan metadatos. Las categorías *Educación* y *Ciclo de vida* contienen metadatos descriptivos. *Técnica* y *Derechos* son administrativos y las categorías *Clasificación* y *Relación* son estructurales. El modelo de datos de LOM es una jerarquía de metadatos, incluyendo metadatos agregados y simples (nodos hoja en la jerarquía). En el esquema base LOMv1.0 sólo los nodos hoja tienen valores individuales definidos a través de sus espacios de valores y tipos de datos asociados. En el esquema base los elementos de datos agregados no tienen valores individuales, por lo tanto, no tienen espacios de valores o tipos de datos. Para cada metadato el esquema define:

- **nombre:** el nombre de referencia del metadato.
- **explicación:** la definición del metadato.
- **tamaño:** el número de valores permitido.
- **orden:** relevancia de la ordenación de los valores.
- **ejemplo:** un ejemplo ilustrativo.
- **espacio de valores:** el conjunto de valores permitidos para el metadato. Normalmente en forma de un vocabulario o una referencia a otro estándar.
- **tipo de datos:** indica si los valores son *LangString*, *Fecha*, *Duración*, *Vocabulario*, *CharacterString* o *No Definido*.

Existen metadatos que contienen una lista de posibles valores (Vocabulario), en vez de un único valor. La lista debe ser de uno de los siguientes tipos:

- **Ordenada:** el orden de los valores en la lista es relevante.
- **No ordenada:** el orden de los valores no tiene relevancia ni significado.

Un vocabulario es una lista “recomendada” de valores apropiados. Se pueden usar también otros valores no incluidos en la lista. Sin embargo, los metadatos que se ajustan a los valores recomendados tendrán el máximo grado de interoperabilidad semántica y la probabilidad de que otros usuarios o sistemas comprendan dichos metadatos será la máxima posible.

También se define un mínimo para los valores máximos permitidos (tamaño) en los siguientes elementos:

- **elementos de datos agregados:** Todas las aplicaciones que procesen instancias de LOM deberán soportar al menos dicho número de entradas. En otras palabras, una aplicación puede imponer un máximo al número de entradas que permite para dicho elemento de datos, pero dicho máximo no puede ser inferior que el valor asignado al menor máximo permitido.
- **elementos de datos de tipo *CharacterString* o *LangString*:** Todas las aplicaciones deberán ser capaces de procesar al menos dicha longitud para el valor del *CharacterString* (bien directamente o en el contenido del *LangString*) del metadato. En otras palabras, una aplicación puede imponer un valor máximo al número de caracteres que procesa para el valor *CharacterString* del metadato, pero dicho máximo no podrá ser inferior al menor máximo permitido definido para el tipo del metadato.

El esquema conceptual de datos definido en este estándar permite la diversidad lingüística tanto de los objetos educativos como de las instancias de metadatos que los describan.

El propósito de este estándar es facilitar la búsqueda, evaluación, adquisición y uso de los objetos educativos, por ejemplo, por alumnos, profesores o procesos automáticos de software. Este estándar también facilita el intercambio y uso compartido de objetos educativos, permitiendo el desarrollo de catálogos e inventarios al tiempo que se toman en consideración la diversidad cultural y los contextos lingüísticos en los que los objetos educativos y sus metadatos serán reutilizados.

## 1.3 Repositorios de Objetos de Aprendizajes

### 1.3.1 Definición

Los OAs además de su característica de reutilización, han de contar con la posibilidad de ser actualizados, combinados, separados, referenciados y sistematizados

de manera que puedan ser clasificados o catalogados y etiquetados para ser ubicados en los correspondientes almacenes o repositorios de contenidos u objetos de aprendizajes, con el fin de que posteriormente puedan ser localizados para su reutilización, o si procede, modificación o reelaboración, mediante las correspondientes estrategias de contraste, comparación, relación y crítica de la información obtenida. Por eso resulta la necesidad de potentes repositorios de objetos de aprendizaje.

Por lo tanto un Repositorio de Objetos de Aprendizaje es una gran colección de OAs, estructurada como una base de datos con metadatos asociados y que generalmente se puede encontrar en la Web. Los objetos y el repositorio son complementarios. Un objeto que no guarde las características necesarias para poder integrarse en un repositorio pierde todas sus virtudes y, a su vez, un repositorio que no cuente con una buena base de objetos deja de ser operativo. La utilización de metadatos facilita la indexación de los OAs.

Algunos ejemplos de repositorios son:

1. MERLOT (Multimedia Educational Resource for Learning and Online Teaching) <http://www.merlot.org/>.
2. CAREO (Campus Alberta Repository of Educational Objects) <http://www.careo.org/>.
3. FLOR (Federación Latinoamericana de Repositorios). <http://www.laclo.espol.edu.ec/>.
4. OER Commons (Open Educational Resources). <http://www.oercommons.org/>.
5. SMETE (Science, Mathematics, Engineering and Technology Education). <http://www.smete.org/smete/>
6. Ariadne. <http://ariadne.cti.espol.edu.ec/FederatedClient>.

En este trabajo se utilizó el repositorio de OAs **Ariadne** ya que resultó ser el más conveniente para la realización de consultas y extracción de objetos relevantes debido a la existencia de un gran número de objetos de aprendizaje en idioma inglés con metadatos bajo el estándar LOM, por trabajar con un lenguaje de consulta expresivo llamado “Lucene”, por tener la posibilidad de realizar búsqueda mediante palabras claves y por ser gratuito. En el siguiente apartado se describen sus características y arquitectura.

### 1.3.1 Repositorio Ariadne

**Ariadne** (nombrada así por la figura de la mitología Griega) es una asociación Europea que permite compartir conocimiento y fomenta la cooperación internacional en enseñanza, abierta a todo el mundo. La infraestructura ha sido desarrollada en Bélgica y Suiza para la producción de contenido de aprendizaje reutilizable. El núcleo de esta infraestructura es una biblioteca digital distribuida, que contiene componentes educacionales reusables, es decir es un repositorio de objetos de aprendizaje distribuido, llamado *Knowledge Pool System* (KPS) que actualmente es usado en contextos

académicos y empresariales. Los usuarios finales interactúan con éste a través de clientes Java, web services y aplicaciones web que permiten a los usuarios insertar documentos y sus metadatos asociados en el KPS, y buscar documentos relevantes. Las aplicaciones web interactúan con éste a través de los servicios web de **Ariadne**. Desde el punto de vista técnico el KPS consiste de una base de datos relacional que contiene los metadatos, y de un repositorio de documentos que contiene los objetos educativos. **Ariadne** no se enfoca en objetos de aprendizaje de ningún campo particular.

**Ariadne** consiste en una arquitectura de 3 capas:

- En la capa inferior, el repositorio permite buscar, publicar y recuperar OAs.
- Una API, que está conectada a servicios web, permite la manipulación de esta capa inferior.
- Aplicaciones tales como *Moodle plugin* o *ALoCom Office plugin* permiten acceder de manera transparente a la información a través de aplicaciones externas.

Un cliente puede navegar en **Ariadne** a través de la interface de consulta SQI (Simple Query Interface). Por otro lado **Ariadne** permite búsquedas federadas; algunos de los repositorios de esta federación con los que trabaja son: LACLO-FLOR, Merlot y OER Commons.

**Ariadne** entre sus estándares de modelos de datos, utiliza el estándar LOM. En la Figura 1.5 podemos ver una de las interfaces de búsqueda de **Ariadne**.



Figura 1.5: Interfaz de búsqueda de **Ariadne**



Esta interfaz es un buscador que permite encontrar OAs en el repositorio local o federado y navegar los resultados por categorías como así también ver los metadatos de los objetos. En la imagen podemos observar que a la izquierda presenta criterios de búsqueda que permiten filtrar los OAs recuperados por tipo de recurso, formato, contexto, idioma y autor; a la derecha presenta una lista de repositorios externos para la realización de búsquedas federadas. Además si se cuenta con los permisos se puede descargar el objeto.

## 1.4 Recursos Lingüísticos

### 1.4.1 Wordnet

WordNet es un sistema léxico construido manualmente por George Miller y sus colegas en el laboratorio de Ciencias Cognitivas de la Universidad de Princeton [Miller, 2010]. Originario de un proyecto cuyo objetivo era construir un diccionario en el que se pudiera buscar de manera conceptual en lugar de solo alfabéticamente, WordNet evolucionó en un sistema que refleja las actuales teorías psicolingüísticas acerca de cómo los seres humanos organizan sus recuerdos léxicos. El objeto básico es un conjunto de sinónimos exactos llamados *synsets*. Por definición, cada *synset* representa un sentido o significado diferente de una palabra, es decir, cada *synset* representa un concepto.

Hay cuatro divisiones principales en WordNet, para sustantivos, verbos, adjetivos y adverbios. Hay un rico conjunto de enlaces que permite la relación entre palabras, palabras y *synsets*, y también entre *synsets*. Dentro de una división, los *synsets* están organizados por las relaciones léxicas definidas sobre ellos. Para los sustantivos, la única división de WordNet utilizada en este trabajo, las relaciones léxicas incluidas son: sinonimia y hiperonimia/hiponimia (relación IS-A). La relación IS-A es la relación dominante, y organiza a los *synsets* en un conjunto de aproximadamente diez jerarquías.

La versión 3.0 contiene 117798 sustantivos, organizados en 82115 *synsets* como se observa en la siguiente tabla (Figura 1.6).

POS	Unique Synsets		Total
	Strings		Word-Sense Pairs
Noun	117798	82115	146312
Verb	11529	13767	25047
Adjective	21479	18156	30002
Adverb	4481	3621	5580
Totals	155287	117659	206941

Figura 1.6: Número de palabras y conceptos para la división de sustantivos en WordNet 3.0

Este recurso se utiliza en el **proceso Keywords** que se describe en la Sección 3.2.

## 1.5 Otros Recursos

### 1.5.1 StopWords

Las *stopwords* son palabras irrelevantes y sin significado que no deberían ser consideradas palabras claves de un texto determinado. Ejemplos de stopwords son los artículos, pronombres, preposiciones, etc. Las mismas son filtradas antes o después del procesamiento de datos en lenguaje natural (texto).

Este recurso se utiliza en el **proceso Keywords** que se describe en la Sección 3.2.

### 1.5.2 N-gramas

Un **N-grama** es una subsecuencia de  $n$  elementos de una secuencia dada. Los espacios en blanco son considerados como carácter. Se añaden espacios en blanco al principio y al final de la secuencia con el fin de ayudar a la igualación entre el comienzo y el final de la palabra. De este modo los N-gramas de inicio y final son considerados de igual forma que si estuviesen por el medio.

Por ejemplo:

La palabra TEXT estará compuesta por los siguientes N-gramas:

- Para  $N = 2$  **Bi-gramas:** \_T, TE, EX, XT, T\_
- Para  $N = 3$  **Tri-gramas:** \_\_T, \_TE, TEX, EXT, XT\_, T\_\_

Existen algunos idiomas que tienen la característica de poseer alguna palabra en su lenguaje de forma que dicha palabra ocurre con más frecuencia que otras palabras. Una de las maneras más comunes para expresar esta idea se conoce como la *ley de Zipf* [Zipf, 1949].

Particularmente esta ley es verdad para la ocurrencia de N-gramas más comunes para un determinado idioma. Por ende los N-gramas son utilizados frecuentemente por algoritmos de detección de idiomas.

Este recurso se utiliza en el **proceso Idiomas** que se describe en la Sección 3.3.

## 2. Calidad en los objetos de aprendizajes

El interés por la generación de objetos de aprendizaje con una cuidada calidad de contenido actualmente va en aumento en muchas instituciones educativas y corporaciones tanto públicas como privadas a nivel nacional e internacional, esto debido a que ven en este nuevo medio informático una poderosa forma de armar cursos de educación a distancia y auto-aprendizaje que seguramente serán fundamentales en los procesos educativos del futuro. En una investigación desarrollada en la Universidad Autónoma de Aguascalientes, a partir del año 2006 se ha comenzado un profundo estudio sobre la calidad en objetos de aprendizaje, tanto para determinar claramente la amplitud del problema como para proponer estrategias en un aseguramiento óptimo de la calidad de contenido de los mismos.

La calidad en objetos de aprendizaje puede abordarse desde distintas perspectivas y en diversos momentos de su propio ciclo de vida. Por un lado tenemos la parte tecnológica del OA, la cual es semejante a la de cualquier otra aplicación de software. La calidad del software es el cumplimiento de los requisitos de funcionalidad, desempeño explícitamente establecidos y documentados, y también el cumplimiento de las características implícitas que se esperan de todo software desarrollado profesionalmente [Pressman, 2000]. Además podemos indicar que un producto de software posee calidad de diseño y calidad de concordancia. Con relación a los objetos de aprendizaje tenemos que la calidad de diseño se refiere a todo lo relacionado con el diseño del objeto de aprendizaje como es la selección de metadatos adecuados y de los elementos que lo deben integrar (textos explicativos, gráficas, fotografías, música, texto hablado, videos, animaciones, evaluaciones, módulos de experimentación, hipervínculos) para que sea eficaz en el proceso de aprendizaje dependiendo del contexto en que se aplique. La calidad de concordancia se refiere al grado de cumplimiento de los estándares y recomendaciones definidas para el OA.

Con relación a la calidad en la parte educativa del OA, tenemos que esta se cumplirá en mayor grado cuanto más se logre el objetivo de obtener un aprendizaje significativo en el usuario, que en este caso es el alumno o estudiante. Un aspecto medular para lograr lo anterior es asegurando la calidad de contenido del OA.

Un problema importante en el manejo de recursos educativos es el de la baja eficiencia en la recuperación de información, como lo manifiestan Ben Ryan y Steve Walmsley en [Ryan, Walmsley, 2003] dónde señalan que:

“Si usted no puede buscar un recurso educativo porque no tiene el metadato, o una búsqueda devuelve varios cientos o miles de resultados, usted no puede reusar el recurso porque no puede localizarlo o no puede decidir qué recurso es pertinente a sus necesidades debido al tiempo requerido para evaluar los resultados de la búsqueda.”

Actualmente no se cuentan con suficientes medios efectivos para asegurar la calidad de contenido en OA, por ello a continuación se detallan los problemas o peligros que pueden encontrarse:

- Posibilidad de enseñar contenidos obsoletos.
- Peligro de exponer al estudiante a material malintencionado.
- Desconfianza en la adquisición de contenidos educativos electrónicos, con lo que se reduce la posibilidad de comercialización de los mismos.

Algunos de los beneficios de un estudio detallado de la calidad de contenido en OA serían los siguientes:

- Crear estándares de calidad para OA.
- Contar con metadatos más robustos y descriptivos desde el punto de vista de la calidad de contenido.
- Incrementar posibilidades de éxito tanto en cursos presenciales como a distancia.
- Facilitar la búsqueda de información por medios electrónicos.
- Creación automatizada de contenidos educativos.
- Facilitar la comercialización de contenidos educativos por medios electrónicos, con lo que las universidades pueden hacerse de recursos comercializando parte del trabajo académico y de investigación que se realiza.
- La posibilidad de armar en forma automatizada cursos considerando las necesidades y requerimientos del estudiante como son la edad, complejidad, manejo de medios, etc. [IEEE, 2002]

A continuación se detallan los aspectos e ideas principales de cuatro propuestas que abordan la medición de la calidad de los OAs desde diferentes puntos de vista. La mayoría de las propuestas necesitan obligatoriamente la participación del usuario para determinar de alguna manera la calificación final del OA.

**Propuesta 1: La Determinación de la Calidad del Contenido de un Objeto de Aprendizaje**

Según [Velázquez et al, 2006a], los autores consideran que es necesario estar consientes que se enfrentan a un producto informático y educacional de manera simultánea, por lo que la calidad de producto en este caso debe considerar los distintos aspectos de un desarrollo de software que emplea el paradigma de objetos y saben que deben existir consideraciones relacionadas a un producto de tipo educativo. En base a lo anterior y con la finalidad de facilitar el estudio de la calidad de un OA distinguen aspectos técnicos, pedagógicos, de contenido y estéticos, como se observar en la Figura 2.1.

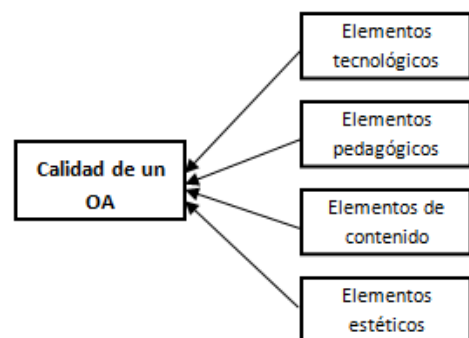


Figura 2.1: Elementos que determinan la calidad de un OA

Elementos Técnicos: Los autores determinan que los aspectos técnicos pueden abordarse desde la perspectiva de ingeniería de software considerando que se encuentran frente a un producto de software que se desarrolla empleando el paradigma de objetos. Entonces atribuyen a los OAs las ventajas de los productos realizados bajo este paradigma como los son la *reutilización* y la *adaptabilidad*.

Elementos Pedagógicos: Con relación a los aspectos pedagógicos, de contenido y estéticos se encontraron con el problema de que la definición de calidad para estos elementos se considera subjetiva en gran medida, por lo que les fue necesario determinar qué aspectos se van a poder cuantificar y cuales deberán permanecer subjetivos. No todas las métricas pueden ser plenamente objetivas. En la Figura 2.2 se listan los elementos pedagógicos que considerarán.

Elementos de Contenido: Consideran aquellos que les dan información sobre la complejidad del tema y el nivel de detalle con que se aborda en el OA. Los mismos se observan en la Figura 2.3.

Elementos Estéticos: Consideran que los aspectos estéticos de un OA se refieren a la presentación de la información (fuentes, colores, tamaño, en sí todos los elementos de formato) y la disposición de la misma (acomodo simétrico o asimétrico, etc.) como se puede observar en la Figura 2.4. Cuando hacen referencia a un OA mencionan que la calidad va a estar determinada al igual que en cualquier otro producto de software por el grado de cumplimiento de los requerimientos del usuario. Tomando como base este principio establecen dos postulados:

- 1.- La calidad de contenido de un OA solo se puede determinar del contexto de un determinado tipo de usuario, ya que estará determinada por el grado de cumplimiento de los requerimientos de este determinado tipo de usuario.
- 2.- En caso de que no se ubique el OA en el contexto de un determinado tipo de usuario, no es posible hablar de calidad de contenido (ya que no tienen requerimientos de referencia), en este caso simplemente se refieren a la medición de los elementos de contenido de un OA. En otras palabras, cuando no

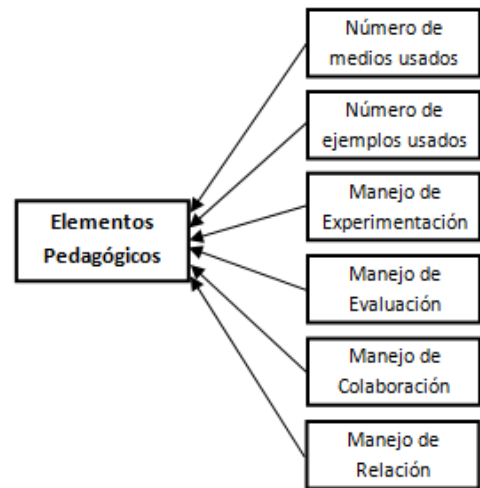


Figura 2.2: Elementos pedagógicos de un OA

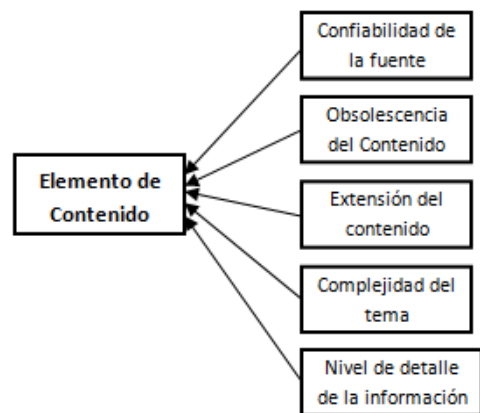


Figura 2.3: Elementos de contenido de un OA

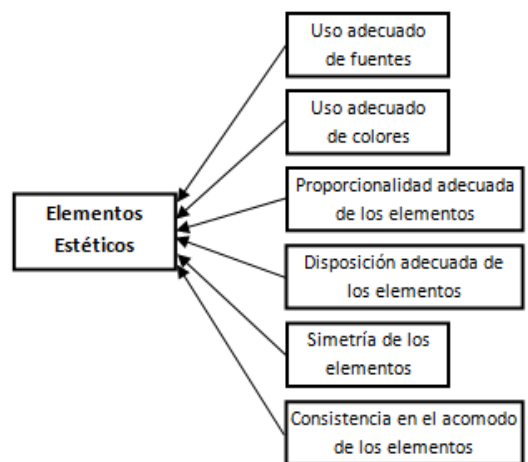


Figura 2.4: Elementos estéticos de un OA

se conoce el contexto del usuario solo pueden hacer una valoración de los elementos del mismo.

El contexto del usuario estará determinado por especificaciones explícitas e implícitas. Dentro de las explícitas se encuentran todas aquellas que el usuario puede expresar en forma consciente y dan una idea clara de sus necesidades de instrucción actuales. Las implícitas son aquellas que el usuario no necesariamente va a poder manifestar en forma consciente, como puede ser el estilo de aprendizaje.

La determinación de la calidad de un OA es llevada a cabo por expertos. En la Tabla 2.1 se puede observar la manera de calificar los OA.

<b>Clasificación: Elementos tecnológicos</b>				
Elemento a evaluar	Estándar para el contexto	Medición en el OA	Peso	Calificación
Reutilización	Recomendado	Si	50%	Bueno
Adaptabilidad	Recomendado	Si	50%	Bueno
<b>Calificación de los elementos tecnológicos 10 (Peso 10%)</b>				
<b>Clasificación: Elementos pedagógicos</b>				
Suma del número de fotografías, diagramas y tablas (Elementos visuales)	3-6	3	30%	Bueno
Personajes interactivos (Elementos visuales)	Indistinto	No	0%	N/A
Número de animaciones sobre conceptos físicos, biológicos, químicos (Elementos visuales)	Indistinto	0	0%	N/A
Suma del número de animaciones de procedimientos técnicos y videos (Elementos visuales)	1	0	10%	Deficiente
Elementos auditivos	Indistinto	No	0%	N/A
Ejemplos	1 por concepto	4	20%	Bueno
Experimentación	Recomendado	No	10%	Deficiente
Preevaluación	Indistinto	No	0%	N/A
Evaluación	Recomendado	Si	20%	Bueno
Colaboración	Recomendado	No	5%	Deficiente
Relación	Recomendado	No	5%	Deficiente
<b>Calificación de los elementos pedagógicos 7 (Peso 40%)</b>				
<b>Clasificación: Elementos de contenido</b>				
Caducidad de información	Indistinto	Indefinido	0%	N/A
Caducidad de conceptos	Años-indefinido	Indefinido	30%	Excelente
Complejidad	Indistinto	Media	0%	N/A
Nivel de detalle de la información	Suficiente-Muy detallada	Detallada	70%	Bueno
<b>Calificación de los elementos de contenido 10 (Peso 40%)</b>				
<b>Clasificación: Elementos estéticos</b>				
Simetría el acomodo de los elementos	Recomendado	Si	80%	Bueno
Número de fuentes que emplea	2	1	20%	Deficiente
<b>Calificación de los elementos estéticos 8 (Peso 10%)</b>				
<b>Calificación general del objeto de aprendizaje 8.6</b>				

Tabla 2.1: Formulario para la determinación de la calidad de un OA

### **Propuesta 2: Estudio de los elementos de la calidad Pedagógica y de Contenido en el Metadato LOM**

Según [Velázquez et al, 2006b], los autores se basaron en la *propuesta 1*, y proponen expandir el conjunto de metadatos que define el estándar LOM agregando metadatos que de alguna manera midan la calidad de los elementos.

Mencionan que la calidad de un OA es determinada por:

- Elementos tecnológicos
- **Elementos pedagógicos**
- **Elementos de contenido**
- Elementos estéticos

Expanden los metadatos correspondientes a Elementos pedagógicos y de contenido.

Elementos pedagógicos:

En la Tabla 2.2 se observa una comparación entre los elementos propuestos por los autores para determinar la calidad pedagógica de un OA y los elementos del metadato de LOM.

<b>Elemento propuesto</b>	<b>Equivalente en IEEE LOM</b>
Formato	4.1 Format
Tipo de Interactividad	5.1 InteractivityType
Tipo de Recurso Educativo	5.2 LearningResourceType
Nivel de Interactividad	5.3 InteractivityLevel
Tiempo típico de aprendizaje	5.9 TypicalLearningTime
Objetivo Pedagógico	9.1 Purpose
Número de medios usados	No hay equivalente en LOM
Número de ejemplos usados	No hay equivalente en LOM
Tipo de ejemplos usados	No hay equivalente en LOM
Manejo de experimentación	No hay equivalente en LOM
Manejo de evaluación	No hay equivalente en LOM
Tipo de evaluación	No hay equivalente en LOM
Manejo de colaboración	No hay equivalente en LOM
Manejo de relación	No hay equivalente en LOM

Tabla 2.2: Metadatos LOM y nuevos metadatos agregados

Con relación a los elementos propuestos puntualizan lo siguiente: El *número de medios usados*, *número de ejemplos usados*, *tipo de ejemplos usados*, *manejo de experimentación*, *manejo de evaluación*, *tipo de evaluación*, *manejo de colaboración* y *manejo de relación*; tendrán un uso en OA agregados, principalmente a nivel de tema; esto permitirá a cada organización manejar la granularidad al grado que deseé.

En los elementos *número de medios usados* y *número de ejemplos usados* emplean valores numéricos enteros para expresar su contenido.

En lo referente a los elementos *manejo de experimentación*, *manejo de evaluación*, *manejo de colaboración* y *manejo de relación*, emplean valores de verdadero o falso, esto permitirá saber si encuentran algún tipo de práctica que permita que el usuario experimente el concepto aprendido, si es que existe alguna evaluación, si se permite realizar trabajo colaborativo y si el OA se encuentra vinculado con algún otro recurso que permita incrementar el aprendizaje.

Con relación al *tipo de evaluación* señalan por ejemplo evaluaciones de reactivos fijos (siempre las mismas preguntas), evaluaciones de reactivos aleatorios (las preguntas van cambiando cada vez que el usuario realiza la evaluación) y evaluaciones

por práctica (este es un tipo complejo de evaluación, en la cual se tomará en cuenta el desempeño del estudiante al realizar una determinada tarea, algunas formas de realizar esto es por medio del uso de simulación e inteligencia artificial).

Elementos de contenido:

Comprende todos aquellos que dan información sobre la complejidad del tema y el nivel de detalle con que abordan en el OA. En la Tabla 2.3 se observa una comparación entre los elementos propuestos y los elementos del metadato de LOM.

<b>Elemento propuesto</b>	<b>Equivalente en IEEE LOM</b>
Estatus del OA	2.2 Status
Entidad (Parte del elemento de confiabilidad de la fuente)	2.3.2 Entity 3.2.2 Entity
Certificación del elemento (Parte del elemento de Confiabilidad de la fuente)	<b>No hay equivalente en LOM</b>
Nivel de detalle de la información (Se recomienda que el usuario pueda seleccionar el nivel deseado)	5.4 SemanticDensity
Complejidad del tema	5.8 Difficulty
Ámbito de la información o Clasificación taxonómica (Parte del elemento de Obsolescencia del contenido. Se recomienda que el usuario pueda seleccionar el ámbito deseado)	9.2 Taxon Path 9.2.1 Source 9.2.2 Taxon 9.2.2.1 Id 9.2.2.2 Entry
Extensión del contenido	<b>No hay equivalente en LOM</b>

Tabla 2.3: Metadatos LOM y nuevos metadatos agregados

El elemento de certificación de la fuente servirá para indicar si el OA ya se encuentra certificado y bajo que estándar se ha realizado esto.

Como se observa, los datos de la comparación entre la propuesta de los elementos que determinan la calidad en OA y el metadato de LOM, determinan que en el metadato de LOM no se tienen contemplados buena parte de los elementos de calidad pedagógicos.

Entonces con la ayuda de expertos o instituciones se podrá calificar los OA utilizando la extensión de los metadatos LOM de manera que la calidad esté contemplada.

**Propuesta 3: Learning Object Review Instrument**

En [Nesbit et al, 2003], los autores proporcionan un marco de evaluación de OA basado en el análisis de nueve dimensiones. A cada dimensión la evalúan mediante una escala de cinco niveles. En [Vargo et al, 2003] demostraron la fiabilidad de este instrumento. Las nueve dimensiones son: *Calidad de Contenido, Alineamiento de los objetivos de aprendizaje, Retroalimentación y adaptación, Motivación, Diseño de Presentación, Usabilidad en la interacción, Accesibilidad, Reusabilidad, Cumplimiento de Estándares.*

La evaluación es llevada a cabo por un grupo de evaluadores quienes mediante actividades de evaluación individual y luego discusiones grupales califican al OA.



#### **Propuesta 4: Evaluación pedagógica de Reeves**

En [Reeves, 1997], los autores proponen un marco para la evaluación de educación basada en computadores. Consta de 14 dimensiones pedagógicas basadas en teorías y conceptos de aprendizaje. Han usado estas dimensiones para evaluar cursos en ambientes de e-learning y, si se considera a un curso como un OA con alto nivel de agregación, los autores señalan que este modelo puede usarse para evaluar la calidad desde el punto de vista pedagógico. Las 14 dimensiones se refieren a aspectos del diseño e implementación de recursos que afectan el aprendizaje. Estas dimensiones son: *epistemológica, filosofía pedagógica, sustento psicológico, orientación a objetivos, validez experimental, rol del instructor, flexibilidad de programa, valor del error, motivación, adaptación a diferencia a individuales, control de aprendizaje, actividades de usuario, aprendizaje cooperativo y sensibilidad cultural.*

Al igual que en la *propuesta 3* la evaluación la desarrolla un grupo de evaluadores quienes mediante actividades de evaluación individual y luego discusiones grupales califican al OA.

#### **Propuesta de análisis de calidad y confiabilidad de los OAs**

En este trabajo proponemos analizar la calidad y confiabilidad de los OAs en el momento que son recuperados luego de realizar una consulta a un repositorio de OAs, con el objetivo de que los OAs de mayor calidad y confiabilidad escalen las primeras posiciones y sean personalizados al perfil de usuario que realiza la consulta. Un OA se considerará de calidad si sus metadatos son de calidad y confiable si las calificaciones de uso otorgadas por los usuarios son significativas en la escala propuesta para su medición. Los OAs serán analizados desde los siguientes puntos de vista:

- Calidad
  - Metadatos: *Análisis detallado de todos los metadatos definidos bajo el estándar LOM. Básicamente consiste en verificar/validar siguiendo el estándar LOM la existencia, tamaño, formato, pertenencia de cada uno de los metadatos que describen al OA.*
  - Idiomas: *Validación y análisis de concordancia de los idiomas asignados en los metadatos General.language y MetaMetadata.language.*
  - Keywords: *Validación y análisis de concordancia de las palabras claves asignadas en el metadato General.keyword.*
  
- Confiabilidad
  - Calificaciones de usuarios finales: *Análisis de las puntuaciones otorgadas por los usuarios que hacen uso del OA considerando el perfil asignado y la autoevaluación dado sus conocimientos y experiencias sobre los temas que trata el OA.*

La diferencia que se intenta remarcar en este trabajo es la de priorizar el análisis automático/semiautomático sobre el análisis manual de expertos o usuarios finales, con el fin de que el usuario no tenga participación al momento de calificar la calidad desde el punto de vista estructural y de contenido sino que la tenga solo desde el punto de vista pedagógico.

### 3. Propuesta

En este trabajo se propone analizar la calidad y confiabilidad de los OAs al momento que son recuperados luego de realizar una consulta a un repositorio de OAs, con el fin de que los OAs de mayor calidad y confiabilidad escalen las primeras posiciones y sean personalizados al perfil de usuario que realiza la consulta. Un OA se considerará de calidad si sus metadatos son de calidad y confiable si las calificaciones de uso otorgadas son significativas en la escala propuesta para su medición. Se propone analizar la existencia, formato, tamaño, pertenencia de todos los metadatos que describen al OA siguiendo el estándar de metadatos LOM, también se propone analizar la concordancia de palabras claves e idiomas asignados en los metadatos *General.language*, *MetaMetadata.language* y *General.keyword*.

Las ideas propuestas este trabajo son implementadas en una aplicación web, la cual utiliza la estrategia de flujo que se detalla en la Figura 3.1:

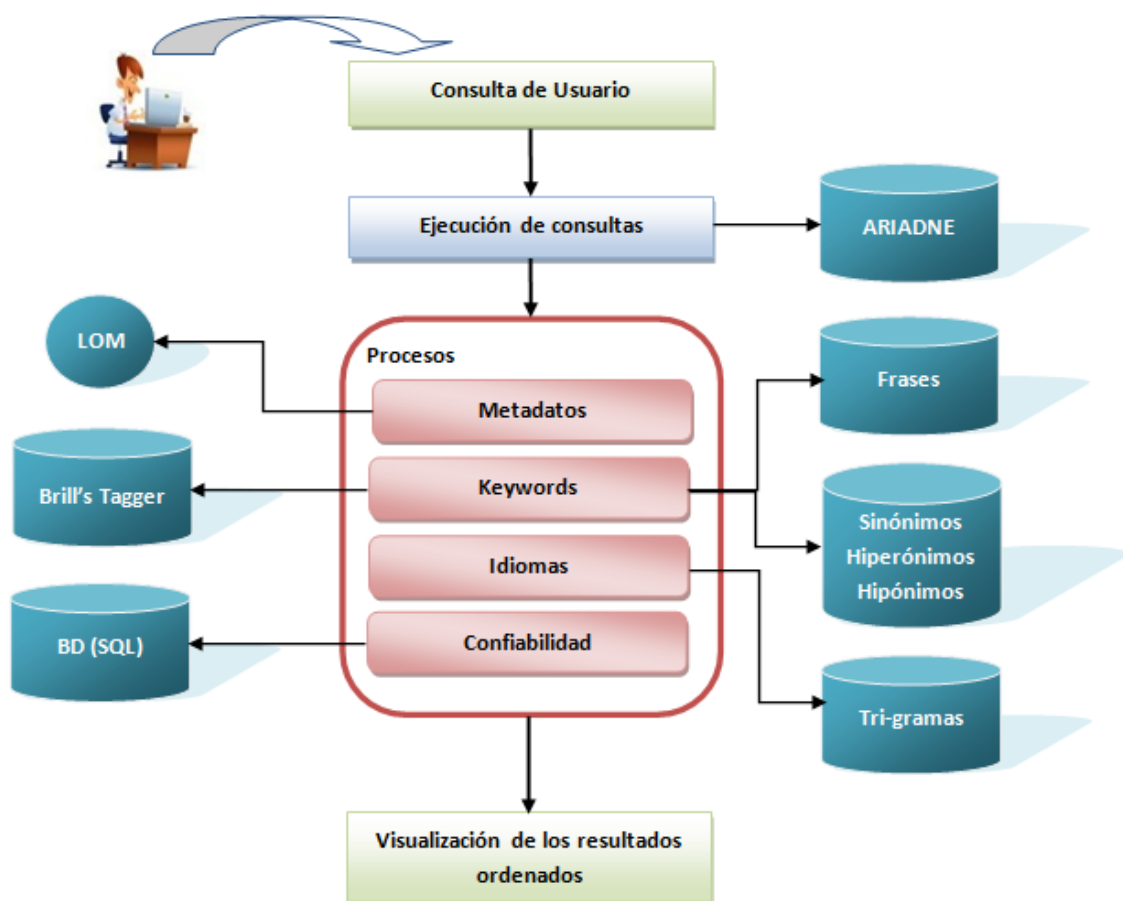


Figura 3.1: Estrategia para determinar la calidad total de los OAs

**Consulta de Usuario:** Este módulo permite al usuario realizar consultas mediante palabras claves, iniciar/cerrar sesión al sistema, crear usuarios y visualizar estadísticas de búsqueda. En la Figura 3.2 se puede observar la interfaz de usuario.

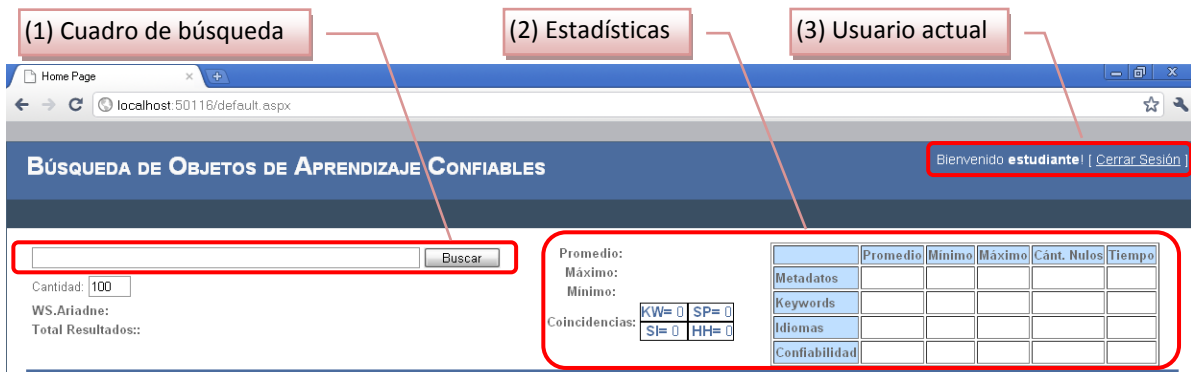


Figura 3.2: Interfaz de usuario para realizar búsqueda mediante palabras claves

- (1) Cuadro de Búsqueda: El usuario ingresa las palabras claves a buscar.
- (2) Estadística: Se despliegan estadísticas referentes a la búsqueda.
- (3) Usuario Actual: Se visualiza el usuario actual. En la Figura 3.3 se observa el formulario de inicio de sesión.

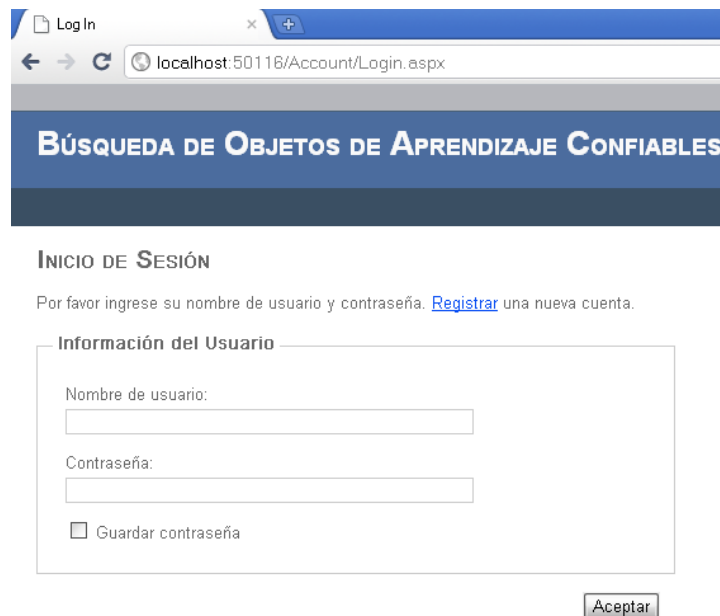


Figura 3.3: Formulario de Inicio de Sesión

En la propuesta solo es utilizado el perfil o profesión del usuario, por ende solo esta información es requerida. Los perfiles contemplados en este trabajo son **Teacher** y **Student**. En la Figura 3.4 se puede observar el formulario de *Nuevo Usuario*.

Register

localhost:50116/Account/Register.aspx?ReturnUrl=

## BÚSQUEDA DE OBJETOS DE APRENDIZAJE CONFIABLES

### NUEVO USUARIO

Utilice el formulario para crear la nueva cuenta.

Nombre de usuario:

Contraseña:

Confirmar contraseña:

Profesión: 

- Teacher
- Teacher
- Student

Figura 3.4: Formulario de Nuevo Usuario

**Ejecución de consultas:** Este módulo realiza la conexión mediante la interface de consulta SQI al repositorio de OAs **Ariadne**, ejecuta la consulta en el lenguaje de consulta *Lucene* y almacena los resultados en memoria para su posterior análisis.

**Procesos:** Aplica los procesos *Metadatos*, *Keywords*, *Idiomas* y *Confiabilidad* a los resultados (OAs) obtenidos luego de la consulta del usuario con el fin de obtener la calidad o puntaje total de los mismos.

- **Metadatos:** Los metadatos pertenecientes a las distintas categorías son calificados siguiendo el estándar LOM, obteniendo así un valor que represente la calidad del metadato, por ende también poder obtener una calificación total de las categorías. Además las categorías son ponderadas con el fin de personalizar los OAs de acuerdo al perfil del usuario y obtener un valor numérico que represente la calidad de metadatos del OA.
- **Idiomas:** En los metadatos de cada OA, existen dos metadatos que determinan el idioma del contenido del OA, y el idioma de la instancia de metadatos asociada. Se validan si los idiomas asignados son correctos utilizando un algoritmo de detección de idiomas. Luego se obtiene un valor numérico que representa la calidad del metadato Idiomas del OA.
- **Keywords:** En los metadatos de cada OA, existe un metadato que determina las keywords o palabras claves del OA. Estas hacen referencia a las palabras claves o términos que más abundan en el contenido del OA y en los valores asignados en los metadatos de tipo *LangString* como los son *General.title* y *General.description*. En este trabajo, sólo en el caso que el contenido del OA sea una página web o archivo HTML, se validan si las keywords son correctas, es decir si realmente describen el contenido del OA. También se validan si las keywords describen los valores asignados en los metadatos *General.title* y *General.description*. Además para poder

obtener un valor más preciso en la calificación, se expanden las keywords utilizando sinónimos, hiperónimos, hipónimos y sus respectivas palabras en plural y singular.

- **Confiabilidad:** Se analiza las opiniones y puntuaciones de los usuarios que hagan uso del OA. Los aspectos que se tienen en cuenta son: *Facilidad de Uso*, *Calidad de Contenido*, *Potencialidad de efectividad*, *Descripción del objeto* y *Autoevaluación*. De la misma manera que en *Metadatos*, el valor de votación es ponderado teniendo en cuenta el perfil del usuario.

Los procesos antes descriptos menos *Confiabilidad* descartan aspectos pedagógicos. Sin embargo en este proceso como se detalla en la Sección 3.4 los aspectos pedagógicos sí son tenidos en cuenta al momento de puntuar *Potencialidad de efectividad*, ya que se debe calificar si el OA realmente cumple con las expectativas de aprendizaje del usuario, si fue efectivo, si cumple con los objetivos para los cuales fue creado.

**Visualización de los resultados ordenados:** Por medio de este módulo, el usuario visualiza los OAs ordenados por la *calidad total* de los OA dónde la *calidad total* es el promedio de los puntajes obtenidos en los procesos aplicados. En la Figura 3.5 se puede observar un ejemplo de un objeto de aprendizaje como resultado de búsqueda de la palabra clave “*java programming*”, los datos sombreados son solo a modo estadístico y de ejemplo.

The screenshot shows search results for the query "java programming". The results are highlighted with red boxes and numbered annotations:

- (1) Puntaje Total: Points to the overall score of 0.59200.
- (2) Puntajes de los procesos: Points to the breakdown of scores: Calidad=0.31134, Idiomas=1.00000, Metadatos=0.25817, Confiabilidad=0.80000.
- (3) Coincidencias: Points to the list of keywords and synonyms.
- (4) Palabras claves META y HTML: Points to the meta and HTML keywords.
- (5) Extensión de palabras claves: Points to the expanded keywords section.

Key data from the screenshot:

- Title:** Introduction to OO Programming in Java - Arrays and For Loops
- Puntaje:** 0,59200
- Calidad:** Keywords=0.31134, Idiomas=1.00000, Metadatos=0.25817, Confiabilidad=0.80000
- Coincidencias:** [ Keywords: 7 ===== Singular/Plural: 1 ===== Sinonimos: 1 ===== Hiper-Hiponimos: 1 ]
- KeywordsMETA:** introduction, oo, programming, java, awt
- KeywordsHTML:** introduction, oo, programming, java
- KeywordsHTMLBody:** visual\_aid\_programming\_object\_oriented\_java\_arrays\_iteration\_loops\_software\_licence\_item\_object-oriented\_programming\_introduction\_ukoer\_loop\_array\_definite\_g622\_oo\_programming
- KeywordsMETAKeywords:** ukoer, software\_objects\_visual\_guide, for\_loop, for\_loops, array, arrays, iteration, definite\_iteration, g622, oo\_programming, object-oriented, object-oriented\_programming, java, object\_oriented\_programming, for\_loop\_visual\_aid, for\_loops\_visual\_aid, array\_visual\_aid, arrays\_visual\_aid, iteration\_visual\_aid, software\_object\_visual\_aid, definite\_iteration\_visual\_aid, g622\_visual\_aid, oo\_programming\_visual\_aid, object-oriented\_visual\_aid, object-oriented\_programming\_visual\_aid, java\_visual\_aid, object\_oriented\_programming\_visual\_aid, oo\_visual\_aid, object\_oriented\_visual\_aid
- Singular-PluralMETAKeywords:** ukoers, softwares\_object\_visual\_guides, for\_loops, for\_loop, arraies, array, iterations, definite\_iterations, g622s, javas, objects\_oriented\_programming, for\_loops\_visual\_aid, for\_loop\_visual\_aid, arraies\_visual\_aid, array\_visual\_aid, iterations\_visual\_aid, softwares\_objects\_visual\_aid, definite\_iterations\_visual\_aid, g622s\_visual\_aid, oo\_programming\_visual\_aid, object-oriented\_visual\_aid, object-oriented\_programming\_visual\_aid, javas\_visual\_aid, objects\_oriented\_programming\_visual\_aid, oo\_visual\_aid, objects\_oriented\_visual\_aid
- SinonimosMETAKeywords:** loop - 1 de 6
- Hiper-HipoMETAKeywords:** programming - 1 de 71
- Metadatos:** Format: LearningResourceType: IdiomaOA:en IdiomaMM:en
- Description:** This reading material forms part of the "Arrays and For Loops" topic in the Introduction to OO Programming in Java module.

Figura 3.5: Ejemplo de un resultado de búsqueda

- (1) Puntaje Total: Se visualiza el puntaje total del objeto obtenido como promedio de los puntajes de los procesos. (**Puntaje = 0,59200**).
- (2) Puntajes de los procesos: Se visualiza el puntaje que obtuvo el objeto en cada uno de los procesos.  
(**Keywords = 0,31134 – Idiomas = 1 – Metadatos = 0,25817– Confiabilidad = 0,8**)

- (3) Coincidencias: Se muestran la cantidad de coincidencias de las palabras claves (keywords), de palabras en singular/plural, de sinónimos, y de hiperónimos e hipónimos.  
 ( $Keywords = 7 - Singular/Plural = 1 - Sinónimos = 1 - Hiper-Hipónimos = 1$ )
- (4) Palabras claves META y HTML: Se visualizan las palabras claves extraídas de los metadatos *General.title* y *General.description* y también de los tags *title* y *body* correspondientes al HTML de la página web a la cual el objeto hace referencia.
- (5) Extensión de palabras claves: Se visualizan las extensiones de las palabras claves y sus coincidencias, en negrita en la Figura 3.5.

### Estadísticas

La Figura 3.6 muestra un ejemplo de las estadísticas sobre los resultados de una consulta del usuario.

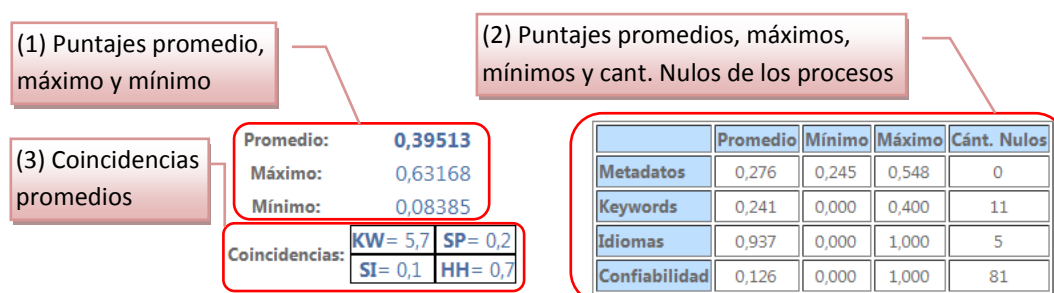


Figura 3.6: Estadísticas de los resultados de búsqueda

- (1) Puntajes promedio, máximo y mínimo: Puntajes máximo, promedio y mínimo obtenido de los puntajes totales de todos los resultados de búsqueda. Notar que el puntaje total de un OA es el promedio de los puntajes obtenidos en los procesos.
- (2) Puntajes promedios, máximos, mínimos y cant. Nulos de los procesos: Puntajes máximo, promedio, mínimo y cantidad de nulos obtenidos al aplicar los procesos a todos los resultados de búsqueda. Un OA tiene puntaje nulo en un determinado proceso cuando obtuvo calificación 0.
- (3) Coincidencias promedios: Promedio de coincidencias de palabras claves (KW), sinónimos (SI), palabras en singular/plural (SP) e hiperónimos e hipónimos (HH) en los resultados de búsqueda.

En los siguientes apartados se describen en profundidad los procesos antes mencionados.

### 3.1 Proceso Metadatos

La personalización de los resultados propuesta en este trabajo es incrementado por la calidad de los metadatos de los OAs según el perfil del usuario que realiza la búsqueda.

Por cada OA recuperado, se busca calcular un valor numérico que determine la calidad de la instancia de metadatos asociada a él, teniendo en cuenta los Metadatos incompletos y la calidad del mismo siguiendo las especificaciones definidas en el estándar LOM [IEEE, 2002].

#### 3.1.1 Sistema de Puntuación

Proponemos un sistema de puntuación que se basa en una escala *Likert* [Bertran, 2004] de cinco puntos, donde si el metadato a analizar no existe en la instancia de metadatos del OA, la puntuación para el metadato será 0. Los puntajes a continuación descriptos solo serán calculados si el tamaño del metadato a analizar no es superior al valor máximo permitido. En el caso que sea mayor, la puntuación del metadato será 0 en el caso contrario continuará un análisis dependiendo de tipo de datos del mismo.

En la Figura 3.7 se puede visualizar el proceso de análisis propuesto de cada uno de los metadatos:

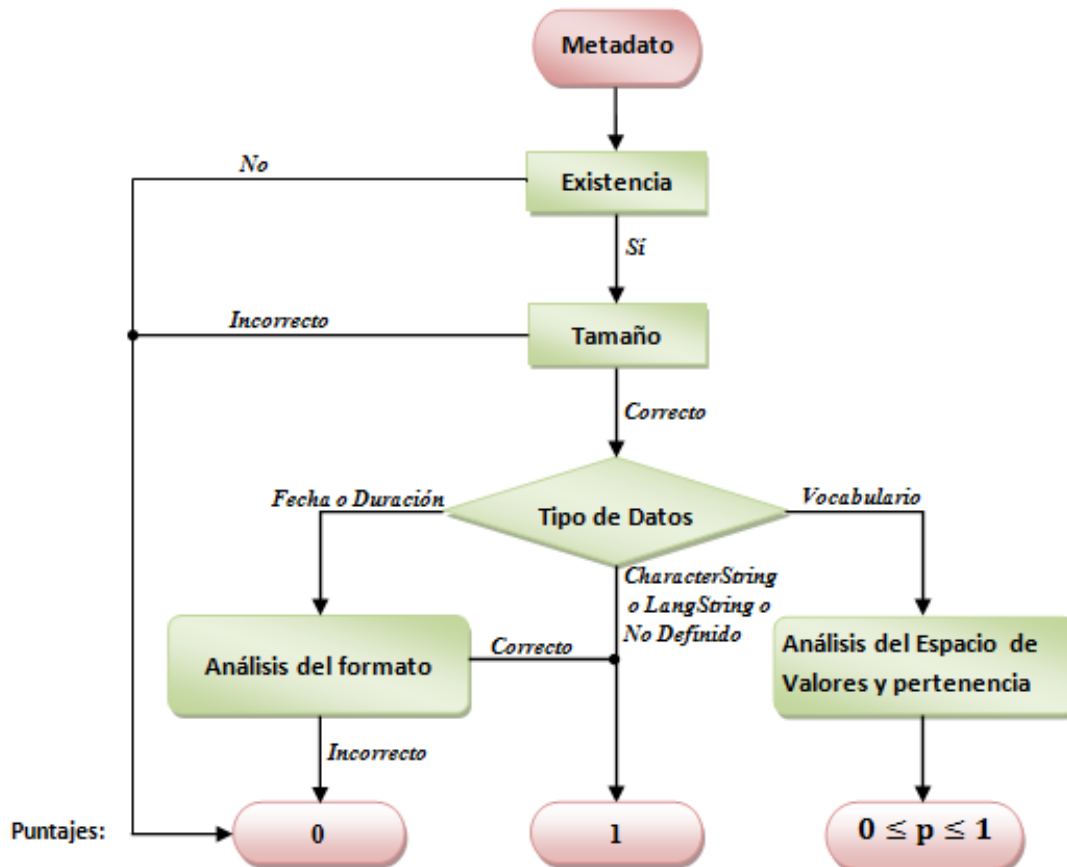


Figura 3.7: Proceso de análisis de los metadatos

## 1. Existencia

Determina si el metadato existe o no en la categoría analizada. En el caso que el metadato no exista el puntaje del mismo es *malo* = 0.

## 2. Tamaño

Verifica si el tamaño del metadato es correcto o no con respecto al definido en el estándar LOM. En el caso que el tamaño sea incorrecto el puntaje es *malo* = 0.

## 3. Tipo de Datos

Detecta el tipo de dato del metadato y deriva en los siguientes módulos alternativos. En el caso de que el tipo de dato detectado sea LangString o CharacterString o No Definido el puntaje es *exitoso* = 1.

## 4. Análisis del espacio de valores y pertenencia

Los metadatos con espacio de valores enumerados (vocabulario) se componen de *source* y *value*.

Por ejemplo:

```
<lifeCycle>
  ....
  <status>
    <source>LOMv1.0</source>
    <value>draft</value>
  </status>
  ....
</lifeCycle>
```

El *source* determina el namespace del *value*. En el ejemplo *source* = LOMv1.0 (namespace definido en el estándar LOM), en este caso siguiendo el estándar el espacio de valores (EV) es: *draft, final, revised, unavailable*. LOMv1.0 no es el único namespace disponible, los propios usuarios o implementadores pueden crear sus propios namespaces siempre teniendo en cuenta que los valores del vocabulario definido no entren en conflicto con el del estándar LOMv1.0 de manera de no disminuir la interoperabilidad de los OAs. En este trabajo validaremos solo el namespace LOMv1.0, pero igualmente consideraremos la posibilidad de otros namespace puntuándolo con un valor menor distinto de 0.

Teniendo en cuenta lo anterior descrito proponemos la siguiente puntuación (**p**):

- **Exitoso = 1:** Si *source* = LOMv1.0 y *value* está incluido en el EV.
- **Sobresaliente = 0,75:** Si *source* = null y *value* está incluido en el EV.



- **Muy Bueno = 0,5**: Si *source* = LOMv1.0 y *value* no está incluido en EV.
- **Bueno = 0,25**: Si *source* != LOMv1.0 o *source* = null y *value* no está incluido en EV.
- **Malo = 0**: En otro caso.

## 5. Análisis de Formato

Verifica si el valor almacenado en el metadato de tipo Fecha o Duración tiene el formato correcto siguiendo el estándar en que se basa LOM (ISO8601:2000) [W3C, 2004] dónde las fechas válidas deben tener el formato que se detalla a continuación.

El espacio léxico de las Fechas consiste en secuencias de longitud finita de caracteres de la forma:

***yyyy '-' mm '-' dd 'T' hh: 'mm': 'ss' ('.' 's +)? (zzzzzz)?***

Dónde:

- *yyyy* es un número de cuatro o más dígitos, opcionalmente negativa que representa el año. Si son más de cuatro dígitos los ceros a la izquierda están prohibidos. "0000" también está prohibido.
- '-' son los separadores entre las partes de fecha;
- El primer *mm* es un número de dos dígitos que representa el mes.
- *dd* es un número de dos dígitos que representa el día.
- 'T' es un separador que indica que se va a especificar la hora.
- *hh* es una cifra de dos dígitos que representa la hora, "24" se permite si los minutos y segundos son cero, y el valor Fecha representado es el primer instante del día siguiente. El valor no puede ser superior a 23.
- ':' Es un separador entre las partes de la porción de la hora del día.
- El segundo *mm* es un número de dos dígitos que representa el minuto.
- *ss* es un número entero de dos dígitos que representa el segundo conjunto.
- '.' s + (si existe) representa las fracciones de segundo.
- *zzzzzz* (si existe) representa la zona horaria.

Los tiempos (duración) válidos según el estándar se observa a continuación en la representación léxica de duración. *n* se sustituye por el valor correspondiente.

***PnYn MnDTnH nMnS***

Dónde:

- *Y* representa el número de años.
- *M* representa el número de meses.
- *D* representa el número de días.
- "T" es el separador entre la fecha y la hora.

- $H$  representa el número de horas.
- $M$  representa el número de minutos.
- $S$  representa el número de segundos. El número de segundos que puede incluir dígitos decimales de precisión arbitraria.

En caso de que el formato del metadato sea correcto el puntaje es exitoso = 1. En caso contrario el puntaje es *malo* = 0.

Haciendo uso del flujo de trabajo antes descripto y promediando los valores obtenidos por cada metadato de cada categoría, se obtiene un valor numérico menor o igual a 1 que determina la calidad de la categoría. Se propone la siguiente definición:

*El estándar de metadatos LOM define 9 categorías de metadatos: General, Technical, Life Cycle, Educational, Meta-Metadata, Rights, Relation, Annotation y Classification.*

*Sea  $c$  una categoría cualquiera ( $c \in 1...9$ ), y sea  $M_c$  el conjunto de sus metadatos.*

Por ejemplo:

*$M_2 =$  Conjunto de metadatos de la categoría Technical.*

*La calidad de la categoría  $c$  viene dada por:*

$$0 \leq \text{Calidad}C_c = \frac{\sum_{m \in M_c} \text{Calidad}M_m}{\#M_c} \leq 1$$

*Dónde  $\text{Calidad}M_m$  es la calidad del metadato  $m$  obtenido mediante el sistema de puntuación propuesto y  $\#M_c$  la cardinalidad del conjunto.*

### 3.1.2 Ponderación de las Categorías

Para obtener un valor numérico personalizado de la calidad de metadatos teniendo en cuenta el perfil del usuario nos basamos en lo detallado en [Woo et al, 2004]. En el mismo se realizaron dos encuestas, en una se trata de analizar cuáles son los metadatos más relevantes para los estudiantes (criterios de búsquedas) al momento de realizar una búsqueda y en la otra cuál es la información más solicitada por profesores al momento de analizar un OA. Se obtuvieron los siguientes resultados que se explican a continuación.

Los estudiantes sugirieron que se podría buscar los OAs de la misma manera que se busca un elemento en un catálogo de la biblioteca. Revelaron que en el criterio deseado para la búsqueda incluyeron palabras claves, resultados/competencias, requisitos de Plan de Estudios, tipo de medios de comunicación, sector, estado o país, creadores y editores. La Tabla 3.1 muestra los criterios de búsqueda en orden descendente de preferencia, y los campos del estándar IEEE LOM (Learning Object Metadata) asociados a ellos. Los criterios son puntuados en una escala *Likert* de cinco puntos.

Search Criteria	Mean	IEEE LOM Field
Subject	4.73	(none)
Keyword	4.13	Field 1.5 General.Keywords Field 9.4 Classification.Keyword
Outcomes/Competencies	3.98	
Syllabus requirements	3.85	
Media Type	3.70	Field 4.1 Technical.Format
Sector (School, VET, Higher Education)	3.30	Field 5.6 Education.Context
State/Country	2.60	
Institution of creator(s)	2.43	
Creator(s)	2.40	Field 2.3.1 LifeCycle.Contribute Fields 2.3.2 LifeCycle.Role/Entity

Tabla 3.1: Criterio de búsqueda y metadatos asociados

Los autores mencionan que al comparar los criterios deseados con los campos disponibles en LOM, es dudoso que la actual estructura de metadatos satisfaga las necesidades educativas de los estudiantes al momento de realizar las búsquedas. Los criterios deseados son más que una lista de deseos. Su disponibilidad se traduce en eficiencia para los usuarios potenciales de OA en el proceso de búsqueda. Si los metadatos proporcionados no son relevantes para las necesidades del usuario, la incorporación de OA se verá seriamente restringida.

En este trabajo decidimos generalizar los metadatos solicitados por los estudiantes obteniendo las siguientes categorías:

- 1) Metadato *keyword*. Categoría **General** y **Clasificación**.
- 2) Metadato *format*. Categoría **Técnica**.
- 3) Metadato *context*. Categoría **Educación**.
- 4) Metadatos *contribute* y *role/entity*. Categoría **Ciclo de Vida**.

Dada la generalización propuesta podemos indicar que las categorías *General*, *Clasificación* y *Técnica* obtuvieron un alto puntaje, lo que significa que es necesario contar con este tipo de información.

Los metadatos se utilizan no sólo con fines de búsqueda, sino también para ayudar a los profesores a tomar decisiones entre los OA. En el trabajo detallado en [Woo et al, 2004], los participantes sugirieron que al tomar decisiones sobre la idoneidad de un OA, querrían la información relativa a los requisitos técnicos y las instrucciones para el uso del objeto, su audiencia estudiantil, los objetivos y los resultados asociados y un historial de uso anterior.

En la Tabla 3.2 se observa los Metadatos más solicitados por los Profesores.

Useful Information on leaning objects	Mean	IEEE LOM Field
1. Technical information such as hardware and software requirements, file size, download times	4.28	Field 4.2 Technical.Size Field 4.4 Technical.Requirements Field 4.6 Technical.Other Platform Requirements
2. The student group for which it was developed (e.g. age, expertise, special needs)	4.20	(none) (Field 5.10 Education.Description?)
3. The teaching and learning context associated with its use e.g., its role, pre-requisite knowledge required, supporting materials	4.13	(none) (Field 5.10 Education.Description?)
4. The specific teaching aims and learning outcomes for which it was developed	4.00	Field 9.3 Classification.Description
5. Technical instructions for students on how to use the learning object (if applicable)	3.83	Field 4.5 Technical.Installation Remarks
6. A review provided by my teaching peers	3.40	Field 2.3.3 LifeCycle.Contribute.Date
7. The life history of the object – details of creation, publication and last modification	3.40	Field 2.3.3 LifeCycle.Contribute.Date
8. A review provided by a curriculum / syllabus authority	3.23	Field 8 Annotations
9. A record of previous uses of the learning object	3.13	Field 8 Annotations

Tabla 3.2: Metadatos más solicitados por los profesores (Teacher)

Decidimos también generalizar los metadatos solicitados por los profesores obteniendo las siguientes categorías:

- 1) Metadatos *size, requeriments y other plataform requeriments*. Categoría **Técnica**.
- 2) Metadatos *description*. Categoría **Clasificación**.
- 3) Metadatos *contribute y date*. Categoría **Ciclo de Vida**.
- 4) Metadatos *entity, date y description*. Categoría **Anotación**.

Los metadatos de las categorías Técnica y Clasificación obtuvieron un alto puntaje, lo que indica que es necesario contar con este tipo de información de manera que los profesores la puedan utilizar para su selección.

Entonces dados estos resultados se decidió ponderar en esta tesina la importancia en la calidad de los metadatos de cada categoría siguiendo el orden antes descripto. Dónde el orden de las categorías es el siguiente: (ordenado del más ponderado al menos ponderado)

*Ponderación en criterios de búsquedas de **estudiantes**:*

- 1) General
- 2) Clasificación
- 3) Técnica
- 4) Educación
- 5) Ciclo de vida

*Ponderación en criterios de búsquedas y selección de los **profesores**:*

- 1) Técnica
- 2) Clasificación
- 3) Ciclo de Vida
- 4) Anotación

Cabe destacar que la calidad de la instancia de metadatos asociada a un OA es determinada por la calidad específicamente de cada una de sus categorías de metadatos y la calidad de cada categoría es determinada por el promedio de la calidad de cada uno de los metadatos que la componen. Básicamente la calificación total de metadatos del OA es el promedio ponderado del puntaje en cada una de las categorías, donde la ponderación viene dada por las preferencias del perfil actual que hace uso del sistema, es decir hay ciertas categorías que pueden ser más importantes para un perfil que para otro.

Por lo tanto teniendo en cuenta las preferencias de los usuarios con perfil **Student** al momento de realizar una búsqueda y las preferencias de los usuarios con perfil **Teacher** al momento de filtrar y seleccionar OAs obtenidas en [Woo et al, 2004]

podemos determinar de manera personalizada por cada perfil de usuario un valor de calidad de metadatos del OA diferente.

Entonces siguiendo [Woo et al, 2004] proponemos lo siguiente:

Definimos un “peso” para cada categoría preferencial en los usuarios con perfil **Student** (Stud):

- |    |              |   |                      |  |
|----|--------------|---|----------------------|--|
| 1. | $p_1^{Stud}$ | → | <b>General</b>       | dónde $1 < p_5^{Stud} < p_4^{Stud} < p_3^{Stud} < p_2^{Stud} < p_1^{Stud}$<br>y<br>$p_6^{Stud} = p_7^{Stud} = p_8^{Stud} = p_9^{Stud} = 1$ |
| 2. | $p_2^{Stud}$ | → | <b>Clasificación</b> |  |
| 3. | $p_3^{Stud}$ | → | <b>Técnica</b>       |  |
| 4. | $p_4^{Stud}$ | → | <b>Educación</b>     |  |
| 5. | $p_5^{Stud}$ | → | <b>Ciclo de Vida</b> |  |
| 6. | $p_6^{Stud}$ | → | Anotación            |  |
| 7. | $p_7^{Stud}$ | → | Meta-Metadatos       |  |
| 8. | $p_8^{Stud}$ | → | Derecho              |  |
| 9. | $p_9^{Stud}$ | → | Relación             |  |

Entonces la calidad de metadatos para el OA  $i$  en el perfil **Student** es el siguiente promedio ponderado:

$$CalidadMetadatos_i = \frac{\sum_{c=1}^9 p_c^{Stud} * CalidadC_c}{\sum_{c=1}^9 p_c^{Stud}}$$

Dónde:

$CalidadMetadatos_i$  : Calidad de metadatos del OA  $i$ .

$p_c^{Stud}$  : Peso en la categoría  $c$  en el perfil **Student**.

$CalidadC_c$  : Calidad de la categoría  $c$ .

De la misma manera es calculado para el perfil **Teacher**.

De esta forma la calidad de los metadatos en los OAs recuperados es personalizada al perfil del usuario obteniendo como resultados que los OAs más relevantes con una cuidada calidad y los que cumplen con sus preferencias sean visualizados en las primeras posiciones.

## 3.2 Proceso Keywords

En este proceso proponemos analizar la coincidencia existente entre las palabras claves y sus relaciones detectadas en el metadato *General.keyword* del OA con las palabras claves extraídas del texto (HTML) de la página web al cual hace referencia el OA en caso de existir y las extraídas de los metadatos *General.description* y *General.title*. En caso que el OA no haga referencia a una página web (HTML) la extracción se realizó solo sobre los metadatos *General.description* y *General.title* del mismo. Luego mediante un sistema de puntuación se determina la *calidad de keywords* del OA en cuestion.

El proceso *keywords* propuesto se divide en 3 etapas, dónde la tercer etapa unifica la salidas de las demás etapas:

1. Extracción y extensión de las palabras claves obtenidas del metadato *General.keyword*
2. Extracción de las palabras claves del HTML y de los metadatos *General.description* y *General.title*
3. Cálculo de coincidencias

En la Figura 3.8 se observa en detalle las etapas junto a los recursos y/o librerías externas que utilizan, dónde el resultado CPCT es la “coincidencia de palabras claves total”.



Figura 3.8: Etapas del proceso keywords

La etapa “*Extracción y extensión de las palabras claves obtenidas del metadato General.keyword*” utiliza el recurso externo Wordnet que se detalla en [Miller, 2010], la estructura de metadatos LOM y las librerías de Brill’s Tagger descrita en [Brill, 1993]. La etapa “*Extracción de palabras claves del HTML y metadatos de tipo LangString*” utiliza el recurso externo StopWords que se detalla a continuación, la estructura de metadatos LOM y la librería AgilityHTMLPack que se detalla en [Html Agility Pack, 2010].

## **Etapa 1: Extracción y extensión de las palabras claves obtenidas del metadato *General.keyword***

En el estándar LOM el metadato *General.keyword* es de tipo *LangString*, no ordenado, múltiple y permite como máximo 10 elementos. En él se asignan las keywords o palabras claves que describen al OA. A continuación se describen los pasos en forma secuencial que se realizan.

### **1. Extracción de palabras claves**

Por cada OA recuperado se extraen las keywords o palabras claves del metadato *General.keyword* con el fin de generar un conjunto que deberían hacer referencia al contenido del mismo.

En esta etapa también se detectan las frases en las palabras claves utilizando una estructura de frases generada desde un diccionario de frases extraído de WordNet [Miller, 2010]. De existir alguna, la misma será tratada como frase y no como una simple palabra clave en las siguientes etapas.

La estructura de frases está almacenada en un archivo XML que generamos a partir de las fuentes de frases del idioma Inglés publicadas en WordNet [Miller, 2010].

La estructura propuesta del archivo XML es la siguiente:

```
<frases>
  ....
  <frase>flight_line</frase>
  <frase>flight_manuever</frase>
  <frase>flight_of_stairs</frase>
  <frase>flight_of_steps</frase>
  <frase>flight_path</frase>
  <frase>flight_simulator</frase>
  <frase>flight_strip</frase>
  <frase>flight_surgeon</frase>
  ....
</frases>
```

### **2. Extensión de palabras claves**

Dadas las palabras claves extraídas del metadato *General.keyword*, por cada una se genera una familia de sinónimos, hiperónimos e hipónimos asociados.

Para generar satisfactoriamente las familias de sinónimos, hiperónimos e hipónimos se utiliza una estructura de palabras almacenadas en un archivo XML estructuralmente ágil para su navegación y consulta que se generó consultando las fuentes *index.noun* y *data.noun* publicadas en WordNet [Miller, 2010].

La estructura propuesta del archivo XML es la siguiente:

```
<words>
  ....
  <word name="olfaction">
    <sinonimos>
      <sinonimo>smell</sinonimo>
      <sinonimo>sense_of_smell</sinonimo>
      <sinonimo>olfactory_modality</sinonimo>
    </sinonimos>
    <hiperhiponimos>
      <hiperhiponimo>modality</hiperhiponimo>
      <hiperhiponimo>sense_modality</hiperhiponimo>
      <hiperhiponimo>sensory_system</hiperhiponimo>
      <hiperhiponimo>exteroception</hiperhiponimo>
      <hiperhiponimo>nose</hiperhiponimo>
    </hiperhiponimos>
  </word>
  ....
</words>
```

El nodo que se visualiza arriba permite dada una palabra (word) obtener todos los sinónimos, hiperónimos e hipónimos asociados.

También se analiza el estado morfológico de cada palabra clave extraída de manera de determinar si está en singular o plural y aplicándole un algoritmo de Tagueo Morfológico (Brill's Tagger) que se detalla en [Brill, 1993] se obtiene su conjunto de palabras asociadas. Es decir si la palabra clave está en singular, entonces la palabra incluida en el conjunto estará en plural y viceversa.

Entonces por cada OA analizado se obtiene:

- Un conjunto palabras claves
- Un conjunto de sinónimos asociados
- Un conjunto de hiperónimos e hipónimos asociados
- Un conjunto de palabras en singular y plural.

## **Etapa 2: Extracción de palabras claves del HTML y metadatos de tipo LangString**

En el estándar LOM la descripción y título del OA son asignados en los metadatos *General.description* y *General.title* respectivamente. La página Web (URL) o unidad educativa a la cual hace referencia el OA viene dada por el valor del metadato *Technical.location*. A continuación se describen los pasos en forma secuencial que se realizan.

### **1. Extracción de valores de los metadatos**

Por cada OA recuperado se extraen los valores de los metadatos *General.title*, *General.description* y *Tecnica.location* que serán utilizado en los siguientes pasos.



## 2. Extracción contenido HTML

En el caso que exista una dirección URL que haga referencia a un documento HTML o página web como valor del metadato *Technical.location*, se utilizan librerías propias del framework 4.0 y la librería AgilityHTMLPack [Html Agility Pack, 2010] para obtener el contenido o texto dentro de la página web asociada.

## 3. Extracción de palabras claves

Dado el contenido extraído de la página web (en caso de existir) y los valores de los metadatos *General.description* y *General.title* se procede a extraer las palabras claves utilizando un algoritmo que básicamente cuenta las ocurrencias de las palabras en el texto. También para realizar esta tarea se intentó utilizar un servicio llamado *TextWise* [TextWise, 1994] pero no tuvo éxito ya que el mismo no tiene en cuenta los stopwords del idioma, por ende las palabras claves obtenidas son insignificantes. Debido a esto se optó por codificar un algoritmo de extracción de palabras claves que sí tenga en cuenta los stopwords.

Con el fin de determinar si dada una palabra es un stopwords o no lo es, generamos un archivo XML de stopwords del idioma inglés con la siguiente estructura:

```
<stopwords>
  <stopword>a</stopword>
  <stopword>able</stopword>
  <stopword>about</stopword>
  <stopword>above</stopword>
  <stopword>abst</stopword>
  <stopword>accordance</stopword>
  <stopword>according</stopword>
  <stopword>accordingly</stopword>
  <stopword>across</stopword>
  ....
</stopwords>
```

El algoritmo codificado consiste en recorrer el texto, palabra por palabra, determinar si es o no un stopwords y contar el número de apariciones. Luego serán consideradas solo aquellas palabras claves que tienen densidad mayor al N%.

Por ejemplo:

Sean

$w_i$  = Número total de veces que aparece la palabra  $i$  en el texto.

$w$  = Número total de palabras en el texto.

Entonces el valor de la densidad de la palabra  $i$  es:  $D_i = \frac{w_i}{w}$

esto es, el número de apariciones de la palabra  $i$  en el texto sobre el número total de palabras.

Entonces en esta tesina proponemos que las palabras claves sean extraídas de la siguiente manera:

- **Metadatos:**
  - **General.title:** Se extraen las palabras claves con densidad mayor al **20%**.
  - **General.description:** Se extraen las palabras claves con densidad mayor al **10%**.
  
- **HTML**
  - **Title:** Se extraen las palabras claves del tag *title* del documento HTML con densidad mayor al **20%**.
  - **Body:** Se extrae las palabras claves del tag *body* del documento HTML con densidad mayor al **2%**.

Los valores de densidad fueron extraídos de [Rodríguez, 2009] de manera de que se refleje que las palabras claves obtenidas de los títulos son más significativas, por ende la densidad es mayor. Entonces por cada OA obtenemos un conjunto de palabras claves que son la unión de los conjuntos antes descriptos.

### Etapa 3: Cálculo de coincidencias

En la Figura 3.9 se pueden observar los conjuntos generados en las etapas anteriores:

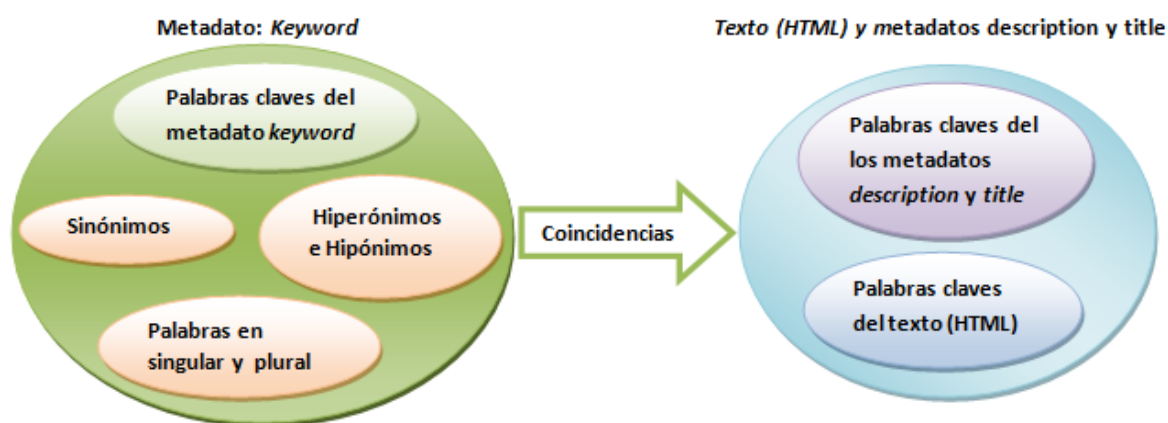


Figura 3.9: Conjuntos de palabras a analizar

En esta etapa proponemos calcular:

- Porcentaje de palabras claves obtenidas en la etapa 1 (Metadato: *General.keyword*) que están incluidas o coinciden en el conjunto de palabras claves obtenido en la etapa 2 (Texto (HTML)) y metadatos *General.description* y *General.title*.

- Porcentaje de sinónimos obtenidos en la etapa 1 que están incluidas o coinciden en el conjunto de palabras claves obtenido en la etapa 2.
- Porcentaje de hiperónimos e hipónimos obtenidas en la etapa 1 que están incluidas o coinciden en el conjunto de palabras claves obtenido en la etapa 2.
- Porcentaje de palabras en singular o plural relacionadas que está incluidas o coinciden en el conjunto de palabras claves obtenidos en la etapa 2.

Obtenidos estos porcentajes, se calcula un valor numérico que indica la coincidencia de las palabras claves. Para ello se tomó la decisión de utilizar y proponer el siguiente sistema de puntuación para indicar la coincidencia.

### 3.2.3.1 Sistema de Puntuación

La idea principal del sistema de puntuación propuesto es priorizar la coincidencia de las palabras claves. Es decir si todas las palabras claves coinciden entonces la calidad de keywords es óptima.

En el caso de que no sea óptima entonces se incrementa el valor obtenido usando la coincidencia de las demás palabras relacionadas (sinónimos, hiperónimos, hipónimos y palabras en singular/plural) de manera que la calidad aumente o se mantenga estable.

Para ello proponemos definir:

$$coincPC = \frac{cantPalabrasClavesIncluidas}{cantKeysTotal} \leq 1$$

$$coincSP = \frac{cantSingularPluralIncluidas}{cantSingularPluralTotal} \leq 1$$

$$coincS = \frac{cantSinonimosIncluidos}{cantSinonimosTotal} \leq 1$$

$$coincHH = \frac{cantHiperHiponimosIncluidos}{cantHiperHiponimosTotal} \leq 1$$

$$incremento = \frac{coincSP + 0,9*coincS + 0,5*coincHH}{2,4} \leq 1$$

$$CPCT = \begin{cases} 1 & coincPC = 1 \\ \frac{2*coincPC + incremento}{2 + incremento} & \text{En otro caso} \end{cases} \leq 1$$

Dónde:

**CPCT**: Coincidencias de palabras claves total.

*coincPC*: Porcentaje de palabras claves incluidas.

*coincSP*: Porcentaje de palabras en singular y plural incluidas.

*coincS*: Porcentaje de sinónimos incluidos.

*coincHH*: Porcentaje de hiperónimos e hipónimos incluidos.

*incremento*: Promedio ponderado de las extensiones (sinónimos, hiperónimos e hipónimos y palabras en singular/plural).

Para demostrar que **CPCT** es correcta debemos probar que:

$$\frac{2 * \text{coincPC} + \text{incremento}}{2 + \text{incremento}} \geq \text{coincPC} \quad \text{siempre que } \text{coincPC} \neq 1$$

Entonces despejando *incremento* tenemos:

$$\begin{aligned} 2 * \text{coincPC} + \text{incremento} &\geq \text{coincPC} * (2 + \text{incremento}) \\ 2 * \text{coincPC} + \text{incremento} &\geq 2 * \text{coincPC} + \text{coincPC} * \text{incremento} \\ \text{incremento} - \text{coincPC} * \text{incremento} &\geq 2 * \text{coincPC} - 2 * \text{coincPC} \\ \text{incremento} - \text{coincPC} * \text{incremento} &\geq 0 \\ \text{incremento}(1 - \text{coincPC}) &\geq 0 \\ \Rightarrow \text{incremento} &\geq 0 \end{aligned}$$

- ✓ Si  $\text{incremento} = 0 \Rightarrow \text{CPCT} = \text{coincPC} \Rightarrow$  se mantiene la prioridad.
- ✓ Si  $0 < \text{incremento} \leq 2,4$  (Máximo:  $0,9 + 0,5 + 1 = 2,4$ )  
 $\Rightarrow \text{coincPC} < \text{CPCT} = \text{coincPC} + \text{porcentaje} \leq 1$   
 $\Rightarrow$  se mantiene la prioridad de palabras claves devolviendo un valor menor a uno pero incrementado en un porcentaje.

### 3.3 Proceso Idiomas

El estándar LOM define dos metadatos que contemplan los idiomas del OA.

- *General.language*: Hace referencia al idioma en que debería estar escrito el contenido del OA, es decir a la unidad de contenido educativo, ya sea página web, documento, imagen, audio, entre otros. a la cual hace referencia.
- *MetaMetadata.language*: Hace referencia al idioma en que debería estar escrito los metadatos de tipo LangString (*General.title* y *General.description*) de la instancia de metadatos asociada al OA.

El proceso de Idiomas propuesto consiste básicamente en calcular la distancia a la que se encuentran los idiomas asignados en los metadatos de los detectados mediante un *algoritmo de detección de idiomas* en el contenido del OA en caso de ser una página web o archivo HTML y en los valores asignados en los metadatos *General.title* y *General.description*, para luego mediante un sistema de puntuación que se describe en las secciones posteriores calcular un valor numérico que represente la *calidad de idiomas* del OA.

En la Figura 3.10 se observan las distintas etapas del proceso, dónde el resultado CIT representa la “calidad de idiomas total”.

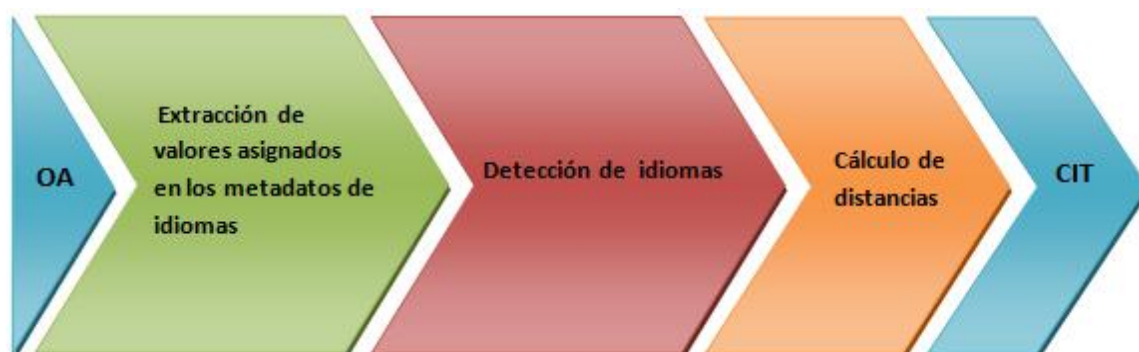


Figura 3.10: Etapas del proceso Idiomas

#### **Etapa 1: Extracción de valores asignados en los metadatos de idiomas**

Por cada OA recuperado se extraen los valores asignados en los metadatos *General.language* y *MetaMetadata.language* para luego ser utilizado en las demás etapas.

#### **Etapa 2: Detección de Idiomas**

En esta etapa se detectan los idiomas más probables de los valores asignados en los metadatos de tipo LangString (*General.title* y *General.description*) y además los idiomas del contenido de la página web relacionada al OA en caso de existir. En esta

tesina se propone detectar los idiomas del contenido de OA solo en el caso que sea una página web o archivo HTML. Para llevar a cabo esta tarea se codificó un algoritmo de detección de idiomas. A continuación se detallan las bases del algoritmo.

## **Características de algoritmos de detección de Idiomas**

Una técnica de destinada a la tarea de procesar, clasificar el texto, así como identificar el idioma permitiendo un cierto nivel de tolerancia a errores textuales, es el basado en la descomposición del texto en N-gramas [Cavnar, Trenkle, 1994]. Como se explicó en la Sección 1.5 un **N-grama** es una subsecuencia de  $n$  elementos de una secuencia dada.

Existen algunos idiomas que tienen la característica de poseer alguna palabra en su lenguaje de forma que dicha palabra ocurre con más frecuencia que otras palabras. Una de las maneras más comunes para expresar esta idea se conoce como la **ley de Zipf**. Particularmente esta ley es verdad para la ocurrencia de N-gramas más comunes para un determinado idioma.

La idea principal es: *Dado un texto de prueba hallar el conjunto de N-gramas junto con sus frecuencias y calcular las distancias entre este conjunto hallado contra todos los conjuntos de N-Gramas de muestra correspondientes a cada uno de los idiomas a evaluar, y se considera el idioma donde la distancia sea mínima. Es decir el idioma más probable será aquel dónde su conjunto de N-Gramas relacionado esté más cerca al conjunto de N-Gramas del texto de prueba. Cabe mencionar que los conjuntos de N-Gramas de muestra son conjuntos hallados desde textos de muestra específicos durante el entrenamiento del algoritmo.*

Cavnar y Trenkle mencionan los siguientes métodos que utilizan métricas de distancia diferente:

1. N-grama Rank-Order
2. N-grama de cálculo de frecuencias

Los dos métodos al igual que como se mencionó anteriormente utilizan un conjunto de N-Gramas de muestra para cada idioma. Este proceso es realizado durante el entrenamiento del método.

En [Ahmed et al, 2004] se realizó una experimentación dónde recogieron muestras de textos en 12 idiomas para obtener los conjuntos de N-Gramas de los diferentes idiomas y así poder obtener el  $N$  óptimo de N-gramas a considerar. En los conjuntos de N-gramas eliminaron caracteres que no son de mucha utilidad en la

diferenciación entre los idiomas. En la Tabla 3.3 se puede ver los cálculos realizados en la recolección de estos conjuntos de N-Gramas.

Language	Number of training files	Size of all training files	Total N-grams	N-gram Sizes
Danish	20	88K	41,485	2 to 7
Dutch	20	67K	30,276	2 to 7
English	20	81K	36,633	2 to 7
French	20	92K	42,108	2 to 7
German	20	80K	35,524	2 to 7
Italian	20	65K	29,878	2 to 7
Polish	20	104K	41,116	2 to 7
Portuguese	20	67K	33,574	2 to 7
Romanian	20	105K	40,625	2 to 7
Spanish	20	65K	32,983	2 to 7
Swedish	20	89K	38,591	2 to 7
Tagalog	20	65K	29,316	2 to 7
<b>Totals</b>	<b>240</b>	<b>968K</b>	<b>432,109</b>	

Tabla 3.3: Ejemplos de entrenamientos y total y tamaño de N-gramas

Como resultado en esta experimentación los autores obtuvieron que  $N=3$  es el valor óptimo a considerar para los N-Gramas. Entonces  $N=3$  es el valor que utilizaremos en esta tesina.

### *N-grama Rank-Order*

En este método cada conjunto de N-Gramas correspondiente a los diferentes idiomas hallado en el entrenamiento está ordenado por frecuencia de aparición de los N-gramas que los componen. En la Tabla 3.4 se ilustra la forma de calcular la distancia entre conjuntos de N-Gramas.

	Language profile	Test document profile	Out of place
Most Frequent	TH	TH	0
	ER	ING	3
	ON	ON	0
	LE	ER	2
	ING	AND	1
Least Frequent	AND	ED	No-match = max distance
Test Document Distance from Language = Sum of out-of-place values			

Tabla 3.4: Cálculo de distancia utilizando Rank-Order

La columna “Language profile” de la Tabla 3.4 hace referencia al conjunto de bi-gramas de muestra del idioma inglés hallado en el entrenamiento, ordenados por frecuencia de aparición, es decir el bi-grama TH aparece con más frecuencia que el bi-grama ON. La columna “Test document profile” hace referencia al conjunto ordenado de bi-gramas hallado del texto de prueba. La columna “Out of place” muestra la distancia de *Manhattan* entre bi-gramas del conjunto correspondiente al texto de prueba y del conjunto de bi-gramas de muestra.

El método es simple y funciona bien en la identificación del idioma en secuencia de textos de 300 bytes o más. Es insensible a los errores tipográficos. Una desventaja de este método es que hace un uso intensivo de recursos y consume tiempo ya que utiliza conjuntos de N-Gramas ordenados y esta operación disminuye el rendimiento.

## *N-grama de cálculo de frecuencias*

Este método no utiliza los conjuntos de N-Gramas ordenados sino que utiliza las frecuencias de ocurrencias de los N-gramas como se describe en [Ahmed et al, 2004], que se detalla en el siguiente apartado, junto con la extensión realizada para esta tesina.

### **Algoritmo de detección de idioma propuesto**

El algoritmo para detección de idiomas propuesto se basó en el algoritmo de **N-grama de cálculo de frecuencias**, el cual se extendió para que devolviera no solo el idioma más probable sino una lista de idiomas, ordenados del más probable al menos probable.

Los idiomas contemplados en la detección son:

- |                 |                   |                |
|-----------------|-------------------|----------------|
| 1. Inglés (en)  | 4. Italiano (it)  | 7. Alemán (de) |
| 2. Español (es) | 5. Holandés (nl)  |                |
| 3. Francés (fr) | 6. Portugués (pt) |                |

Dado los resultados que obtuvieron los autores en la experimentación se utilizó **N=3** en el algoritmo, por ende se utilizan tri-gramas (3-gramas).

El algoritmo se entrenó con textos de muestra específicos para cada idioma publicados en [Barber, 2009], dando como resultado siete conjuntos de tri-gramas. Los mismos fueron agrupados en un único archivo XML. Cada tri-grama tiene asociado un valor numérico que determina la frecuencia de aparición del mismo en el texto y el idioma al cual pertenece.

La estructura del archivo XML propuesta es la siguiente:

```
<trigrams>
  <trigram name="ame" language="es">64</trigram>
  <trigram name="med" language="es">53</trigram>
  ....
  <trigram name="rém" language="pt">8</trigram>
  <trigram name="pou" language="pt">33</trigram>
  ....
  <trigram name="ale" language="it">94</trigram>
  <trigram name="fac" language="it">60</trigram>
  ....
  <trigram name="nie" language="fr">36</trigram>
  <trigram name="nou" language="fr">69</trigram>
  ....
  <trigram name="ark" language="en">23</trigram>
  <trigram name="rke" language="en">11</trigram>
  ....
  <trigram name="rbü" language="de">3</trigram>
  <trigram name="bür" language="de">13</trigram>
  ....
  <trigram name="eli" language="nl">159</trigram>
  <trigram name="jkn" language="nl">2</trigram>
  ....
</trigrams>
```

El tri-grama "ame" aparece 64 veces en el texto de muestra escrito en idioma "es" (Español)



El algoritmo propuesto utiliza otra métrica de la mencionada en [Ahmed et al, 2004] para el cálculo de distancia, nombrada en [Barber, 2009]. A continuación se describe la métrica de cálculo de distancia utilizada en la propuesta, como así también la extensión de la misma.

Sea:

**ConjuntoTGramas (CTG):** Conjunto de tri-gramas junto con sus frecuencias encontradas en el texto de prueba.

**Frec(i,t):** Frecuencia (número) del tri-grama **i** en el conjunto de tri-gramas **t**.

**Total(k):** Número total de ocurrencias de tri-gramas del idioma **k**. (Calculado sobre el texto de muestra)

$$\text{IdiomaMasProbable} = \max_{j:1..N} \left( \sum_{\text{trig} \in \text{CTG}} \frac{\text{Frec}(\text{trig}, j)}{\text{Total}(j)} * \frac{\text{Frec}(\text{trig}, \text{CTG})}{\#\text{CTG}} \right)$$

Dónde **j: 1...N** hace referencia al conjunto de tri-gramas junto con sus frecuencias del idioma **j**. (Tomados de un texto de muestra).

Entonces utilizando la definición anterior además de obtener el idioma más probable permite también obtener la lista de idiomas más probables. En esta tesina utilizaremos la lista de idiomas más probables.

### **Etapa 3: Cálculo de distancias**

Esta etapa tiene como función calcular la distancia entre el idioma asignado y el idioma más probable utilizando el sistema de puntuación propuesto, el mismo se detalla a continuación.

### **Sistema de Puntuación**

Como se señaló anteriormente en el estándar LOM el metadato *General.language* hace referencia al idioma utilizado por defecto en el contenido del OA y el metadato *MetaMetadata.language* determina dada la instancia de metadatos, el idioma por defecto en que deberían estar escritos los valores de los metadatos de tipo *LangString* (Por ejemplo: *General.title* y *General.description*).

La idea básicamente es calcular qué tan lejos está el idioma asignado en los metadatos del idioma detectado (usando el algoritmo de detección de idiomas propuesto) contando los lugares o sea la *distancia*.

Por ejemplo:

Supongamos que el idioma asignado en el metadato es **EN**.

La lista de idiomas devueltos por el algoritmo es = **DE | FR | NL | EN | IT | PT | ES**

Esto significa que el idioma más probable detectado en el OA es DE = Alemán, el segundo más probable es FR=Francés, etc. De esta manera podemos ver que el idioma asignado EN esta a distancia 3 de ser el más probable.

$$\overbrace{\text{DE | FR | NL}}^{\text{Distancia} = 3} \text{ | EN | IT | PT | ES}$$

Para calcular la coincidencia o aproximación del idioma asignado proponemos lo siguiente:

$$CI = 1 - \frac{\text{diferencia}}{\text{longListaIdiomas}} \leq 1$$

Dónde:

**CI:** Coincidencia o aproximación del idioma asignado.

**diferencia:** Distancia del idioma más probable detectado (posiciones o lugares).

**longListaIdiomas:** Longitud de la lista de idiomas devuelta. En el ejemplo es 7.

Entonces proponemos para calcular la coincidencia total de idiomas del OA la siguiente definición.

Sea:

$CI^M$ : Coincidencia o aproximación del idioma asignado en el metadato *MetaMetadata.language*.

$CI^G$ : Coincidencia o aproximación del idioma asignado en el metadato *General.language*.

$$CIT(*) = \begin{cases} \frac{CI^M + CI^G}{2} & CI^M \neq 0 \text{ y } CI^G \neq 0 \\ CI^M & CI^M \neq 0 \text{ y } CI^G = 0 \\ CI^G & CI^M = 0 \text{ y } CI^G \neq 0 \\ 0 & \text{En otro caso} \end{cases} \leq 1$$

(\*) Es la calidad total de idiomas del OA utilizando las coincidencias calculadas para los dos casos.

### 3.4 Proceso Confiabilidad

El proceso Confiabilidad propuesto provee un sistema de votación de manera que los usuarios finales participen en la calificación del OA puntuándolo teniendo en cuenta diferentes aspectos o puntos de vistas del OA y así determinar la confiabilidad del OA.

En Merlot [Merlot] se propone calificar a los OA en las siguientes dimensiones:

1. **Facilidad de Uso (Usabilidad):** La cuestión básica que persigue esta dimensión es cuán fácil es el uso del objeto durante la primera vez que se utiliza. Los planteamientos que ayudarían a evaluar esta dimensión están relacionados al valor estético y la provisión de realimentación a las respuestas de los usuarios. Fácil navegación, interfaz de usuario intuitiva y calidad de la interfaz de ayuda.
2. **Calidad del contenido:** Esta dimensión apunta a la veracidad, certeza, presentación balanceada de las ideas y apropiado nivel de detalle. Claridad de contenido, la representación del contenido debe facilitar su correcta comprensión.
3. **Potencial de efectividad:** Esta dimensión se refiere a la alineación entre los objetivos de enseñanza, actividades, evaluaciones y características del estudiante o profesor.

En esta tesina proponemos agregar las siguientes dimensiones:

4. **Descripción del objeto:** Esta dimensión trata la descripción en general del objeto en forma detallada y ordenada, metadatos incompletos, información relevante y faltante.
5. **Autoevaluación:** Esta dimensión permite que el usuario autoevalúe su conocimiento o experiencias sobre el tema o temas que trata el OA.

En la Figura 3.11 se observa la interfaz gráfica que permite realizar la votación. Las puntuaciones también se basan en una escala de *Likert* que van desde 1 a 5 y N/A, siendo 5 la mejor puntuación y N/A (no vota o no contesta) la peor puntuación. Cabe recalcar que los usuarios al momento de calificar *Potencial de efectividad* deben considerar aspectos pedagógicos para determinar un puntaje acorde a los resultados obtenidos.



Figura 3.11: Interfaz gráfica para realizar la votación

## Ponderación de los perfiles

La idea principal es ponderar la votación utilizando el perfil del usuario actual que realiza la votación. En esta tesina consideramos que un usuario con un perfil o profesión **Teacher** debe tener un voto con más peso que el voto de una persona con perfil **Student**. Por lo tanto proponemos utilizar esta idea definiendo los siguientes pesos.

Sea  $pesos_k$  el conjunto de pesos donde  $k \in \{Teacher, Student\}$

Proponemos los siguientes pesos:

$$pesos_{Teacher} = \{1; 1; 1; 1\}$$

$$pesos_{Student} = \{0,75; 0,75; 0,75; 0,75\}$$

Dónde cada valor en la posición  $i$  del conjunto corresponde al peso de la pregunta  $i$  de la encuesta. De esta manera los votos de un usuario con perfil **Teacher** tienen más peso que el voto de un usuario con perfil **Student**.

Por ejemplo:

$pesos_{Teacher} [2] = 1$ , esto significa que el peso para la pregunta *Calidad de Contenido* es 1 en el perfil **Teacher**.

A continuación proponemos un sistema de puntuación que utiliza los pesos definidos.

## Sistema de puntuación

Además de ponderar la votación utilizando el perfil del usuario actual también se tiene en cuenta la autoevaluación obtenida del usuario sobre sus conocimientos y experiencias personales sobre el tema como valores de pesos. Es decir un usuario con un perfil o profesión **Teacher** debe tener un voto con más peso que el voto de una persona con perfil **Student** y además proponemos que un usuario dónde en su autoevaluación obtiene un puntaje mayor que otro usuario debería influir en mayor medida en el peso del voto que el usuario con autoevaluación menor.

En este sistema de puntuación decidimos que el valor que determina la confianza final del OA es el promedio ponderado de todas las opciones.

Para facilitar el cálculo del promedio ponderado proponemos que cada puntaje tenga asociado un valor numérico menor a 1.

Proponemos utilizar los siguientes números asociados:

NA	→	0
1	→	0,2
2	→	0,4
3	→	0,6
4	→	0,8
5	→	1

Sea:

$puntaje_{Auto}$ : Puntaje obtenido en la autoevaluación sobre el tema o temas que trata el OA. (Dimensión 5).

Entonces se propone que la confiabilidad del OA sea medida por:

$$COA = \frac{\sum_{i=1}^4 \mathit{puntaje}_{Auto} * \mathit{pesos}_k[i] * \mathit{puntaje}_i}{\sum_{i=1}^4 \mathit{puntaje}_{Auto} * \mathit{pesos}_k[i]} \quad k: \text{Perfil de usuario}$$

Dónde:

**COA**: Confiabilidad del OA.

**$\mathit{puntaje}_i$** : Puntaje obtenido o votado (números asociados) en la pregunta  $i \in \{1, 2, 3, 4\}$

**$\mathit{pesos}_k[i]$** : Peso de la pregunta  $i$  para el perfil  $k$ .

Los votos obtenidos son guardados en una Base de Datos local y promediados con la votación ya existente para el OA en cuestión. Luego dado que los datos son almacenados en una Base de Datos local podrán ser consultados en búsquedas posteriores de OAs si así lo requiera para calificar nuevamente al OA recuperado.

## 4. Implementación de la propuesta

### 4.1 Tecnologías utilizadas

Para el desarrollo e implementación del presente trabajo se utilizaron tecnologías Microsoft. Puntualmente la propuesta consiste en una aplicación web desarrollada bajo framework 4.0 en lenguaje C# utilizando Microsoft Visual Web Developer 2010 express y como servidor de base de datos SQL Server express 2008. Como host de la aplicación web se utilizó Internet Information Services 7.0 (IIS7). Para almacenar la información extraída de WordNet, Tri-gramas y Stopwords se utilizaron archivos XML (eXtensible Markup Language).

### 4.2 Repositorios utilizados

Los repositorios junto a sus características que se tuvieron en cuenta para la propuesta fueron:

1. **OER Commons:** Este repositorio se basa en el estándar OAI-PMH [Lagoze et al, 2002] para la recuperación de objetos mediante consultas en la URL. El problema es que las consultas realizadas con este estándar no permiten la búsqueda mediante palabras claves sino por rangos de fechas y identificador del objeto entre otras.
2. **Ariadne:** Este repositorio permite realizar las consultas vía REST o SOAP, y es elegido por contar con un lenguaje de consulta bastante expresivo llamado "Lucene" que permite realizar búsquedas por palabras claves y además por contener un importante número de objetos de aprendizaje en idioma Inglés.

Notar que la interface de consulta SQI solo se encarga de enviar y recibir las consultas, no de la estructura de las mismas. Por lo tanto distintos repositorios que utilicen a SQI para su interoperabilidad podrán tener lenguajes de consulta con diferente grado de expresividad.

Por lo tanto, decidimos que **Ariadne** es el más conveniente para la realización de consultas y extracción de OAs.

### 4.3 Prototipo

#### 4.3.1 Arquitectura

La arquitectura de la aplicación web fue diseñada en capas. En la figura 4.1 se puede observar la arquitectura.



Figura 4.1: Arquitectura

➤ La primera capa (**Presentación**) está conformada por clases que permiten desplegar y visualizar la interfaz gráfica al usuario. Es la capa con la cual el usuario interactúa e ingresa datos.

➤ La segunda capa (**Admin. de Procesos y Admin. de Repositorios**) es una capa de lógica de negocios (LN) que contiene clases Manager o Administrador, las cuales tienen la función de administrar, decidir y lanzar peticiones a la capa inferior para ejecutar los distintos procesos y recuperar los OAs. También aquí contiene las clases necesarias para autenticarse y crear la sesión de conexión al repositorio de OAs **Ariadne**.

➤ La tercera capa es también una capa de LN pero con menos nivel de abstracción que la segunda capa. Esta capa interactúa en forma directa con la capa de **Accesos a Datos**, envía diferentes peticiones de consulta, ya sea a XMLs o BDs o Web Services (WS).

➤ La cuarta capa (**Acceso a datos**), contiene las clases necesarias para manipular archivos XMLs, para conectarse y realizar consultas a la base de datos SQL Server express 2008 y contiene las clases para utilizar WS correspondientes al repositorio de OAs **Ariadne**.

➤ La capa **Entidades** es una capa transversal a las demás capas, en la cual se definen las clases de las entidades utilizadas en todas las capas. Por ejemplo, entidades como: LearningObject, FileHTML, LOM y Preferencia.

Todas las llamadas son realizadas por llamadas a procedimientos desde las capas superiores a las inferiores. En el prototipo aquí presentado no fue contemplado que las capas inferiores se comuniquen con las capas superiores, en caso que existiese la utilización de eventos es una opción apropiada.

## 5. Experimentación

Para probar el prototipo se realizaron diferentes experimentaciones, las mismas podemos dividirlas en dos aspectos fundamentales o dos preguntas que deberemos responder:

**Pregunta 1:** ¿Los objetos de aprendizaje de mayor calidad tienden a ubicarse en las primeras posiciones luego de realizar una búsqueda mediante palabras claves?

**Pregunta 2:** ¿Los objetos de aprendizaje devueltos luego de una búsqueda se adaptan y se personalizan al perfil del usuario que realiza la consulta?

A continuación se detallan cada una de las experimentaciones.

Cabe recordar que la personalización de los resultados de búsquedas por perfiles es efectuada en el *proceso Metadatos* donde se prevé que cada perfil tiene mayor preferencia por determinadas categorías de metadatos que otras. Entonces la ponderación de las categorías de metadatos determina la personalización de los resultados.

### 5.1 ¿Los objetos de aprendizaje de mayor calidad tienden a ubicarse en las primeras posiciones?

Debemos verificar si los OAs de calidad devueltos en una búsqueda escalan las primeras posiciones en el ranking de OAs. Para ello realizamos 20 búsquedas al repositorio de objetos **Ariadne** con diferentes palabras claves devolviendo 100 resultados por cada búsqueda descartando cualquier tipo de personalización por perfil de usuario.

Las palabras claves utilizadas fueron las siguientes:

- |  |                          |
|--|--------------------------|
| 1. Biology marine                      | 6. History German        |
| 2. Ecology                             | 7. Modern architecture   |
| 3. Artificial intelligence programming | 8. Operative system Unix |
| 4. Computer science                    | 9. TCP                   |
| 5. Cancer disease                      | 10. Tectonics plates     |

Por cada palabra clave o consulta se realizaron dos búsquedas, donde en una fue considerada la calidad de los OA (prototipo) y en la otra no, es decir por cada palabra clave obtuvimos dos conjuntos o casos de OAs:



Caso **A**: Resultados crudos devueltos por el repositorio.

Caso **B**: Resultados devueltos por el prototipo.

De los 100 resultados devueltos por cada búsqueda, es decir la búsqueda del caso **A** y la búsqueda del caso **B** se seleccionaron los primeros 20 resultados y se calculó el promedio total de calidad de los objetos de manera de obtener un valor numérico que represente la calidad del conjunto devuelto en las primeras posiciones. En la Tabla 5.1 se observan las calificaciones obtenidas en las diferentes búsquedas para los dos casos y en la Figura 5.1 el gráfico de barras correspondiente.

Casos Búsquedas	A	B
1	0,3745	0,4823
2	0,3196	0,5314
3	0,4654	0,5670
4	0,4820	0,5482
5	0,4579	0,5207
6	0,4423	0,5262
7	0,3490	0,5143
8	0,3827	0,5569
9	0,3645	0,4972
10	0,4088	0,6094
<b>Promedio</b>	<b>0,4047</b>	<b>0,5355</b>

Tabla 5.1: Calificaciones promedios obtenidas en las 10 búsquedas

Para más detalles sobre las calificaciones de las distintas búsquedas dirigirse al Apéndice 8.1.1.

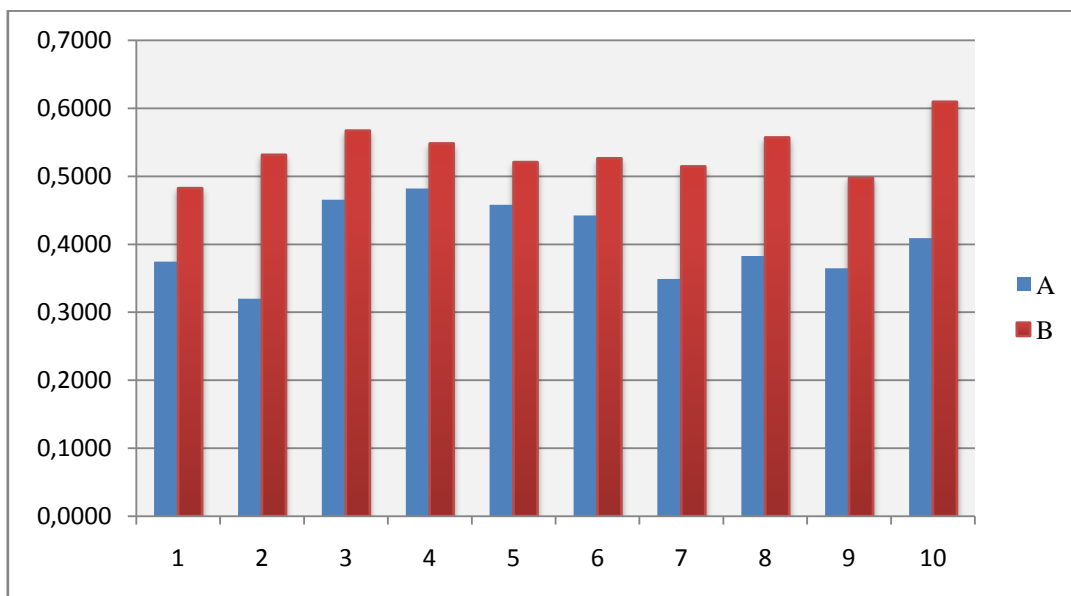


Figura 5.1: Gráfico de las calificaciones obtenidas para los diferentes casos

Como se observa en la Figura 5.1 (Caso **B**) los objetos de mayor calidad tienden a escalar las primeras posiciones dando como consecuencia una mayor calificación en promedio de los objetos de aprendizaje recuperados. Además al comparar los promedios generales de los dos casos, vemos  $PromB = 0,5355 > PromA = 0,4047$  por ende en promedio se satisface lo anterior.

## 5.2 ¿Los objetos de aprendizaje se adaptan y se personalizan al perfil de usuario que realiza la consulta?

Debemos verificar o validar que los resultados de búsquedas se adaptan al perfil del usuario que realiza la búsqueda. Para ello se contó con la ayuda de usuarios con diferentes perfiles y se consideró el feedback otorgado por cada uno para establecer si el prototipo satisface, en algún grado, los diferentes perfiles.

Los perfiles que se consideraron son **Teacher** y **Student**.

En la Tabla 5.2 podemos observar los diferentes usuarios y los objetivos que desean alcanzar.

Usuario	Perfil	Objetivo
<b>Profesor1</b>	Teacher	Preparar el curso de programación en java para un nivel Inicial
<b>Profesor2</b>	Teacher	Preparar el curso de números complejos para un nivel Inicial
<b>Estudiante1</b>	Student	Aprender las bases, conceptos y herramientas con las que cuenta Java. Paradigma POO.
<b>Estudiante2</b>	Student	Aprender conceptos, teoremas, operaciones y definiciones sobre números complejos

Tabla 5.2: Usuarios con diferentes perfiles y objetivos

Se realizaron las siguientes búsquedas

- ✓ El usuario **Profesor1** realizó una búsqueda con la consulta “java”.
- ✓ El usuario **Estudiante1** realizó una búsqueda con la consulta “java”.
- ✓ El usuario **Profesor2** realizó una búsqueda con la consulta “complex numbers”.
- ✓ El usuario **Estudiante2** realizó una búsqueda con la consulta “complex numbers”.

Aquí las pruebas también se dividieron en dos casos:

**Caso A: Realización de búsquedas con el perfil asignado:** Se realizaron las siguientes acciones o casos de uso para luego ser analizados y evaluados:

- I. El usuario *Profesor1* realiza la búsqueda con la consulta “java” con su perfil asignado **Teacher** y ordena a su criterio los resultados (ranking).
- II. El usuario *Profesor2* realiza la búsqueda con la consulta “complex numbers” con su perfil asignado **Teacher** y ordena a su criterio los resultados (ranking).
- III. El usuario *Estudiante1* realiza la búsqueda con la consulta “java” con su perfil asignado **Student** y ordena a su criterio los resultados (ranking).
- IV. El usuario *Estudiante2* realiza la búsqueda con la consulta “complex numbers” con su perfil asignado **Student** y ordena a su criterio los resultados (ranking).

**Caso B: Análisis de resultados de búsquedas del perfil opuesto:** Se realizaron las siguientes acciones o casos de uso para luego ser analizados y evaluados:

- I. El usuario *Profesor1* ordena a su criterio los resultados de búsquedas obtenidos por el perfil opuesto. O sea analiza los resultados obtenidos por *Estudiante1*.
- II. El usuario *Profesor2* ordena a su criterio los resultados de búsquedas obtenidos por el perfil opuesto. O sea analiza los resultados obtenidos por *Estudiante2*.
- III. El usuario *Estudiante1* ordena a su criterio los resultados de búsquedas obtenidos por el perfil opuesto. O sea analiza los resultados obtenidos por *Profesor1*.
- IV. El usuario *Estudiante2* ordena a su criterio los resultados de búsquedas obtenidos por el perfil opuesto. O sea analiza los resultados obtenidos por *Profesor2*.

El análisis de los resultados consistió en evaluar la cercanía entre el ranking devuelto por el prototipo del ranking propio que realizó el usuario utilizando sus propios criterios de selección y elección de OA, para ello utilizamos la distancia de *Manhattan*, la cual es adecuada para capturar distancias entre posiciones. ***La distancia es calculada entre las posiciones de los diez (10) primeros objetos ordenados por el usuario con las posiciones correspondientes en la lista devuelta por el prototipo.***

A continuación se detalla una breve descripción de la distancia de *Manhattan*.

Sea:

$OAP = [OR_1, OR_2, OR_3, \dots, OR_n]$  la lista de objetos de aprendizaje devueltos por el prototipo.

$OAU = [OA_1, OA_2, OA_3, \dots, OA_n]$  la lista de objetos de aprendizaje ordenados por el usuario (ranking).

Entonces la distancia entre el ranking otorgado por el prototipo y el ranking realizado por el usuario es:

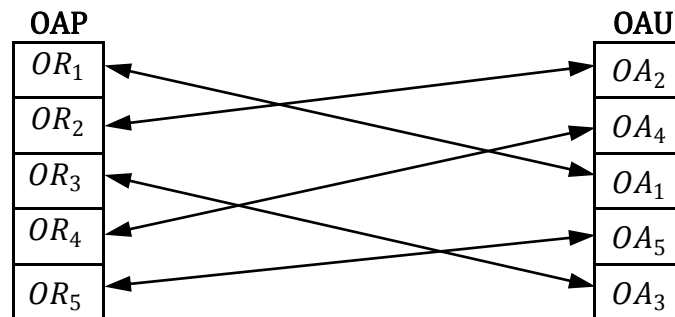
$$\begin{aligned} DistM(OAP, OAU) &= \sum_{i=1}^n |OR_i - OA_i| \\ &= |1 - OA_1| + |2 - OA_2| + \dots + |n - OA_n| \end{aligned}$$

Se considera peor caso para cada usuario cuando:

$$OA_1 = OR_n ; OA_2 = OR_{n-1} ; OA_3 = OR_{n-3} ; \dots ; OA_n = OR_1$$

Por ejemplo:

Sea  $n=5$ .



Entonces la distancia es calculada entre las diferentes posiciones.

$$DistM(OAP, OAU) = |1 - 5| + |2 - 4| + |3 - 3| + |4 - 2| + |5 - 1| = 8$$

$$\text{Pero Caso} = |1 - 5| + |2 - 4| + |3 - 3| + |4 - 2| + |5 - 1| = 12$$

En la Tabla 5.3 se observa la distancia de *Manhattan* entre los 10 primeros resultados de búsqueda obtenidos con el prototipo entre los diferentes perfiles (Teacher y Student) asignados a los usuarios participantes de la experimentación.

	Estudiante1	Estudiante2
Profesor1	29	-
Profesor2	-	20

Tabla 5.3: Distancia de Manhattan entre los resultados de búsqueda entre perfiles

En ambos, el peor caso en distancia entre resultados es 50. Para más detalles sobre el cálculo de distancia correspondiente a la Tabla 5.3 dirigirse al Apéndice 8.1.2. También en el Apéndice 8.1.2 se podrá ver los resultados de búsquedas recuperados.

En las tablas Tabla 5.4 y Tabla 5.5 se puede observar los cálculos de distancias obtenidos por los diferentes usuarios. La columna **A** indica los valores de distancias obtenidos contra la salida original del prototipo luego de la realización de las búsquedas de los usuarios con sus correctos perfiles sobre los 10 primeros OA (caso **A**). En la columna **B** hace referencia al caso **B**, dónde se intenta validar que los resultados de búsquedas realmente son personalizados y se adaptan al perfil asignado al usuario, por lo tanto en esta columnas se observan los resultados obtenidos en las búsquedas que realizaron los usuarios con su perfil asignado y en el análisis de los resultados del perfil opuesto.

Distancia Usuario	A	B	Peor Caso
	Teacher	Student	
Profesor1	14	19	50
Profesor2	7	20	50
<b>Promedio</b>	<b>10,5</b>	<b>19,5</b>	<b>50</b>

Tabla 5.4: Distancias de los usuarios con perfil **Teacher** en el caso A y B

Distancia Usuario	B	A	Peor Caso
	Teacher	Student	
Estudiante1	30	10	50
Estudiante2	22	6	50
<b>Promedio</b>	<b>26</b>	<b>8</b>	<b>50</b>

Tabla 5.5: Distancias de los usuarios con perfil **Student** en el caso A y B

En el Apéndice 8.1.3 se puede observar los resultados de búsquedas de los diferentes usuarios junto a las distancias obtenidas con el ranking realizado.

Como puede observarse en la Tabla 5.4, los resultados ordenados por los profesores en promedio se encuentran a menor distancia **10,5** de los resultados devueltos por el perfil **Teacher** que de los resultados devueltos por el perfil **Student**, siendo **Teacher** el perfil asignado a los profesores podemos señalar que los resultados tienden a personalizarse perfil del usuario que realiza la búsqueda. Lo mismo ocurre para los estudiantes, como se observa en la Tabla 5.5, los resultados ordenados por los estudiantes están a distancia **8** en promedio que los resultados devueltos por el perfil **Student**.

## 6. Discusiones

En este trabajo se ha presentado la propuesta de una aplicación web que permite encontrar objetos de aprendizaje de una cuidada calidad desde el punto de vista de sus metadatos. Se ha utilizado *WordNet* como recurso lingüístico para la extensión de palabras claves en el idioma inglés. El prototipo se ha desarrollado sobre el framework 4.0, como entorno de desarrollo se utilizó Microsoft Visual Developer 2010 express y como servidor de base de datos a SQL server 2008 express. Se optó utilizar *Ariadne* como repositorios de objetos de aprendizaje. Además cabe destacar que se han utilizado archivos XML para la administración y consulta de datos dentro de la aplicación como así también web services externos que provee el repositorio para entablar la conexión con él y realizar consultas. En las siguientes secciones se discutirán algunos puntos.

### 6.1 Repositorios

*Ariadne* fue el repositorio elegido entre los analizados por contar con varias ventajas, entre ellas trabajar con un lenguaje de consulta expresivo llamado “Lucene”, por tener la posibilidad de realizar búsqueda por palabras claves, por ser gratuito y además por contar con un gran número de objetos de aprendizaje en idioma inglés con metadatos bajo el estándar LOM en comparación con otros repositorios. A pesar de la ventajas hubo algunos inconvenientes durante las pruebas del prototipo, nos encontramos con que muchos objetos de aprendizaje no tenían ningún identificador que los identifique, poseían links rotos, algunos ni siquiera hacían referencia a un recurso digital, o sea solo eran un conjunto de metadatos que describían algo que no existía.

### 6.2 Prueba del Prototipo

Los procesos aplicados a los resultados de búsquedas generó que los objetos de aprendizaje de mayor calidad escalen las primeras posiciones en los resultados de búsqueda, como se mostró en las experimentaciones en promedio los objetos devueltos por el prototipo cuentan con mayor calidad en sus metadatos que los objetos devueltos en forma predeterminada por el repositorio.

Fue difícil la evaluación de la personalización de los resultados por perfil de usuario, ya que lo que se busca es alcanzar los mejores resultados en el proceso de enseñanza-aprendizaje, lo cual es difícil de evaluar. Para ello se realizó una evaluación por parte de usuarios (estudiantes y profesores) del sistema, respecto a su visión de los objetos devueltos por el prototipo, considerando si los objetos le resultan útiles y si el orden es el correcto. La experimentación determinó que los resultados devueltos para el perfil *Student* se adaptan de manera más eficiente a los estudiantes que a los profesores. Igualmente notar que los objetos devueltos para el perfil *Teacher* también fueron efectivos pero en menor medida. En conclusión los usuarios encontraron un ranking de objetos que tiende ser más acorde a sus intereses y necesidades en los devueltos en su

perfil asignado y no en el perfil opuesto, por ende los resultados fueron personalizados correctamente.

### 6.3 Trabajo Futuro

- ✓ En el proceso *Keywords* extraer palabras claves en textos escritos en idiomas diferentes al inglés, como así también en textos multi-idiomias.
- ✓ En el proceso *Idiomias* extender los idiomas soportados agregando los conjuntos de tri-gramas con sus respectivas frecuencias para cada nuevo idioma.
- ✓ Agregar un nuevo proceso que analice aspectos puramente pedagógicos de manera de obtener un análisis en detalle desde este punto de vista.
- ✓ Análisis en profundidad de:
  - Disponibilidad del recurso digital al cual hace referencia el OA.
  - Correctitud de la URL (formación) que apunta al recurso digital.
  - Concordancia entre lo asignado en el metadato *Technical.format* con el recurso educativo en sí del OA.
  - Análisis de confiabilidad de la fuente del OA. Nivel de confianza en los creadores o fuentes de contenidos debido a su reconocimiento, prestigio o experiencia en el área del conocimiento.
- ✓ Ampliar los aspectos utilizados para la personalización de resultados de búsquedas. Es decir no solo considerar el perfil sino también, por ejemplo: nivel educativo, edad, estilo de aprendizaje, preferencias, etc.
- ✓ Ampliar la batería de repositorios de objetos de aprendizaje de manera de poder buscar en varios simultáneamente.
- ✓ Medición potencial para aumentar la “reusabilidad” en los OA. Para poder aumentar los niveles de reusabilidad de OAs en repositorios de libre acceso sería necesario facilitar la introducción (o la obtención) automática e inteligente de metadatos bien sea por parte del creador del OA como también por parte de los evaluadores de dichos OAs. Para ello, sería necesario poner a disposición de estos actores herramientas inteligentes que permitan deducir metadatos a partir del propio OA, de los comentarios de sus creadores o revisores o de otras fuentes de información.

## 7. Bibliografía

[Ahmed et al, 2004] Bashir Ahmed, Sung-Hyuk Cha, and Charles Tappert. "Language Identification from Text Using N-gram Based Cumulative Frequency Addition". Proceedings of Student/Faculty Research Day, CSIS, Pace University, May 7th, 2004

[Barber, 2009] Ian Barber. Language detection with n-grams. <http://phpir.com/language-detection-with-n-grams>. Nov. 2009.

[Bertran, 2004] Dane Bertran. Likert Scales. University of Windsor. March 2004.

[Brill, 1993] Brill Eric. A simple rule-based part of speech tagger. Department of Computer Science University of Pennsylvania Philadelphia, Pennsylvania 19104. 1993

[Caplan, 2003] Caplan, P. Metadata Fundamentals for All Librarians. Chicago: American Library Association, 2003.

[Cavnar, Trenkle, 1994] W.B. Cavnar and J.M. Trenkle. N-gram based text categorization. In Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval, pages 161–169. 1994

[Duncan, 2003] Duncan, C. Granularization. En A. Littlejohn (Ed.), *Reusing Online Resources: A Sustainable Approach to E-learning* (pp 12-19). Open & Flexible Learning Series. London, Sterling: VA Taylor & Francis. 2003.

[García, 2005] García Aretio, Lorenzo. Editorial del BENED. Objetos de Aprendizajes: Características y repositorios. Abril de 2005.

[Hillman, 2003] Hillman, D.. Using Dublin Core. <http://dublincore.org/documents/usageguide>. 2003

[Html Agility Pack, 2010] Html Agility Pack 1.4.0 Binaries. Agile HTML parser that builds a read/write DOM and supports plain XPATH or XSLT. Ms-PL. <http://htmlagilitypack.codeplex.com/>. May 2010.

[IEEE, 2002] IEEE-LOM: IEEE 1484.12.1, New York: Institute of Electrical and Electronics Engineers, Learning Technology Standards Committee. Standard for Learning Object Metadata. <http://ltsc.ieee.org/wg12>. 2002.

[Lagoze et al, 2002] Lagoze Carl, Van de Sompel Herbert. The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html>. Version 2.0 of 14-06-2002.

[Merlot] Multimedia Educational Resource for Learning and Online Teaching. <http://www.merlot.org>



[**Miller, 2010**] WordNet. George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41. Christiane Fellbaum (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press. Princeton University "About WordNet." WordNet. Princeton University. <http://wordnet.princeton.edu> . 2010.

[**Nesbit et al, 2003**] John Nesbit, Karen Belfer, Tracey Leacock. Learning Object Review Instrument (LORI) User Manual E-Learning Research and Assessment Network. 2003.

[**Pressman, 2000**] Roger S. Pressman, 6ª Edición, Mc Graw Hill. Ingeniería de Software. Impreso en México. 2000.

[**Reeves, 1997**] Reeves, T.C.: Evaluating What Really Matters in Computer-Based Education. Available. <http://www.educationau.edu.au/archives/cp/reeves.htm>. 1997.

[**Rodriguez, 2009**] Ismael Rodriguez. Keywords - El Elemento Más Importante en SEO. <http://www.miempresainlinea.com/blog/post/keywords-seo.aspx>. September 28, 2009.

[**Ryan, Walmsley, 2003**] Ben Ryan, Steve Walmsley. Implementing metadata collection: a project's problems and solutions. Learning technology. [http://lutf.ieee.org/learn\\_tech/issues/january2003/index.html#3](http://lutf.ieee.org/learn_tech/issues/january2003/index.html#3). Vol. 5, no. 1, Jan. 2003.

[**TextWise, 1994**] TextWise. Semantic Signatures. Your Document digital DNA. <http://www.textwise.com> . 1994.

[**Vargo et al, 2003**] Vargo, J., Nesbit, J.C., Belfer, K., Archambault, A. Learning Object Evaluation: ComputerMediated Collaboration and Inter-Rater Reliability. International Journal of Computers and Application 25(2003) 198-205. 2003.

[**Velázquez et al, 2006a**] Velázquez A, César; Muñoz A., Jaime; Álvarez R. Francisco; Garza G., Laura. "La Determinación de la Calidad del Contenido de un Objeto de Aprendizaje". Centro de Ciencias Básicas, Universidad Autónoma de Aguascalientes, México Av. Universidad No. 940, Col. 2006.

[**Velázquez et al, 2006b**] Velázquez A, César; Muñoz A., Jaime; Álvarez R. Francisco; Pinales D. Francisco. "Estudio de los elementos de la calidad Pedagógica y de Contenido en el Metadato de LOM". Centro de Ciencias Básicas, Universidad Autónoma de Aguascalientes, México Av. Universidad No. 940, Col. 2006.

[**W3C, 2004**] W3C. XML Schema Part 2: Datatypes Second Edition. W3C Recommendation. <http://www.w3.org/TR/xmlschema-2/>. 28 October 2004.

[**Wiley, 2002**] Wiley, D. Connecting Learning Objects to Instructional Design Theory: A definition, a metaphor, and a taxonomy. In D. A. Wiley (ed.) *Instructional Use of Learning Objects*. Editorial Association for Instructional Technology, 2002.

[**Woo et al, 2004**] Woo Karen, Maree Gosper, Donna Gibbs, Tim Hand, Sharon Kerr and David Rich. User Perspectives on Learning Object Systems.  
<http://ausweb.scu.edu.au/aw04/papers/refereed/woo/paper.html>. 2004.

[**Zipf, 1949**] Zipf George Kingsley. Human Behavior and the Principle of Least-Effort. Editorial, Addison-Wesley. 1949.

## 8. Apéndice

### 8.1 Experimentaciones

#### 8.1.1 Consultas: Calificaciones promedios

En las tablas a continuación se observan las calificaciones promedios obtenidas en los primeros 20 resultados de búsquedas en cada una de las 10 búsquedas distintas que se realizaron.

Dónde:

**A:** Resultados crudos devueltos por el repositorio.

**B:** Resultados devueltos por el prototipo

Consulta	A	B	Consulta	A	B	Consulta	A	B
<b>biology marine</b>	0,4497	0,6434	<b>ecology</b>	0,1333	0,7164	<b>artificial intelligence programming</b>	0,4539	0,6847
	0,5473	0,5593		0,1333	0,6044		0,2948	0,6212
	0,4482	0,5473		0,1333	0,5946		0,5021	0,5584
	0,4297	0,5072		0,4250	0,5486		0,4983	0,5584
	0,0593	0,4976		0,1333	0,5408		0,5021	0,5584
	0,6434	0,4913		0,4297	0,5404		0,4983	0,5584
	0,4297	0,4572		0,1333	0,5331		0,5021	0,5584
	0,4281	0,4572		0,1333	0,5303		0,5021	0,5584
	0,4175	0,4572		0,1333	0,5164		0,4983	0,5584
	0,4500	0,4572		0,1333	0,5101		0,4948	0,5584
	0,4262	0,4572		0,4297	0,5082		0,3847	0,5584
	0,4262	0,4572		0,4297	0,5049		0,5035	0,5584
	0,0757	0,4572		0,6044	0,5045		0,5035	0,5584
	0,4281	0,4572		0,4297	0,5045		0,4084	0,5584
	0,0757	0,4572		0,4297	0,5045		0,5101	0,5555
	0,4175	0,4572		0,4297	0,5045		0,5101	0,5555
	0,4175	0,4572		0,4297	0,5000		0,5054	0,5555
	0,0482	0,4572		0,4297	0,4958		0,5555	0,5555
	0,4500	0,4572		0,4297	0,4937		0,5555	0,5555
	0,4227	0,4572		0,4297	0,4839		0,1245	0,5555
<b>Promedio</b>	<b>0,3745</b>	<b>0,4823</b>	<b>Promedio</b>	<b>0,3196</b>	<b>0,5320</b>	<b>Promedio</b>	<b>0,4654</b>	<b>0,5670</b>

Consulta	A	B
computer science	0,4584	0,6850
	0,5249	0,6519
	0,5408	0,6223
	0,5130	0,6135
	0,5037	0,5892
	0,4507	0,5461
	0,4727	0,5408
	0,5249	0,5302
	0,4584	0,5249
	0,5130	0,5249
	0,4727	0,5249
	0,4621	0,5249
	0,4489	0,5249
	0,4667	0,5130
	0,5249	0,5130
	0,4507	0,5118
	0,4507	0,5054
	0,4584	0,5054
0,4667	0,5054	
0,4777	0,5054	
<b>Promedio</b>	<b>0,4820</b>	<b>0,5482</b>

Consulta	A	B
cáncer disease	0,4012	0,6250
	0,4297	0,6149
	0,5249	0,6048
	0,4175	0,5723
	0,4175	0,5408
	0,4281	0,5370
	0,4281	0,5370
	0,5408	0,5317
	0,4281	0,5249
	0,1333	0,5249
	0,4951	0,5243
	0,4175	0,5183
	0,6149	0,5183
	0,5183	0,5183
	0,5183	0,4951
	0,5183	0,4572
	0,5370	0,4539
	0,5317	0,4418
0,4297	0,4363	
0,4281	0,4363	
<b>Promedio</b>	<b>0,4579</b>	<b>0,5207</b>

Consulta	A	B
history german	0,6612	0,6713
	0,3739	0,6612
	0,3437	0,6373
	0,4700	0,6016
	0,4175	0,5812
	0,4281	0,5630
	0,4700	0,5197
	0,3977	0,5183
	0,3937	0,5054
	0,6713	0,5054
	0,4700	0,5054
	0,4577	0,4938
	0,1323	0,4700
	0,4462	0,4700
	0,4175	0,4700
	0,4281	0,4700
	0,4700	0,4700
	0,4700	0,4700
0,4577	0,4700	
0,4700	0,4700	
<b>Promedio</b>	<b>0,4423</b>	<b>0,5262</b>

Consulta	A	B
modern architecture	0,3966	0,8247
	0,1333	0,5256
	0,3966	0,5101
	0,3966	0,5101
	0,4453	0,5054
	0,4885	0,5054
	0,1481	0,5054
	0,1889	0,5054
	0,4215	0,5054
	0,5256	0,5054
	0,3966	0,5054
	0,1667	0,5054
	0,1333	0,5054
	0,3728	0,4892
	0,3728	0,4885
	0,4700	0,4885
	0,4885	0,4825
	0,4453	0,4735
0,1333	0,4723	
0,4595	0,4723	
<b>Promedio</b>	<b>0,3490</b>	<b>0,5143</b>

Consulta	A	B
operative system unix	0,4059	0,7767
	0,5974	0,6767
	0,1109	0,6212
	0,1109	0,5974
	0,3967	0,5934
	0,1109	0,5934
	0,3967	0,5479
	0,1109	0,5479
	0,1722	0,5357
	0,4572	0,5192
	0,5192	0,5162
	0,4572	0,5155
	0,5934	0,5155
	0,1109	0,5155
	0,6767	0,5129
	0,4908	0,5126
	0,4568	0,5107
	0,4908	0,5102
0,4975	0,5101	
0,4908	0,5101	
<b>Promedio</b>	<b>0,3827</b>	<b>0,5569</b>

Consulta	A	B
TCP	0,1177	0,5821
	0,1202	0,5729
	0,4572	0,5102
	0,5129	0,5101
	0,5102	0,5101
	0,1424	0,5101
	0,0711	0,5092
	0,4572	0,5054
	0,4572	0,5054
	0,4572	0,4905
	0,4572	0,4883
	0,2216	0,4841
	0,1084	0,4767
	0,4572	0,4767
	0,4572	0,4767
	0,4572	0,4767
	0,4572	0,4732
	0,4572	0,4721
0,4572	0,4572	
0,4572	0,4572	
<b>Promedio</b>	<b>0,3645</b>	<b>0,4972</b>

Consulta	A	B
plates tectonics	0,4553	0,7307
	0,4303	0,7307
	0,4012	0,6968
	0,4528	0,5963
	0,4683	0,5963
	0,4059	0,5963
	0,4059	0,5963
	0,1333	0,5963
	0,4637	0,5963
	0,0579	0,5963
	0,4500	0,5963
	0,5357	0,5963
	0,5054	0,5963
	0,5054	0,5963
	0,4175	0,5963
	0,5101	0,5963
	0,5101	0,5963
	0,1333	0,5725
0,4281	0,5725	
0,5054	0,5357	
<b>Promedio</b>	<b>0,4088</b>	<b>0,6094</b>

### 8.1.2 Distancias de resultados entre perfiles

A continuación se observan los resultados de búsquedas obtenidos en los diferentes perfiles.

*Búsqueda: "java". Ranking. (Profesor1 y Estudiante1)*

Pos.	Teacher	Student
1	Introduction to OO Programming in Java - Intro to module and to OO programming	Introduction to OO Programming in Java - Intro to module and to OO programming
2	Java 2 (Java%2B2.ppt)	Object-Oriented Software Design - Introduction and Review of Java
3	Conceptos básicos de JAVA	Java 2 (Java%2B2.ppt)
4	Object-Oriented Software Design - Introduction and Review of Java	Conceptos básicos de JAVA
5	Java Básico - Características	Introduction to OO Programming in Java - Classes and arithmetic
6	Manejo de Eventos: Ejemplo No. 1 (PruebaFrame.java)	Introduction to OO Programming in Java - Creating Classes
7	Java vs C	Java Básico – Características
8	Manejo de Eventos: Ejemplo No. 1 (ManejoBoton.java)	Manejo de Eventos: Ejemplo No. 1 (PruebaFrame.java)
9	Introduction to OO Programming in Java - Classes and arithmetic	Distributed Web Systems - Java Servlets
10	HolaMundo con NetBeans Java (holamundo.pdf)	Java vs C

Distancia de Manhattan entre resultados:

$$d = |1 - 1| + |2 - 3| + |3 - 4| + |4 - 2| + |5 - 7| + |6 - 8| + |7 - 10| + |8 - \text{NE}[3]| + |9 - 5| + |10 - \text{NE}[1]| = 0 + 1 + 1 + 2 + 2 + 2 + 3 + 5 + 4 + 9 = 29$$

Búsqueda: “*complex numbers*”. Ranking (*Profesor2 y Estudiante2*).

Pos.	Teacher	Student
1	Introduction to Complex Numbers	An introduction to complex numbers
2	Visualising Complex Numbers	Visualising Complex Numbers
3	Adding and subtracting complex numbers	Adding and subtracting complex numbers
4	Multiplying complex numbers	Multiplying complex numbers
5	Multiplying complex numbers in polar form	Multiplying complex numbers in polar form
6	Dividing complex numbers	Dividing complex numbers
7	Number systems	Number systems
8	Complex numbers	The Argand Diagram and Polar Form
9	The Argand Diagram and Polar Form	Complex numbers
10	An introduction to complex numbers	Computer generated exercises - complex numbers

Distancia de Manhattan entre resultados:

$$d = |1 - \text{NE}[10]| + |2 - 2| + |3 - 3| + |4 - 4| + |5 - 5| + |6 - 6| + |7 - 7| + |8 - 9| + |9 - 8| + |10 - 1| = 9 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1 + 1 + 9 = 20$$

*En ambas búsquedas el peor caso se da a distancia:*

$$d = |1 - 10| + |2 - 9| + |3 - 8| + |4 - 7| + |5 - 6| + |6 - 5| + |7 - 4| + |8 - 3| + |9 - 2| + |10 - 1| = 9 + 7 + 5 + 3 + 1 + 1 + 3 + 5 + 7 + 9 = 50$$

### 8.1.3 Distancias y análisis de ranking de los usuarios

A continuación se observan los rankings de resultados de búsquedas realizados por los usuarios juntos con los cálculos de distancias.

Búsqueda: “java”. [Ranking]

**Usuario:** *Profesor1*

**Perfil de búsqueda asignado:** *Teacher*

Pos.	Profesor1: A su criterio	Prototipo: Perfil - Teacher	Prototipo: Perfil - Student
1	Introduction to OO Programming in Java - Intro to module and to OO programming	Introduction to OO Programming in Java - Intro to module and to OO programming	Introduction to OO Programming in Java - Intro to module and to OO programming
2	Conceptos básicos de JAVA	Java 2 (Java2.ppt)	Object-Oriented Software Design - Introduction and Review of Java
3	Java 2 (Java2.ppt)	Conceptos básicos de JAVA	Java 2 (Java2.ppt)
4	Object-Oriented Software Design - Introduction and Review of Java	Object-Oriented Software Design - Introduction and Review of Java	Conceptos básicos de JAVA
5	Java Básico - Características	Java Básico - Características	Introduction to OO Programming in Java - Classes and arithmetic
6	HolaMundo con NetBeans Java (holamundo.pdf)	Manejo de Eventos: Ejemplo No. 1 (PruebaFrame.java)	Introduction to OO Programming in Java - Creating Classes
7	Introduction to OO Programming in Java - Classes and arithmetic	Java vs C	Java Básico - Características
8	Manejo de Eventos: Ejemplo No. 1 (ManejoBoton.java)	Manejo de Eventos: Ejemplo No. 1 (ManejoBoton.java)	Manejo de Eventos: Ejemplo No. 1 (PruebaFrame.java)
9	Manejo de Eventos: Ejemplo No. 1 (PruebaFrame.java)	Introduction to OO Programming in Java - Classes and arithmetic	Distributed Web Systems - Java Servlets
10	Java vs C	HolaMundo con NetBeans Java (holamundo.pdf)	Java vs C

*Distancia de Manhattan entre resultados:*

✓ **Resultados: Profesor1: A su criterio y Prototipo: Perfil - Teacher**

$$d = |1 - 1| + |2 - 3| + |3 - 2| + |4 - 4| + |5 - 5| + |6 - 10| + |7 - 9| + |8 - 8| + |9 - 6| + |10 - 7| = 0 + 1 + 1 + 0 + 0 + 4 + 2 + 0 + 3 + 3 = \mathbf{14}$$

✓ **Resultados: Profesor1: A su criterio y Prototipo: Perfil - Student**

$$d = |1 - 1| + |2 - 4| + |3 - 3| + |4 - 2| + |5 - 7| + |6 - \text{NE}[1]| + |7 - 5| + |8 - \text{NE}[3]| + |9 - 8| + |10 - 10| = 0 + 2 + 0 + 2 + 2 + 5 + 2 + 5 + 1 + 0 = \mathbf{19}$$

*Cabe destacar que en el peor caso en distancia para 10 resultados es de 50.*

Pos.	Estudiante1: A su criterio	Prototipo: Perfil - Student	Prototipo: Perfil - Teacher
1	Introduction to OO Programming in Java - Intro to module and to OO programming	Introduction to OO Programming in Java - Intro to module and to OO programming	Introduction to OO Programming in Java - Intro to module and to OO programming
2	Object-Oriented Software Design - Introduction and Review of Java	Object-Oriented Software Design - Introduction and Review of Java	Java 2 (Java2.ppt)
3	Conceptos básicos de JAVA	Java 2 (Java2.ppt)	Conceptos básicos de JAVA
4	Java Básico - Características	Conceptos básicos de JAVA	Object-Oriented Software Design - Introduction and Review of Java
5	Introduction to OO Programming in Java - Classes and arithmetic	Introduction to OO Programming in Java - Classes and arithmetic	Java Básico - Características
6	Introduction to OO Programming in Java - Creating Classes	Introduction to OO Programming in Java - Creating Classes	Manejo de Eventos: Ejemplo No. 1 (PruebaFrame.java)
7	Java 2 (Java2.ppt)	Java Básico - Características	Java vs C
8	Manejo de Eventos: Ejemplo No. 1 (PruebaFrame.java)	Manejo de Eventos: Ejemplo No. 1 (PruebaFrame.java)	Manejo de Eventos: Ejemplo No. 1 (ManejoBoton.java)
9	Java vs C	Distributed Web Systems - Java Servlets	Introduction to OO Programming in Java - Classes and arithmetic
10	Distributed Web Systems - Java Servlets	Java vs C	HolaMundo con NetBeans Java (holamundo.pdf)

*Distancia de Manhattan entre resultados:*

✓ *Resultados:* **Estudiante1: A su criterio** y **Prototipo: Perfil - Student**

$$d = |1 - 1| + |2 - 2| + |3 - 4| + |4 - 7| + |5 - 5| + |6 - 6| + |7 - 3| + |8 - 8| + |9 - 10| + |10 - 9| = 0 + 0 + 1 + 3 + 0 + 0 + 4 + 0 + 1 + 1 = 10$$

✓ *Resultados:* **Estudiante1: A su criterio** y **Prototipo: Perfil - Teacher**

$$d = |1 - 1| + |2 - 4| + |3 - 3| + |4 - 5| + |5 - 9| + |6 - \text{NE}[1]| + |7 - 2| + |8 - 6| + |9 - 7| + |10 - \text{NE}[1]| = 0 + 2 + 0 + 1 + 4 + 5 + 5 + 2 + 2 + 9 = 30$$



Búsqueda: “complex numbers”. [Ranking]

Usuario: *Profesor2*

Perfil de búsqueda asignado: *Teacher*

Pos.	Profesor2: A su criterio	Prototipo: Perfil - Teacher	Prototipo: Perfil - Student
1	Introduction to Complex Numbers	Introduction to Complex Numbers	An introduction to complex numbers
2	Visualising Complex Numbers	Visualising Complex Numbers	Visualising Complex Numbers
3	Adding and subtracting complex numbers	Adding and subtracting complex numbers	Adding and subtracting complex numbers
4	Multiplying complex numbers	Multiplying complex numbers	Multiplying complex numbers
5	Multiplying complex numbers in polar form	Multiplying complex numbers in polar form	Multiplying complex numbers in polar form
6	Dividing complex numbers	Dividing complex numbers	Dividing complex numbers
7	An introduction to complex numbers	Number systems	Number systems
8	Complex numbers	Complex numbers	The Argand Diagram and Polar Form
9	The Argand Diagram and Polar Form	The Argand Diagram and Polar Form	Complex numbers
10	Number systems	An introduction to complex numbers	Computer generated exercises - complex numbers

*Distancia de Manhattan entre resultados:*

✓ *Resultados: Profesor2: A su criterio y Prototipo: Perfil - Teacher*

$$d = |1 - 1| + |2 - 2| + |3 - 3| + |4 - 4| + |5 - 5| + |6 - 6| + |7 - 10| + |8 - 8| + |9 - 9| + |10 - 6| = 0 + 0 + 0 + 0 + 0 + 0 + 3 + 0 + 0 + 4 = 7$$

✓ *Resultados: Profesor2: A su criterio y Prototipo: Perfil - Student*

$$d = |1 - \text{NE}[10]| + |2 - 2| + |3 - 3| + |4 - 4| + |5 - 5| + |6 - 6| + |7 - 1| + |8 - 9| + |9 - 8| + |10 - 7| = 9 + 0 + 0 + 0 + 0 + 0 + 6 + 1 + 1 + 3 = 20$$

Pos.	Estudiante2: A su criterio	Prototipo: Perfil - Student	Prototipo: Perfil - Teacher
1	An introduction to complex numbers	An introduction to complex numbers	Introduction to Complex Numbers
2	Visualising Complex Numbers	Visualising Complex Numbers	Visualising Complex Numbers
3	Adding and subtracting complex numbers	Adding and subtracting complex numbers	Adding and subtracting complex numbers
4	Multiplying complex numbers	Multiplying complex numbers	Multiplying complex numbers
5	Dividing complex number	Multiplying complex numbers in polar form	Multiplying complex numbers in polar form
6	Multiplying complex numbers in polar form	Dividing complex numbers	Dividing complex numbers
7	The Argand Diagram and Polar Form	Number systems	Number systems
8	Computer generated exercises - complex numbers	The Argand Diagram and Polar Form	Complex numbers
9	Complex numbers	Complex numbers	The Argand Diagram and Polar Form
10	Number systems	Computer generated exercises - complex numbers	An introduction to complex numbers

*Distancia de Manhattan entre resultados:*

✓ *Resultados:* **Estudiante2: A su criterio** y **Prototipo: Perfil - Student**

$$d = |1 - 1| + |2 - 2| + |3 - 3| + |4 - 4| + |5 - 6| + |6 - 5| + |7 - 8| + |8 - 10| + |9 - 9| + |10 - 7| = 0 + 0 + 0 + 0 + 1 + 1 + 2 + 0 + 3 = 6$$

✓ *Resultados:* **Estudiante2: A su criterio** y **Prototipo: Perfil - Teacher**

$$d = |1 - 10| + |2 - 2| + |3 - 3| + |4 - 4| + |5 - 6| + |6 - 5| + |7 - 9| + |8 - NE[3]| + |9 - 8| + |10 - 7| = 9 + 0 + 0 + 0 + 1 + 1 + 2 + 5 + 1 + 3 = 22$$