

Aplicación de técnicas de Ingeniería Lingüística en sistemas de e-learning basados en objetos de aprendizaje¹

José R. Hilera, Luis Bengochea, Ricardo Sánchez de Madariaga,
José A. Gutiérrez de Mesa, José J. Martínez

Universidad de Alcalá, Departamento de Ciencias de la Computación
28871 Alcalá de Henares, España
{jose.hilera, luis.bengochea, ricardo.sanchez,
jantonio.gutierrez, josej.martinez}@uah.es

Resumen. Se presentan tres posibles formas de aplicar técnicas derivadas del tratamiento de la información textual al ámbito de los sistemas de e-learning basados en objetos de aprendizaje (*learning objects*) reutilizables: la generación automática de metadatos (LOM, IMS-MD, SCORM, Dublin Core) a partir de recursos didácticos, la generación automática de cuestionarios de evaluación (IMS-QTI), y la construcción de buscadores lingüísticos de objetos didácticos en repositorios normalizados (IMS-DRI) y en repositorios semánticos basados en ontologías.

1 Introducción

La Ingeniería Lingüística tiene como principal objetivo proporcionar medios para ampliar y mejorar la utilización de la Lengua, haciendo de ella una herramienta más eficaz. Desde el punto de vista tecnológico, la Ingeniería Lingüística ayuda a mejorar la utilización de la Lengua en los sistemas informáticos, asimilando, analizando, seleccionando y presentando la información con el objetivo de que las máquinas lleguen a “entender” el lenguaje natural y, de esta forma, se satisfagan las necesidades de información de los usuarios con mayor precisión y se contribuya a superar el problema de exceso de información. Teniendo en cuenta que tales usuarios pueden ser los estudiantes que reciben formación a través de sistemas informáticos de e-learning, parece evidente que entre ambos campos se pueden establecer diferentes interrelaciones, algunas de las cuales se presentan en este trabajo.

El resto del artículo está estructurado de la siguiente forma: en primer lugar se realiza una introducción a los fundamentos de la ingeniería lingüística para, a continuación dedicar tres apartados a diferentes propuestas de aplicación de las técnicas de esta ingeniería. Las propuestas están relacionadas con la posibilidad de generar de forma automática de los metadatos establecidos por especificaciones y estándares de

¹ Este trabajo está soportado por el Ministerio de Industria a través del proyecto FIT-350101-2004-7: “Plataforma para la gestión y explotación de recursos educativos virtuales”, del Programa Nacional de Tecnologías de Servicios de la Sociedad de la Información.

facto, como IEEE-LOM, IMS-MD o SCORM, para los objetos de aprendizaje (*learning objects*); con la generación automática de exámenes con formato estándar (IMS-QTI) a partir de los contenidos de objetos de aprendizaje; y con el desarrollo de sistemas avanzados de búsqueda y anotación de objetos en repositorios semánticos con arquitectura IMS-DRI.

2 Fundamentos de la Ingeniería Lingüística

La secuencia de procesos básicos de la Ingeniería Lingüística consiste fundamentalmente en la introducción de material en un sistema informático proveniente de diferentes fuentes (habla, texto impreso, etc.); el reconocimiento de la lengua en la que está elaborado el material, con distinción de las palabras independientes; la “comprensión” del significado del material en el nivel apropiado a cada aplicación específica; la utilización de la comprensión en aplicaciones de transformación, búsqueda de información o traducción; la generación del medio para presentar los resultados de la aplicación; y, finalmente, la presentación de los resultados a los usuarios humanos mediante un dispositivo periférico adecuado: pantalla, altavoz, impresora, etc. Una parte importante de estos procesos se basa en la comprensión del lenguaje natural, para lo cual se utilizan lematizadores, analizadores morfológicos, analizadores sintácticos y modelos semánticos.

Los *lematizadores* son programas que implementan algoritmos (en inglés, *stemming algorithms*) para extraer el lexema o raíz de las palabras, considerando como un solo término palabras que literalmente son diferentes pero que poseen una raíz común. Esto es muy útil en los Sistemas de Recuperación de Información, para conseguir que palabras con significado similar se reduzcan a los mismos lexemas. Los algoritmos lematizadores se dividen en dos categorías: los algoritmos de supresión de sufijos, que se basan en reglas para suprimir sufijos frecuentes en un determinado idioma; y los algoritmos de variedad de letras sucesoras, que se basan en árboles de letras sucesoras de las palabras, sobre los que se calcula la entropía o variedad máxima donde se separa el lexema [1, 2, 3].

Los *analizadores morfológicos* se utilizan para describir la información de las frases a partir de los elementos básicos de la lengua: morfemas, palabras y partes de la oración, señalando los grupos de elementos que funcionan como un todo, lo que permite identificar las entidades del texto. Los *analizadores sintácticos* agrupan los constituyentes de las frases tomando como entrada el resultado del proceso de análisis morfológico. Permiten extraer componentes más grandes que las palabras de un corpus textual e identificar sintagmas que se pueden agrupar posteriormente en oraciones. Finalmente, los *modelos semánticos* se utilizan para representar el significado del lenguaje en términos de conceptos y de relaciones entre ellos, utilizando la capacidad de abstracción, el razonamiento lógico y la capacidad de organización y estructuración de los datos.

Actualmente se desarrollan líneas de investigación en este campo centradas en la incorporación de los avances en Ingeniería Lingüística, especialmente, al sector de las comunicaciones, de la edición y la enseñanza. En este último caso, se están llevando a cabo proyectos para la generación automática de contenidos docentes utilizando téc-

nicas lingüísticas [4]. Otras líneas de investigación pretenden la incorporación de técnicas de Ingeniería Lingüística al ámbito de los Sistemas de Recuperación de Información, especialmente en Internet, para reducir al mínimo el “nivel de ruido” en las consultas, evitando documentos sin información relevante; y para conseguir que los documentos aparezcan en un orden de prioridad adecuado [5, 6, 7, 8].

Los autores de este artículo estamos trabajando actualmente en la aplicación de técnicas estándar de normalización de textos para conseguir una mayor precisión de los sistemas de recuperación de información, mediante la extracción automática de términos y sintagmas que caracterizan semánticamente a un documento [9, 10]. Se está construyendo un sistema de extracción con el que evaluar distintos algoritmos de asignación de relevancia, como punto de partida para la creación de un sistema de generación automática de cuestionarios de autoevaluación para entornos de enseñanza basados en Internet (e-learning), como evolución de un sistema básico desarrollado recientemente [11].

3 Generación automática de metadatos de objetos de aprendizaje

Para facilitar su reutilización, los objetos de aprendizaje deben ofrecer a sus potenciales usuarios suficiente información sobre su origen, tipo de contenido, complejidad, etc., como para que éstos puedan decidir si dichos objetos son candidatos en un proceso de selección de contenidos para una actividad formativa basada en e-learning. Esta información se conoce como *metadatos* o *descriptores* del objeto de aprendizaje. Existen diferentes propuestas para normalizar los metadatos necesarios para describir un objeto de aprendizaje, como la del IMS Global Learning Consortium [12], o la de ADL a través de su modelo de referencia SCORM [13]; ambas caracterizadas por tomar como referencia el estándar elaborado por la asociación IEEE con el nombre de LOM (*Learning Object Metadata*) [14], proponiendo la utilización de un subconjunto de los 64 metadatos definidos por este estándar, englobados en nueve categorías. Algunos de estos metadatos son: título, idioma, descripción, palabras clave, tamaño, duración, edad recomendada, nivel de dificultad, etc.

El principal inconveniente que existe en la práctica respecto a los metadatos es que es difícil encontrar objetos de aprendizaje, por ejemplo publicados en los repositorios existentes en la Web², que tengan cumplimentados ni siquiera un mínima parte de sus metadatos asociados. Lo habitual es que los únicos datos disponibles sean el título, nombre del autor y la fecha de creación. Los autores no suelen disponer del tiempo necesario para cumplimentar los aproximadamente 60 metadatos restantes que, según LOM, podrían utilizarse para describir el objeto de aprendizaje, a pesar de disponer de editores software de libre uso para ello, como *Reload Editor* (www.reload.ac.uk) o *LRN* (www.microsoft.com/elearn).

Una forma de ayudar a los autores a cumplimentar el mayor número de metadatos, sería ofreciéndoles un contenido por defecto para los mismos, de tal forma que el

² En la siguiente dirección existen enlaces a los principales repositorios:

www.eun.org/ww/en/pub/celebrate_help/useful_links/projects__learning_object_rep.htm

autor solo debería confirmarlo o modificarlo en parte. Aquí es donde la Ingeniería Lingüística puede jugar un importante papel, ya que puede ofrecer un contenido inicial para metadatos que habitualmente no se rellenan, como la descripción del recurso, las palabras clave, etc. Se podrían aplicar técnicas de Ingeniería Lingüística para, a partir de los contenidos educativos de un objeto de aprendizaje, ser capaces de generar, por ejemplo, la descripción del objeto, mediante un proceso de elaboración de resumen automático. El problema que se plantea en estos casos es que un objeto de aprendizaje no consta de un único documento que analizar, sino que puede estar formado por múltiples recursos textuales y multimedia, por lo que se plantea la dificultad añadida de tener que manejar múltiples formatos de archivo.

En este departamento de la Universidad de Alcalá ya se ha desarrollado un software de Ingeniería Lingüística para la extracción de términos clave en documentos en formato Word (.doc) [15], que se está adaptando para generar de forma automática el valor del metadato *Keyword* asociado a objetos de aprendizaje, a partir de los contenidos educativos en formato Word o HTML incluidos en paquetes SCORM con estructura .zip, y ampliando también los formatos admitidos, para reconocer texto en archivos pdf o ppt.

Dicho software permite identificar, dentro del texto completo de los objetos de aprendizaje, los sintagmas que son más relevantes acerca de su contenido. Para ello, es preciso disponer antes de un conjunto de objetos de aprendizaje pertenecientes a un corpus similar al de los objetos estudiados, cuyos sintagmas relevantes sean conocidos de antemano. A partir de dicho conjunto y mediante un proceso de entrenamiento se construye una tabla de discretización de las características asociadas a los términos considerados relevantes en los objetos del conjunto de entrenamiento.

Una vez entrenado el sistema, el proceso de identificación automática de los términos que serán considerados valores de los metadatos de un nuevo objeto de aprendizaje, consiste en lo siguiente: tras una fase de normalización del texto se obtiene una primera relación de sintagmas candidatos, desechando aquellos que no cumplen una serie de condiciones (que su longitud no esté entre un máximo y un mínimo preestablecidos, que comiencen o terminen por palabras vacías, que no alcancen una frecuencia mínima de aparición, etc.). También se somete a los candidatos a una fase de lematización con el objetivo de considerar solamente las raíces de las palabras. A continuación, se establece una baremación discreta de cada sintagma, basada en los valores de su frecuencia en el texto en relación con el conjunto de control ($TF \times IDF$), la longitud desde el comienzo del texto hasta la primera aparición del sintagma y cuántas veces es considerado como relevante entre los objetos del conjunto de control. Finalmente, un clasificador Naïve-Bayes ordenará la lista de sintagmas candidatos de acuerdo con la probabilidad total normalizada de que sean relevantes y de ahí se elegirán los n primeros, siendo n el número de valores con que se desee dotar al metadato *Keyword*.

4 Generación automática de cuestionarios de evaluación a partir de objetos de aprendizaje

La generación automática de enunciados de exámenes a partir de documentos textuales es uno de los campos de estudio de la Ingeniería lingüística. Se han desarrollado técnicas que permiten obtener, por ejemplo, cuestionarios tipo test para comprobar los conocimientos de una persona después de realizar la lectura de una documentación determinada, a partir del texto de dicha documentación. En la Universidad de Alcalá se han aplicado este tipo de técnicas en algunas asignaturas de carácter práctico, generando en tiempo real diferentes cuestionarios con diferente estructura y preguntas sobre una misma materia, para que sean realizados por los alumnos con el fin de validar sus conocimientos sobre esa materia antes de pasar al estudio de la prevista a continuación en el programa docente [11].

Habitualmente los sistemas que realizan este tipo de generación no tienen en cuenta los estándares que en la actualidad se están implantando en las plataformas de e-learning para la gestión de evaluaciones, como la especificación QTI (*Question and Test Interoperability*) del IMS Learning Consortium [16, 17]. Esta norma ha sido creada con el objetivo de ofrecer una estructura básica que describa la forma de representar evaluaciones (*assessments*) y sus calificaciones correspondientes, para que sean intercambiables entre diferentes sistemas de e-learning. Es una idea similar a la de reutilización de los objetos de aprendizaje (*learning objects*), pero en este caso aplicada a la reutilización de objetos de evaluación. La especificación QTI establece, por una parte, la estructura que deben tener los objetos de evaluación y, por otra, la que deben tener los objetos que almacenan los resultados de la evaluación.

Un sistema de generación automática de exámenes tipo test basado en técnicas de Ingeniería Lingüística debería ser capaz de generar archivos con el formato establecido por IMS, a partir del contenido de objetos de aprendizaje. Suponiendo que este artículo formara parte de un objeto de aprendizaje sobre Ingeniería Lingüística, en el listado 1 se muestra un ejemplo de examen con una sola sección y una única pregunta de respuesta múltiple que podría haber sido generada de forma automática a partir del texto del tercer párrafo del apartado segundo del artículo, aplicando un algoritmo de generación similar al implementado en el sistema descrito en [11]. El texto de la pregunta (*¿Qué son los lematizadores?*) y la respuesta correcta (*b. Programas que implementan algoritmos para extraer el lexema o raíz de las palabras*) se obtendrían de la primera frase del párrafo indicado; mientras que el resto de opciones no válidas ofrecidas al alumno (*a. Analizadores morfológicos, c. Analizadores sintácticos*) se extraerían de un contexto próximo a la frase anterior, en este caso, del siguiente párrafo en el mismo apartado.

Los autores han aplicado técnicas de Ingeniería Lingüística en la generación de cuestionarios a partir de textos técnicos [11]. Los algoritmos de generación han funcionado razonablemente bien, aunque están limitados a este tipo de textos, siendo necesaria una labor de investigación en el futuro para conseguir obtener resultados aceptables a partir de documentos no técnicos (por ejemplo, poemas).

5 Sistemas de búsqueda en repositorios de objetos de aprendizaje

El objetivo de los estándares sobre objetos de aprendizaje es facilitar la reutilización de estos objetos en diferentes plataformas de e-learning y por diferentes usuarios interesados en su contenido, bien para aprender (usuario con perfil de alumno) o para utilizarlo en la creación de un proceso de aprendizaje (usuario con perfil de profesor). Para que esta reutilización sea real, es necesario que los objetos de aprendizaje que se desarrollen se almacenen en repositorios a los que puedan acceder, a través del sistema de búsqueda adecuado, los usuarios interesados en ellos.

En este sentido, el IMS Learning Consortium ha elaborado la especificación *IMS Digital Repositories Interoperability (DRI)* [18], que incluye recomendaciones en relación con las funciones que deberían ofrecer estos repositorios, con el objetivo de facilitar la interoperabilidad entre repositorios de diferentes plataformas implementados con tecnologías de almacenamiento diferentes, y de universalizar el acceso a los mismos desde diferentes sistemas o motores de búsqueda de objetos de aprendizaje. Esta especificación también sugiere algunas de las funciones que podrían incluir los sistemas que realizan búsquedas en repositorios, como la traducción de las peticiones realizadas en un lenguaje diferente al del sistema de almacenamiento del repositorio, la conversión de datos necesaria para aprovechar todos los metadatos incluidos en cada objeto de aprendizaje almacenado, o la distribución de las consultas a todos los almacenes o repositorios gestionados por el sistema de búsqueda.

La Ingeniería Lingüística puede ser de utilidad en la implementación de este tipo de sistemas. Así, un sistema desarrollado con las técnicas apropiadas, podría permitir a un usuario realizar peticiones en *Lenguaje Natural*, encargándose el sistema, como establece IMS-DRI, de su traducción a una petición equivalente en el lenguaje interno del repositorio (por ejemplo, el lenguaje XQuery cuando se trate de un repositorio XML) que incluya las referencias necesarias a los metadatos más adecuados según la petición realizada.

Realmente, si se tiene en cuenta que los objetos de aprendizaje almacenados en un repositorio son, al fin y al cabo, recursos de información, el sistema de búsqueda anterior se podría desarrollar aplicando la experiencia acumulada durante décadas en el campo de la Lingüística y de las Ciencias de la Documentación respecto a la creación y utilización de numerosos y avanzados Sistemas de Recuperación de Información y Documentación. En este sentido, recursos lingüísticos como los vocabularios, las bases de datos léxicas o los tesauros de descriptores pueden también utilizarse para la indización de los recursos que se registran en el repositorio (en este caso, objetos de aprendizaje en lugar de documentos); y en su búsqueda posterior a partir de lenguajes de consulta que permitan, por ejemplo, extender una búsqueda de objetos sobre una determinada materia a otros sobre otras materias relacionadas con la anterior en el tesoro de descriptores utilizado en su momento como lenguaje de indización.

En este último caso pueden utilizarse técnicas como la construcción automática de tesauros [19] a partir del corpus analizado (los objetos de aprendizaje en este caso). Esta técnica se basa en la llamada *hipótesis de asociación*, que afirma que las palabras de un corpus relacionadas entre sí tienden a coincidir en ese corpus [20]. Para

evitar problemas con documentos muy largos se utiliza una ventana de texto de tamaño fijo. Para medir el grado de coincidencia entre dos palabras en el corpus se utiliza una variante de la *cantidad de información mutua esperada* (EMIM, *Expected Mutual Information Measure*) [20, 21]. Esta medida es el número de veces que las dos palabras coinciden en alguna ventana de texto del corpus, normalizada por el número de veces que las dos palabras aparecen en todo el corpus.

Cuando un usuario desea extender una búsqueda de objetos sobre una determinada materia a otros sobre materias relacionadas, puede suministrársele, con ayuda del tesauro automático desarrollado sobre todas las materias relacionadas entre sí a las que pertenece el objeto original, una serie de palabras relacionadas con las palabras representativas de este objeto. Estas palabras relacionadas son redundantes en muchos casos, con el objetivo de maximizar el *recall* de la consulta expandida. Con el fin de que este aumento de *recall* no perjudique la precisión, el propio usuario puede eliminar de la consulta expandida las palabras que considere que van a perjudicar la precisión de la nueva búsqueda. Hay que tener en cuenta que en la consulta expandida sólo figurarán palabras que están en los objetos de aprendizaje, ya que el tesauro se ha construido (automáticamente) sobre el propio corpus que se está manejando (los objetos de aprendizaje).

5.1 Ontologías para repositorios de objetos de aprendizaje

Además de recursos lingüísticos tradicionales, como los citados, y al igual que está ocurriendo con los sistemas de recuperación de información más avanzados, también en el caso de los repositorios de objetos de aprendizaje es posible la creación de sistemas de búsqueda semántica, que ayuden al usuario a encontrar más fácilmente aquellos recursos didácticos que realmente se adapten a sus necesidades. La conveniencia de estos nuevos sistemas de búsqueda la han puesto de manifiesto numerosos autores, especialmente a partir de la iniciativa de creación de una Web Semántica que sustituya a la Web Lingüística actual [22].

El elemento fundamental de esta nueva forma de organizar la información es la ontología. Una ontología es “una especificación explícita de una conceptualización” [23], es decir, una representación formal de un dominio de conocimiento. Las ontologías se expresan a través de lenguajes de marcado basados en XML, como OWL (*Ontology Web Language*) [24], lo que facilita su reutilización en diferentes plataformas de anotación y búsqueda de recursos. Estos lenguajes definen etiquetas para representar los diferentes elementos de la ontología. Una ontología puede utilizarse para anotar recursos que posteriormente serán susceptibles de ser localizados por los usuarios que utilicen un buscador semántico que tiene acceso a la ontología.

Recientemente, algunos autores han propuesto la utilización de ontologías en repositorios de objetos de aprendizaje para facilitar la organización y búsqueda de estos recursos didácticos [25, 26]. Teniendo en cuenta estas iniciativas sobre la incorporación de ontologías en los repositorios de objetos de aprendizaje, es posible también la aplicación de técnicas de Ingeniería Lingüística que han demostrado su utilidad en el tratamiento de ontologías, ya que es evidente que existe una estrecha relación entre las ontologías y la Lingüística. Algunos autores afirman que una ontología es un

recurso lingüístico más, como los vocabularios, las taxonomías, los tesauros o las bases de datos léxicas [27]. En una base de datos léxica las unidades léxicas (palabras) son descritas como objetos lingüísticos, pero pueden ser relacionadas con una jerarquía conceptual localizada en una ontología. En el caso de los tesauros, las ontologías pueden considerarse como una evolución de éstos, caracterizadas por un desarrollo semántico más profundo en las relaciones entre conceptos, lo cual es posible por el uso que hacen de la lógica de descripciones [28].

Como afirma Vossen [29], la Ingeniería Lingüística puede resultar de gran utilidad para asociar palabras y expresiones con conceptos de una ontología; esto es, encontrar el significado de las palabras, así como minimizar la ontología y relacionar automáticamente los documentos o términos con las clases de la ontología. Además, siguiendo el razonamiento de este autor, dado que gran parte de la información todavía está en formato textual, aunque se desarrollen agentes “inteligentes” que ayuden al usuario en la búsqueda e interpretación de la información, éstos necesitarán de la Ingeniería Lingüística para “aprender” de esta información textual; y, por otra parte, también hay que tener en cuenta que al final de un proceso de búsqueda se requiere una comunicación lingüística del agente con el usuario.

En cuanto a la posible utilidad de las técnicas lingüísticas en un repositorio basado en ontologías, se sugieren las siguientes: 1) utilizar estas técnicas para crear la ontología del repositorio a partir de información textual, especialmente para extraer los conceptos relacionados con los dominios de conocimiento con los que están relacionados los objetos e aprendizaje almacenados; 2) combinar ambas tecnologías en los procesos de búsqueda de objetos, utilizando el conocimiento del dominio representado en la ontología para limitar los resultados de la búsqueda, aplicando, como sugiere Vossen [29], la Ingeniería Lingüística para extraer información de un texto que coincida con esa ontología, es decir, interpretando sólo expresiones que tengan sentido dentro del marco interpretativo de la ontología.

6 Conclusiones

En este artículo se han presentado algunas propuestas de aplicación de la Ingeniería Lingüística en el campo del e-learning. La primera propuesta se refería a la posibilidad de generar de forma automática algunos metadatos estándar asociados a los objetos de aprendizaje con valores no normalizados; por ejemplo, el metadato *Description* establecido por la especificación IEEE-LOM. Estas técnicas podrían integrarse con las propuestas de otros autores para la generación de bibliotecas de objetos de aprendizaje a partir de contenidos preexistentes [30], que resuelven el problema de la generación de la estructura de los objetos a partir de plantillas y hojas XSLT, pero que no contemplan la posibilidad de generar de forma automática también el contenido de los metadatos.

Otra posible aplicación analizada ha sido la posibilidad de generación de cuestionarios de evaluación compatibles con el estándar IMS-QTI. En este sentido, la propuesta incluida en este trabajo se ha basado en la versión 1.2 de esta especificación, y es intención de los autores adaptarla a la nueva y más completa versión 2.0 de la norma, contemplando todos los posibles tipos de cuestionarios contemplados por la

nueva versión. Además, aunque no lo contempla QTI, nuestra intención es ampliar el sistema informático de evaluación al caso más complejo de generación y corrección automática de exámenes, mediante el análisis del texto de las respuestas expresadas por los alumnos en Lenguaje Natural, a partir de la experiencia de otros autores en este campo [31].

Por último se ha tratado la posible combinación de las técnicas de Ingeniería Lingüística en los repositorios de objetos de aprendizaje y sistemas de búsqueda asociados, para su integración en la arquitectura de repositorios sugerida por IMS-DRI. En este sentido, la propuesta es aplicar las técnicas ya utilizadas en los sistemas de recuperación de información (SRI), teniendo en cuenta que los objetos de aprendizaje no son sino un caso particular de recurso de información. Al igual que ocurre con los SRI, que empiezan a utilizar métodos de búsqueda semántica que mejoren los resultados de las consultas de los usuarios, para hacer realidad, en el caso de Internet, lo que se conoce como Web Semántica, los repositorios de recursos de aprendizaje también comienzan a incorporarse a esta tendencia, por lo que se han sugerido en este trabajo algunas posibles combinaciones de la Ingeniería Lingüística con las ontologías en las que se basarán los repositorios semánticos en el futuro.

Referencias

1. Hafer, M., Weiss, S.: "Word segmentation by letter successor varieties", *Information Storage and Retrieval*, No. 10, 1974, 371-385.
2. Paice, C. D.: "A method for the evaluation of stemming algorithms based on error counting", *Journal of the American Society for Information Science*, Vol. 47, No. 8, 1996, 632-649.
3. Porter, M.F.: "An algorithm for suffix stripping", *Program*, Vol. 14, No. 3, 1980, 130-137.
4. OCENET Consulta: Un portal de consulta en español basado en tecnologías lingüísticas, Oficina del Español en la Sociedad de la Información (Instituto Cervantes), 2003. http://www.cervantes.es/seg_nivel/lect_ens/oesi.
5. Panessi, W., Bordignon, F. R.: "Procesamiento de variantes morfológicas en búsquedas de textos en castellano", *Revista Interamericana de Bibliotecología*, Vol. 24, No. 1, 2001, 69-88.
6. Figuerola, C.G., Gómez, R., López, E.: "Stemming and n-grams in Spanish: An evaluation of their impact on information retrieval", *Journal of Information Science*, No. 26, 2000, 461-467.
7. Frakes, W. B., Baeza-Yates, R.: *Information Retrieval: Data Structures & Algorithms*. Prentice Hall, Englewood Cliffs, NJ, 1992.
8. Kowalski G., Maybury, M. T.: *Information Storage and Retrieval. Theory and Implementation* (2ed), Kluwer Academic Publishers, Berlin, 2000.
9. Climent, S.: "Sistemas de resumen automático de documentos". *Revista Digital de Humanidades*, 2001. http://www.uoc.edu/humfil/digithum/digithum3/catala/Art_Climent_esp/Climent/climent.html.
10. Turney, P.: "Coherent keyphrase extraction via Web mining". *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 2003, 434-439.
11. Barchino, R., Gutiérrez, J.M., Hiler, J.R., García, E.: "Sistema multimedia inteligente para la generación y corrección de exámenes en Internet", *Actas del 2º Simposio Internacional de Informática Educativa* (edición electrónica), Puertollano, España, 2000.

12. IMS Learning Resource Meta Data, IMS Global Learning Consortium, 2001. <http://www.imsproject.org/metadata>.
13. Shareable Content Object Reference Model (SCORM), ADL. <http://www.adlnet.org>.
14. IEEE 1484.12.1-2002 Learning Object Metadata. Institute of Electrical and Electronics Engineers, New York USA (2002). <http://ltsc.ieee.org>.
15. Gómez, A.: Extractor automático de sintagmas relevantes para colecciones de documentos en castellano, TFC (Director: L. Bengochea), Universidad de Alcalá, 2004.
16. IMS Question and Test Interoperability, IMS Global Learning Consortium. <http://www.imsproject.org/question>.
17. Manero, B.: "Estudio de la propuesta IMS de estandarización de enseñanza asistida por computadora", Informe Técnico, Universidad Complutense de Madrid, sept. 2003. <http://eaula.sip.ucm.es/en/publications/informeTecnico.pdf>
18. IMS Digital Repositories Interoperability, IMS Global Learning Consortium. <http://www.imsproject.org/digitalrepositories>.
19. Xu, J., Croft, W., Corpus-Based Stemming using Co-occurrence of Word Variants. ACM Transactions on Information Systems, 16(1) 61-81, 1995.
20. van Rijsbergen, C. J., Information Retrieval, second ed. Butterworths, 1979.
21. Church, K. Hanks, P., Word association norms, mutual information and lexicography. Proceedings of the 27th ACL Meeting, pp. 76-83, 1989.
22. Berners-Lee, T., Hendler, J., Lassila, O.: "The Semantic Web", Scientific American, Vol. 284, No. 5, pp. 34-43, 2001.
23. Gruber, T.R.: "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition*, No. 5, pp. 199-200, 1993.
24. Web Ontology Language (OWL), World Wide Web Consortium. <http://www.w3.org/2004/OWL>.
25. Soto, J., Sánchez-Alonso, S., Sicilia, M.A.: "Flexibility in semantic learning object repositories", Recent Research Developments in Learning Technologies, 2005 (próxima publicación).
26. Sicilia, M.A., Cuadrado, J., Rodríguez, D.: "Ontologies of Software Artifacts and Activities: Resource Annotation and Application to Learning Technologies", Proc. of the 17th Int. Conf. SEKE, Taipei, Taiwan, China, July, 14-16, 2005.
27. Arano, S.: La ontología: una zona de interacción entre la Lingüística y la Documentación. En: <http://www.hipertext-net/web/pag220.htm>
28. García, A.: "Instrumentos de representación del conocimiento: tesauros versus ontologías", *Anales de Documentación*, No. 7, pp.79-95, 2004.
29. Vossen, P.: "Oportunidades para la Ingeniería Lingüística", Revista Digital D'Humanitats, No. 3. <http://www.uoc.es/humfil/articles/esp/vossen/vossen.html>
30. Iriarte, L., Marco, M., Morón, D., Pérez-Sancho, C., Pernías, P.: "Generación de una biblioteca de objetos de aprendizaje a partir de contenidos preexistentes", SPDECE'2004, Guadalajara, 2004.
31. Bucholz, S., Daelemans, W.: "Complex Answers: A Case Study using a WWW Question Answering System". *Journal of Natural Language Engineering*, Vol. 7, No. 4 (2001).