

LookIng4LO

Sistema Informático para la extracción automática de Objetos de Aprendizaje

Regina Motz¹, Claudia Badell², Martín Barrosa³,
Rodolfo Sum⁴
Universidad de la República
Montevideo, Uruguay
[rmotz¹ | cbadell²]@fing.edu.uy,
[marbar81³|rodosum⁴]@hotmail.com

Gabriel Díaz⁵, Manuel Castro⁶
Universidad Nacional de Educación a Distancia
Madrid, España
[gdiaz⁵ | mcastro⁶]@ieec.uned.es

Resumen

El área de e-Learning necesita imperiosamente disminuir los costos de elaboración de los materiales digitales. Una de las estrategias que se siguen es crear el material didáctico como componentes modulares y reutilizables, los llamados Objetos Digitales de Aprendizaje (ODAs). La mayor ventaja de los ODAs es su capacidad de integrarse fácilmente con otros más complejos para formar unidades didácticas o cursos completos. Actualmente existe mucho material digital disponible en diversas fuentes como la Web, documentos pdf, repositorios de ODAs, muchos de ellos con metadatos en formatos abiertos como SCORM, LOM, etc. Pero por otro lado, existe una cantidad todavía mucho mayor de documentos que podrían ser reutilizables si pudieran transformarse en ODAs. Varios trabajos previos apoyan esta transformación, pero realizándola básicamente de forma manual. En este trabajo presentamos la experiencia del uso de la herramienta LookIng4LO, Proyecto de Grado de la Facultad de Ingeniería de la Universidad de la República, para la extracción automática de ODAs, y brindamos una descripción general de la herramienta. Parte de la experiencia fue realizada para un curso de Redes de Comunicaciones Industriales de 3º curso de la carrera de Ingeniería Técnica Industrial de la UNED (Universidad Nacional de Educación a Distancia), con el apoyo del proyecto CYTED-SOLITE.

Palabras claves: e-Learning, Objetos de Aprendizaje, SCORM, Extracción y generación de Metadatos, Procesamiento de Lenguaje Natural, Anotación Semántica, Ontología.

I. INTRODUCCIÓN

Looking4LO es un prototipo desarrollado en el marco del Proyecto de Grado de la Facultad de Ingeniería, UdelaR [1], que recibe como entrada documentos no estructurados (pdf, texto, HTML y paquetes SCORM [3]) y extrae información según un área temática y un conjunto de componentes pedagógicos (definiciones, ejemplos, ejercicios, etc.), empaquetándola en Objetos Digitales de Aprendizaje (ODAs)

[2]. El área temática se define a través de una ontología y los componentes pedagógicos son modelados con reglas para definir patrones de búsqueda. Además, el Sistema genera metadatos que describen el contenido extraído y el origen de dicha información. Los ODAs generados son empaquetados utilizando el estándar SCORM.

Para que la reutilización de los ODAs pueda ser realizada con criterios pedagógicos, estos deben ser extraídos de forma que pertenezcan a tipos básicos de elementos pedagógicos, entre los cuales se encuentran las definiciones, los ejemplos, ejercicios, teoremas, demostraciones, etc. El estándar de metadatos LOM [4] proporciona, entre sus atributos, estos tipos de ODAs pero, sin embargo, encontramos que la exacta definición de estos elementos es muy dependiente del área temática del contenido del material.

Para entender el contexto de uso de la herramienta Looking4LO comenzamos dando una descripción general de la herramienta. Luego, a partir de la Sección IV, analizamos la experiencia de su uso y finalmente en la Sección V brindamos algunas conclusiones y líneas de trabajos futuros.

II. VISIÓN GENERAL DEL SISTEMA LOOKING4LO

El sistema LookIng4LO [1] recibe como entrada un conjunto de documentos no estructurados sobre los que realiza la extracción de ODAs. Se basa en una representación ontológica del área temática por la que interesa anotar la información, y en la definición de un conjunto de Componentes Pedagógicos (ejercicio, definición, ejemplo, etc.) modelados a través de reglas. El resultado obtenido son ODAs extraídos de los documentos de acuerdo a la especificación de los Componentes Pedagógicos junto con un conjunto de metadatos que son generados de forma automática.

En la Figura 1 se presenta un diagrama que contiene los participantes del proceso de generación de ODAs con metadatos.

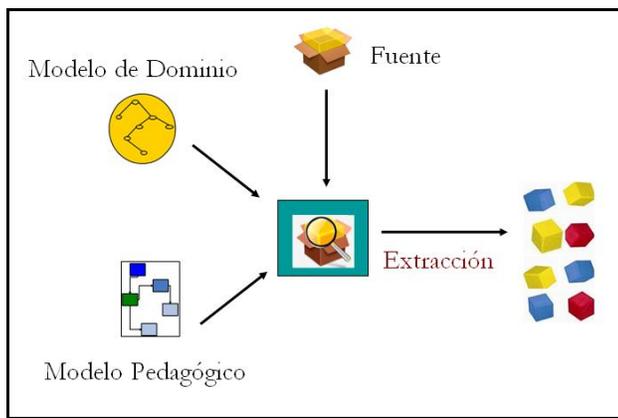


Figura 1 – Visión General del Sistema Looking4LO

Modelo Pedagógico se refiere a una abstracción que sirve para modelar un curso. En el ámbito de la Pedagogía hay involucrados conceptos más complejos que los utilizados en este proyecto. En nuestro caso, se usa para identificar una estructura formada por elementos que cumplen una función o rol dentro de un curso, y llamamos **Componentes Pedagógicos** a estos elementos. Es decir, el término **Componente Pedagógico** se refiere a cualquier material, texto en esta versión del prototipo, que cumpla una función didáctica específica, como ser la definición de un concepto, un ejercicio o problema para resolver, la demostración de un teorema, etc. Por lo tanto, los Componentes Pedagógicos que recibe el Sistema indican qué tipo de información se quiere extraer sobre un tema particular, que está determinado por otro parámetro, el dominio de interés.

Modelo de Dominio se refiere a un área temática de interés o conocimiento. Su función consiste en definir cualquier objeto o entidad que se quiera representar, y se utiliza para modelar el tema sobre el que se busca generar ODA's. Temáticas de interés pueden ser matemática, programación de computadoras, historia, cocina o cualquier otra sobre la que se quiera generar ODA's.

El *Modelo de Dominio* define qué tema se quiere buscar, y los *Componentes Pedagógicos* qué es lo que se busca sobre él. En el prototipo, los Componentes Pedagógicos son modelados mediante un conjunto de reglas, y cada dominio particular se modela a través de una ontología. Una ontología es un modelo de datos que representa un dominio y se utiliza para razonar sobre él. Según la W3C, “Una ontología define los términos a utilizar para describir y representar un área de conocimiento. Las ontologías son utilizadas por las personas, las bases de datos, y las aplicaciones que necesitan compartir un dominio de información (un dominio es simplemente un área de temática específica o un área de conocimiento, tales como medicina, fabricación de herramientas, bienes inmuebles, reparación automovilística, gestión financiera, etc.). Las ontologías incluyen definiciones de conceptos básicos del dominio, y las relaciones entre ellos, que son útiles para los ordenadores [...]. Codifican el conocimiento de un dominio y

también el conocimiento que extiende los dominios. En este sentido, hacen el conocimiento reutilizable”. [8]

Llamamos **documento** o **fuentes** a cualquier elemento digital que contenga material desde donde generar ODA's. Dado que la variedad de fuentes posibles es muy amplia, se diseñó el Sistema de forma que pueda evolucionar a nuevos formatos y estrategias de extracción.

La salida de LookIng4LO es un conjunto de ODA's, donde cada uno de ellos posee metadatos que indican, entre otros, la temática relativa a la ontología de dominio utilizada, y la correspondencia con algún componente pedagógico.

Existen muchas definiciones sobre **ODAs**. Según Wiley (2000) un ODA es “cualquier recurso digital que pueda ser reutilizado como soporte para el aprendizaje” y también lo define como “material educativo diseñado y creado en pequeñas unidades con el propósito de maximizar el número de situaciones educativas en las que pueda ser reutilizado”. [6, 7]

Las principales características de un ODA son que se trata de un objeto digital, que tiene un propósito educativo y es auto-contenido y reutilizable. En el prototipo, un **ODA** es modelado como un elemento que contiene texto, más una estructura (árbol n-ario) de metadatos que lo describe. Cada elemento de esta estructura de metadatos, tiene un nombre, valor y un conjunto de elementos hijos del mismo tipo. Esta estructura permite manejar metadatos definidos en formato LOM y extensiones realizadas sobre esta.

En el prototipo, un ODA tiene cuatro conjuntos de metadatos que clasifican esta información de acuerdo al origen desde donde es obtenida:

- **Fuente:** metadatos disponibles a nivel de cada fuente o recurso. Se refiere a la información asociada al archivo, como autor, fecha de creación, etc. En el caso de un paquete SCORM, también se refiere a los metadatos disponibles en el archivo *manifest*; entre estos, se distingue tres niveles: globales a todo el paquete, asociados a los recursos y los que aplican específicamente a un archivo contenido en un recurso.
- **Generales:** son generados automáticamente por el Sistema y contienen información sobre el contenido del documento, como el idioma.
- **Específicos:** generados automáticamente y son específicos a un tipo de componente pedagógico. Pueden existir diferentes tipos de *metadatos específicos* para cada tipo de componente pedagógicos (ejercicios, ejemplos, definiciones, etc.). Por ejemplo, *nivel de interactividad* puede ser aplicado a un ejercicio pero no a una definición, *tiempo de lectura* puede ser relevante para una definición o ejemplo, pero tal vez no para un ejercicio.
- **Externos:** se añaden en forma manual por el usuario del Sistema. Se asocian a todos los ODA's generados durante

la ejecución, y para estos, se debe proporcionar su nombre y valor.

Esta clasificación de los metadatos permite mantener en los ODAs toda la información disponible al momento de realizar la extracción, en contrapartida con integrarla utilizando alguna estrategia para resolver automáticamente los conflictos encontrados. Sin embargo, hay una excepción con los metadatos **Fuente**; cuando la fuente es un paquete SCORM se obtienen los mismos integrando los tres niveles de metadatos que este puede contener (generales del SCORM, a nivel de recursos, a nivel de cada recurso). El algoritmo asigna mayor prioridad a los metadatos más “ceranos” al recurso, complementando estos con el nivel superior. Es decir, en caso de conflicto, se mantiene el valor del nivel más bajo, y donde no exista un metadato se lo toma del nivel inmediato superior si está disponible en él. Por lo tanto, se realiza una integración desde lo más general a lo más específico, manteniendo lo específico en caso de conflicto.

En la Figura 2 se muestra un diagrama con la clasificación de las cuatro agrupaciones de metadatos definidos previamente cuando la **Fuente** es un paquete SCORM.

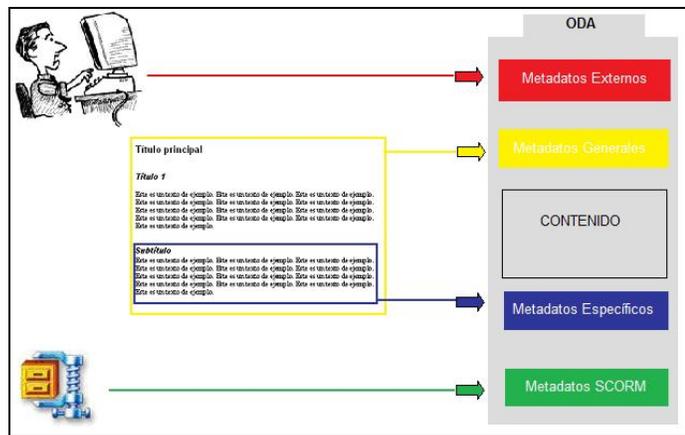


Figura 2 – Clasificación de los metadatos de un ODA

En el centro de la Figura 2, se representa el texto de un documento fuente que pertenece al paquete SCORM, y a su derecha un “ODA” que contiene un segmento de dicho texto (recuadro negro), que se corresponde a uno de los componentes pedagógicos buscados.

El prototipo soporta metadatos multivaluados. Un metadato es multivaluado cuando para un mismo metadato se puede tener más de un valor posible. Por ejemplo, el metadato *autor* es multivaluado ya que un documento puede tener más de un autor.

III. COMPONENTES PEDAGÓGICOS Y METADATOS EXTRAÍBLES

En la Tabla 1 y la Tabla 2, se indican los Componentes Pedagógicos, Metadatos Específicos y Metadatos Generales

implementados. El prototipo extrae un tipo de metadato específico para cada uno de los componentes pedagógicos implementados, y un tipo de metadato general, que aplica a todos. La elección de estos metadatos fue arbitraria, ya que se buscó mostrar la utilidad de la clasificación de metadatos específicos y generales, e implementar su extracción para demostrar su factibilidad. Extender y/o incorporar nuevos metadatos generales/específicos implica implementar nuevas reglas.

Componente Pedagógico	Metadatos Específicos
Definición	tiempo de lectura
Ejemplo	tiene imagen
Ejercicio	nivel de interactividad

Tabla 1 – Componentes Pedagógicos y Metadatos Específicos extraíbles

Metadatos Generales
Autor

Tabla 2 – Metadatos Generales extraíbles

El metadato específico *tiempo de lectura* aplica al componente pedagógico *definición*, y es una estimación del tiempo que requiere leer el contenido del ODA. Se calcula contabilizando la cantidad de palabras del contenido del ODA, dividido por una constante. La constante por la que se divide es 200. La elección de este valor se basa en el análisis realizado en el Proyecto de Grado ODA Asistente Pedagógico [9], del año 2007 del Instituto de Computación de la Facultad de Ingeniería, UdelaR, Montevideo, Uruguay.

El metadato específico *nivel de interactividad* aplica al componente pedagógico *ejercicio*, y asocia un valor entero al ODA. Este valor depende de si el ejercicio debe enviarse por email, a un foro, news o no se requiere ninguna de estas actividades para su resolución. El valor asignado según el nivel de interactividad se basó en el análisis realizado en proyecto ODA Asistente Pedagógico [9]. En la Tabla 3 se muestra el valor asociado al nivel de interactividad según el medio de contacto. En caso de no haber un medio de contacto en el contenido del ejercicio, el valor asignado es cero.

Medio de contacto	Valor de interactividad (1-10)
Email	5
News	8
Foro	9

Tabla 3 – Nivel de Interactividad

El metadato específico *tiene imagen* aplica al componente pedagógico *ejemplo*, y asocia un valor booleano, que es verdadero en caso que el ODA contenga una imagen o figura como parte de su contenido, falso en caso contrario. Este último metadato se implementa en forma parcial ya que el prototipo solo extrae texto, pero se identifica si dentro del texto original se encuentra una imagen.

Autor es un metadato general, es decir, se busca a nivel de todo el documento y no solo en el contenido de un tipo de

componente pedagógico particular. Estos metadatos aplican a todos los ODAs que se extraen. Cuando se identifica el ó los autores de un documento, se extrae también el correo electrónico y página web de cada autor en caso de que esta información esté disponible junto al nombre del autor.

IV. CASO DE ESTUDIO

En esta sección se muestran los resultados obtenidos de la evaluación sobre el prototipo. Se comenzará por presentar la configuración del Sistema al realizar las pruebas, luego se muestran los resultados de la evaluación, y finalmente se incluye las conclusiones alcanzadas en base a la información obtenida.

A. Ontología del Dominio

Como se mencionó previamente, se debe proveer al Sistema con una ontología que modele el área temática sobre la que se va a extraer información. Se construyó una ontología sencilla sobre Redes de Comunicaciones que abarca los principales conceptos del tema.

En la Figura 3 se muestra el diagrama que representa la ontología de prueba sobre Redes de Comunicaciones.

B. Fuentes

Se seleccionó un conjunto de archivos de prueba de diferentes formatos, que tratan sobre Redes de Comunicaciones. Uno de estos archivos es un Trabajo de doctorado de la UNED [5] que introduce el tema de Redes de Comunicaciones, por lo que contiene “material de buena calidad” para la extracción, teniendo en cuenta redacción y contenido. Además se incluyeron ejercicios, exámenes y otros documentos, obtenidos de Internet, abarcando los diferentes formatos que soporta el prototipo.

Previo a la ejecución de las pruebas, se analizó manualmente la muestra de documentos para determinar en cada uno, donde se encuentra la información que se corresponde a una definición, ejercicio o ejemplo. Además, se identificó cuales de los metadatos que el prototipo puede extraer, se encontraban en cada documento.

C. Extracción de Componentes Pedagógicos

Como se mencionó en la Sección II, el prototipo identifica los segmentos de texto donde se trata información relevante para su extracción. El conjunto de componente pedagógicos incluidos en la prueba, así como los metadatos que se deben generar, están descritos en la Sección III.

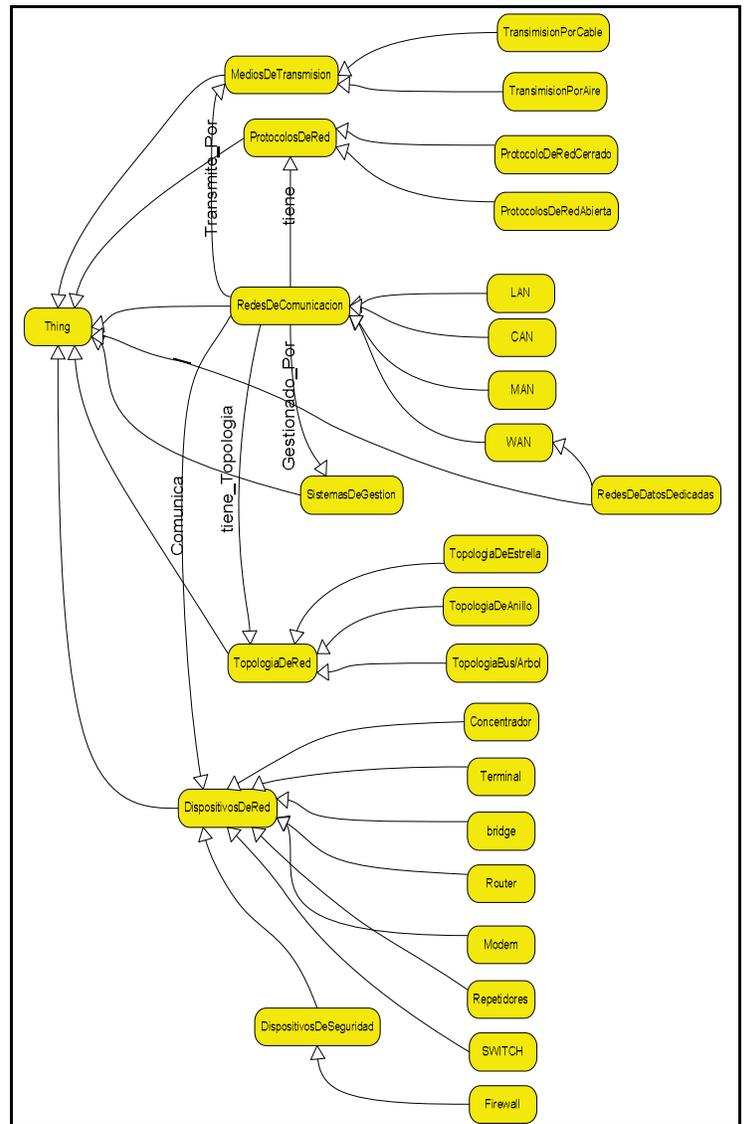


Figura 3– Ontología de prueba: Redes de Comunicaciones

Los resultados se presentan en tablas, que incluyen los valores esperados y obtenidos para cada archivo de entrada. Se utilizó una tabla para cada tipo de componente pedagógico. Además, se indica la cantidad de falsos positivos, así como la razón por la que no se detectaron algunos componentes pedagógicos. Los errores por no detección de *componentes pedagógicos* se clasifican por dos razones posibles: el concepto buscado no se encontraba en la ontología, o no se definieron reglas que permitieran detectar el componente pedagógico. Los falsos positivos corresponden a texto que tiene el mismo patrón que los diseñados para detectar componentes, pero que en realidad no se trata sobre los componentes buscados.

En la Tabla 4 se presentan los resultados correspondientes al Componente Pedagógico *definición*. En la columna Fuente se listan los nombres de los archivos de entrada; a continuación el formato del archivo y luego los resultados esperados y obtenidos.

		Componente Pedagógico: Definición				
Fuente	Formato	Cantidad Esperada	Cantidad Detectada	Cantidad No detectada		Falso Positivo
				Falta concepto Ontología	No clasifica por reglas	
dispositivoRed	TXT	6	5	0	1	0
tiposRed	TXT	5	5	0	0	0
Redes	HTM	17	8	4	5	0
protocol1	HTM	9	3	5	1	1
eirc0602	PDF	0	0	0	0	0
eirc0603	PDF	1	0	0	1	0
eirc0802	PDF	0	0	0	0	0
eirc0803	PDF	0	0	0	0	0
trabajo-Doctorado	PDF	45	11	12	22	6
TOTALES		83	32	21	30	7

Tabla 4 – Resultados en la extracción de Definiciones

La columna *Cantidad Esperada* indica cuantas definiciones se encontraron de forma manual en el documento. *Cantidad Detectada* indica el número de ODAs efectivamente generados, y la suma con los valores debajo de las columnas *Cantidad no Detectada* coincide con la primera. En los casos en que el Sistema genera un ODA que no corresponde (falso positivo), este valor se declara en la última columna.

Las tablas para *ejemplos* y *ejercicios* contienen la misma información. A continuación, en la Tabla 5, se presentan los resultados en la extracción de *Ejemplos*.

		Componente Pedagógico: Ejemplo				
Fuente	Formato	Cantidad Esperada	Cantidad Detectada	Cantidad No detectada		Falso Positivo
				Falta concepto Ontología	No clasifica por reglas	
dispositivoRed	TXT	0	0	0	0	0
tiposRed	TXT	0	0	0	0	0
Redes	HTM	2	0	0	2	0
protocol1	HTM	0	0	0	0	0
eirc0602	PDF	0	0	0	0	0
eirc0603	PDF	0	0	0	0	0
eirc0802	PDF	0	0	0	0	0
eirc0803	PDF	0	0	0	0	0
trabajo-Doctorado	PDF	3	2	0	1	0
TOTALES		5	2	0	3	0

Tabla 5 – Resultados en la extracción de Ejemplos

En la Tabla 6 se presentan los resultados en la extracción de *Ejercicios*.

		Componente Pedagógico: Ejercicio				
Fuente	Formato	Cantidad Esperada	Cantidad Detectada	Cantidad No detectada		Falso Positivo
				Falta concepto Ontología	No clasifica por reglas	
dispositivoRed	TXT	0	0	0	0	0
tiposRed	TXT	0	0	0	0	0
Redes	HTM	0	0	0	0	0
protocol1	HTM	0	0	0	0	0
eirc0602	PDF	3	3	0	0	0
eirc0603	PDF	3	3	0	0	0
eirc0802	PDF	7	7	0	0	0
eirc0803	PDF	7	7	0	0	0
trabajo-Doctorado	PDF	0	0	0	0	0
TOTALES		20	20	0	0	0

Tabla 6 – Resultados en la extracción de Ejercicios

D. Extracción de Metadatos Generales y Específicos

A continuación, se presentan los resultados en la generación de metadatos que forman parte de los ODAs. Esta información está separada de acuerdo a la clasificación y alcance de los metadatos. Por un lado, autor, como representante de los Metadatos Generales que aplican a todos los ODAs extraídos de un mismo documento; y por otro, los Metadatos Específicos que están asociados a un tipo de componente pedagógico. El alcance de estos últimos es local a cada ODA. Como metadatos específicos se tiene, *tiempo de lectura* para las *definiciones*, *tiene imagen* para *ejemplos* y *nivel de interactividad* para los *ejercicios*.

La extracción de la información sobre el/los autores se hace sobre todo el contenido del documento. En la muestra utilizada para las pruebas, la mayoría de los documentos no incluía el autor como parte de su contenido. Esto se representa con un 0 en la columna *Esperados*. En los casos en que se cuenta con el nombre del/los autores del documento, el valor esperado es el número de ODAs que contienen la información de sus autores (de todos los tipos de ODAs). La columna *Encontrados* indica cuantos de estos ODA contienen el metadato con la información correcta. En la Tabla 7 se presentan los resultados para la extracción del metadato general autor.

		Metadatos Generales	
Fuente	Formato	Autor	
		Esperados	Encontrados
dispositivoRed	TXT	0	0
tiposRed	TXT	0	0
Redes	HTM	0	0
protocol1	HTM	0	0
eirc0602	PDF	1	1
eirc0603	PDF	0	0
eirc0802	PDF	0	0
eirc0803	PDF	0	0
trabajo-Doctorado	PDF	0	0
TOTALES		1	1

Tabla 7 – Resultados extracción de Autor

Nótese que el foco en la selección de los archivos de la muestra está orientado hacia la variedad de formatos de archivo y componentes pedagógicos, y no tanto, para evaluar la capacidad de generación de metadatos. De todas maneras, se incluyen varios casos que permiten dar una primera evaluación sobre esta capacidad.

Para los metadatos específicos, la información que se registró es diferente, ya que cada metadato contiene información que es “local” a una ODA, y a su vez, cada tipo de componente pedagógico puede tener sus metadatos específicos particulares.

En la Tabla 8 se presentan los resultados obtenidos en la generación de metadatos específicos.

Fuente	Formato	Metadatos Específicos					
		Definición		Ejemplo		Ejercicio	
		Esperados	Encontrados	Esperados	Encontrados	Esperados	Encontrados
dispositivoRed	TXT	5	5	0	0	0	0
tiposRed	TXT	5	5	0	0	0	0
Redes	HTM	8	8	0	0	0	0
protocol1	HTM	3	3	0	0	0	0
eirc0602	PDF	0	0	0	0	3	1
eirc0603	PDF	0	0	0	0	3	3
eirc0802	PDF	0	0	0	0	7	5
eirc0803	PDF	0	0	0	0	7	6
trabajo-Doctorado	PDF	11	7	2	2	0	0
TOTALES		32	28	2	2	20	15

Tabla 8 – Resultados en la generación de Metadatos Específicos

El metadato *tiempo de lectura*, tiene dos columnas, la primera muestra el número de metadatos que fueron extraídos con valor correcto; y la segunda, indica la cantidad de incorrectos. La suma de ellos, coincide con el total de ODAs extraídos del tipo *definición*.

Para el metadato *ejemplo*, en la primer columna se indica cuantos ODAs de éste tipo contienen imágenes, y en la segunda, cuantos tienen el metadato con la información correcta.

El metadato *nivel de interactividad* de un ejercicio depende de si éste debe ser enviado por mail, a un foro, news o si el alumno simplemente debe realizarlo como práctica. A todos los ODAs del tipo *ejercicio*, se les asocia un metadato con su nivel de interactividad. La columna *Esperados* indica la cantidad de ODAs de tipo *ejercicio* que han sido extraídos, y la columna *Encontrados* indica el número de estos ODAs que tienen el metadato con el valor correcto.

E. Paquetes SCORM

En la evaluación del prototipo se utilizaron paquetes SCORM para verificar el manejo de los metadatos incluidos en su estructura y, por lo tanto, del manejo de los metadatos Fuente. En estos paquetes, se incluyeron documentos tal que su contenido clasificara por distintas reglas especificadas para cada componente pedagógico. La finalidad de esta elección y prueba, es corroborar que se:

- manipula correctamente la estructura de los paquetes
- realiza una correcta extracción de metadatos a partir del archivo manifiesto
- obtiene los mismos ODAs que si se extrajera la información a partir de los recursos del SCORM de forma independiente.

El algoritmo de integración de estos metadatos, asocia a todos los ODAs extraídos del paquete los metadatos del nivel 1. Si en los siguientes niveles (2 y 3) se encuentra nuevamente algunos de estos metadatos, se sobrescribe el valor del metadato del nivel 1 o se incluye también el nuevo para los que son multivaluados. Esto es igual para el nivel 2 y 3. Si

alguno de los niveles “inferiores” contiene nuevos metadatos, estos son agregados a los que se vienen acumulando.

V. CONCLUSIONES Y TRABAJOS FUTUROS

A. Conclusiones

Para la evaluación de LooKIng4LO se seleccionó un conjunto de archivos con formato html, pdf, texto y SCORM, que tratan sobre Redes de Computadores. Se definió una ontología de prueba sobre la misma temática y se ejecutaron las pruebas incluyendo reglas para extraer Componentes Pedagógicos de los tipos definición, ejemplo y ejercicio, y metadatos autor (general), tiempo de lectura, imagen y nivel de interactividad (específicos).

A partir de los resultados, se demostró la capacidad de procesamiento de los diferentes tipos de formato de documento.

Con respecto a los paquetes SCORM se logró extraer todos los metadatos incluidos en su estructura (imsmanifest), y su integración fue correcta, de acuerdo al algoritmo diseñado para esto. La descompresión de los archivos del paquete se realizó sin problemas y la extracción de ODAs desde sus recursos produjo los mismos resultados para componentes pedagógicos y metadatos que si se hubieran procesado de forma independiente.

El conjunto de reglas utilizadas, está diseñado para modelar oraciones y texto con sintaxis “correcta”. Uno de los casos en que se detectó un error en la delimitación, se debe a que la oración previa al fragmento donde se detecta un *componente pedagógico* no contiene un punto (delimitación de la oración). En este caso, la oración anterior también resulta incluida en el contenido.

Como se mencionó, los errores por no detección de información se clasificaron como omisión en la ontología o error en las reglas. En el caso de la ontología, se debe evaluar qué conceptos agregar y no caer en la tentación de incluir términos de forma indiscriminada, que no sean específicos del área modelada, para no obtener correspondencias incorrectas.

La mayor cantidad de errores están asociados a las reglas. En cada caso, esto se debe a la falta de una regla que lo contemple o que una regla no es lo suficientemente robusta y obtiene falsos positivos. La solución a estos errores en algunos casos es directa y simple; en otros puede involucrar más procesamiento para obtener mayor información sobre el contexto permitiendo así tomar “mejores decisiones”. Hay otros casos, en los que la forma de introducir conceptos tiene un grado de sutileza que hace difícil su modelado con reglas. El alcance del análisis realizado en el proyecto no permitió profundizar en estos aspectos.

De todas maneras, se puede afirmar que el rendimiento del prototipo está muy por debajo de la capacidad que se puede

lograr. Las limitaciones que presenta, al momento de esta evaluación, se deben a que su función es la de probar la factibilidad de la solución. Es posible mejorar su configuración y, hasta este momento, no se encontraron restricciones técnicas o tecnológicas que impidan hacerlo.

B. Trabajos futuros - Soporte a otros formatos de fuentes

El Sistema se diseñó de forma extensible. Actualmente se quiere extender para otro tipo de formatos de fuentes y, en ese sentido, se quiere aprovechar la experiencia en repositorios de temas electrónicos del grupo de la UNED y, en particular, de su colaboración con la OCW.

OCW (*Open CourseWare*) [10] es una iniciativa editorial electrónica a gran escala, basada en Internet y fundada conjuntamente por la Fundación William and Flora Hewlett, la Fundación Andrew W. Mellon y el Instituto Tecnológico de Massachusetts (MIT). Entre sus objetivos destacan el intento de proporcionar un acceso libre, sencillo y coherente a los materiales de los cursos del MIT para educadores del sector público, estudiantes y autodidactas de todo el mundo y crear un modelo eficiente, basado en estándares, que otras universidades puedan emular a la hora de publicar sus propios materiales pedagógicos.

En este último sentido la UNED lleva ya varios años colaborando con su propio repositorio electrónico OCW [11, 12], en colaboración con el portal UNIVERSIA. Los materiales depositados en cualquier sitio OCW son de diferentes tipos: planificación de cursos (programas, temarios, objetivos pedagógicos, calendarios, etc.), contenidos (bibliografía, documentos, material audiovisual, material auxiliar, etc.) y distintas actividades pedagógicas (ejercicios, tests, proyectos, prácticas de laboratorio, etc.). Son, por lo tanto, otra buena fuente de datos de prueba para la herramienta, teniendo en cuenta, además, que pasan una serie de filtros de calidad previos en la Universidad.

VI. AGRADECIMIENTOS

Los autores agradecen al Programa Iberoamericano de Ciencia y Tecnología para el desarrollo (CYTED) su soporte para este trabajo mediante el proyecto CYTED-508AC0341 "SOLITE-SOFTWARE LIBRE EN TELEFORMACIÓN" y también al Ministerio Español de Ciencia e Innovación su apoyo mediante el proyecto RedOBER - Proyecto TSI2007-31091-E Objetos Educativos Reutilizables (para el EEES en las especialidades de las Tecnologías de la Información y las Comunicaciones), como también al programa Latin American and Caribbean Collaborative ICT Research (LACCIR)-Proyecto JARDIN.

VII. REFERENCIAS

- [1] Claudia Badell, Martín Barrosa, Rodolfo Sum, Extracción automática de objetos de aprendizaje y sus metadatos
Proyecto de Grado, Instituto de Computación, Facultad de Ingeniería, UDELAR, Montevideo Uruguay (2008)
- [2] APROA Comunidad – FAQ: Sobre Objetos de Aprendizaje
URL: http://146.83.43.182/aproa/1116/article-68380.html#h2_1
[última visita, noviembre 2008].
- [3] SCORM-Shareable Content Object Reference Model
URL: <http://adlnet.gov/scorm/index.aspx>
[última visita, noviembre 2008]
- [4] LOM Learning Object Metadata
URL: <http://ltsc.ieee.org/wg12/>
[última visita, noviembre 2008]
- [5] UNED – Universidad Nacional de Educación a Distancia, Departamento de Ingeniería Eléctrica Electrónica y de Control
URL: <http://www.ieec.uned.es/Asignaturas/Guias/623065.pdf>
[última visita, noviembre 2008]
- [6] Baltasar Fernández Manjón, Pablo Moreno Ger, José Luis Sierra Rodríguez, Iván Martínez Ortiz.
Uso de estándares aplicados a TIC en educación.
URL: <http://ares.cnice.mec.es/informes/16/index.htm>
[última visita, noviembre 2008]
- [7] Wiley D. A..
Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D. A. Wiley (Ed.), *The Instructional Use of Learning Objects: Online Version*, 2000
URL: <http://reusability.org/read/chapters/wiley.doc>
[última visita, diciembre 2008]
- [8] W3C ontología.
URL: <http://www.w3c.es/Traducciones/es/SW/2005/owlfaq>
[última visita, noviembre 2008]
- [9] Natalia De Rogatis, Nicolás Millot, Javier Oliva, ODA Asistente Pedagógico.
Proyecto de Grado, Instituto de Computación, Facultad de Ingeniería, UDELAR(2006)
- [10] Open CourseWare del MIT
URL: <http://ocw.mit.edu/OcwWeb/web/about/about/index.htm>
[última visita, enero 2009]
- [11] Open CourseWare de la UNED-Universia
URL: <http://ocw.innova.uned.es/ocwuniversia/>
[última visita, enero 2009]
- [12] Curso de Redes de Comunicaciones Industriales en la OCW-UNED
URL: <http://ocw.innova.uned.es/ocwuniversia/ingenieria-industrial/redes-de-comunicaciones-industriales>
[última visita, enero 2009]