



DESARROLLO DE UNA APLICACIÓN PARA EVALUAR EL MERCADO DE TRABAJO EN BIBLIOTECONOMÍA Y DOCUMENTACIÓN

Sonia Sanchez-Cuadrado, Jorge Morato, Jose Antonio Moreiro¹

Valentín Moreno

Departamento de Informática

Universidad Carlos III

{ssanhec, jorge, vmoreno}@ie.inf.uc3m.es

¹*Departamento Biblioteconomía y Documentación*

jamore@bib.uc3m.es

RESUMEN

La tarea de evaluar la adecuación de las habilidades y competencias profesionales de los estudiantes de biblioteconomía y documentación es requerida para diseñar los planes de estudio, pero realmente laboriosa. Para facilitar el estudio se ha desarrollado una aplicación software, que con un mínimo mantenimiento, puede presentar estudios de forma coherente y periódica. Se ha desarrollado un tesoro para representar la terminología de ambos dominios: las ofertas de trabajo y los planes docentes de Biblioteconomía y Documentación. Los resultados presentan de forma coherente como el ajuste de los currículos para las ofertas en el mundo laboral. Los resultados obtenidos han sido comparados con estudios previos, basados en medidas para validar la adecuación de nuestra propuesta.

ABSTRACT

The task of comparing the labour markets demands with the abilities and professional competences of the students in Library and Information Science is necessary for the design of the curricula of their studies; nevertheless this work entails a concerted effort for the investigators. A software application has been developed in order to facilitate periodical study. A thesaurus has been developed to represent the terminology in both domains: labour market demands and the university training in Library and Information Science. The result displays in a coherent way the adjustment of curricula for the demands in the labour market. The obtained results have been compared with previous studies, based on surveys, to validate the suitability of our proposal.

PALABRAS CLAVES

Mercado de trabajo, estudios de biblioteconomía y documentación, formación universitaria, tesauros, estudios de mercado



INTRODUCCIÓN

El conocimiento de los perfiles más demandados en las ofertas de empleo en Biblioteconomía y Documentación tiene un interés indiscutible para la mejora de la formación de estos profesionales. Debido a este motivo se han realizado en los últimos años distintos informes y estudios (ANECA, 2004; ECIA, 2004).

La adecuación de los estudiantes en Biblioteconomía y Documentación a estos perfiles también ha sido protagonista de distintos análisis. Con frecuencia, la metodología ha sido realizada mediante dos tipos de estrategias:

1. Elaboración de encuestas y entrevistas personales: El acceso a los encuestados se realiza usualmente mediante técnicas tales como cuestionarios remitidos por correo electrónico (Fujino, 2007) o entrevistas telefónicas (Tejada, 2003). Este tipo de estrategias tiene un sesgo debido a su dependencia a la predisposición y credibilidad del encuestado.
2. Comparación manual de planes de estudios y ofertas y demandas de empleo. Una alternativa a las entrevistas personales es el análisis manual de las ofertas de empleo. Este enfoque se puede encontrar en el trabajo del investigador Moreiro (1999), donde se analizan las demandas y ofertas recibidas por los Servicios Universitarios de Orientación y Planificación Profesional.

Dada la laboriosidad de estas tareas se deben acotar los estudios, bien sea geográficamente, bien temporalmente. Algunos ejemplos de los mismos son:

- Acotación geográfica o según centro de estudios. Un ejemplo de este tipo es el estudio de la inserción laboral de los titulados de la Universidad de Barcelona (Borrego, 2004), también el estudio centrado en España (Moreiro, 2001), su circunscripción a la Universidad Carlos III (Moreiro, 1999), y a la Universidad Complutense (Tejada, 2002), entre otros.
- Acotación temporal. El estudio del empleo en el sector entre los años 2000-2005 (Moreiro, 2008), y también el estudio similar de licenciados en la Universidad de Sao Paulo entre los años 2000-2004 (Fujino, 2007), etc.

En el presente trabajo, se propone un método que permite eludir los sesgos indicados en las encuestas debido a su menor coste manual. En las próximas secciones se muestra el planteamiento del proyecto, así como los primeros resultados provisionales.

OBJETIVOS

El objetivo general es facilitar la comparación de los planes de estudio del ámbito de la Biblioteconomía y Documentación con las demandas de empleo en el área.

El método se ha propuesto para la obtención del siguiente conocimiento:



- Conocer la adecuación de los estudios universitarios al entorno profesional y laboral en el área de la Biblioteconomía y la Documentación.
- Identificar aquellas competencias y actitudes que están subrepresentadas en los planes formativos. De este modo se propone habilitar una herramienta de mejora del currículo docente.

La principal característica que se persigue con la propuesta es conseguir una metodología que permita reutilizar los recursos existentes y que pueda ser automatizada en el mayor grado posible. Al disminuir los costes se pretende automatizar la elaboración de informes que muestren dicha comparación de forma periódica, con los menores sesgos y acotaciones posibles.

TRABAJOS PREVIOS

El enfoque seguido ha sido similar al propuesto por Sánchez-Cuadrado (2007a) para mejorar los sistemas de transferencia de conocimiento entre los callcenter y los analistas de software de SAGE-SP. Mediante procesamiento del lenguaje natural y la creación de diversos tesauros y ontologías se analizan los textos de las llamadas telefónicas detectando estadísticamente las mejoras más demandadas por los usuarios del software de dicha empresa (Sanchez-Cuadrado, 2007b).

Un planteamiento con alguna similitud, pero con objetivos bien distintos, se ha utilizado en el pasado para la creación de ontologías. Así, FCA-Merge (Gómez-Pérez, 2004) recibe como entrada dos ontologías que se desean fusionar y un conjunto de documentos cuyo contenido trata sobre el dominio de las ontologías. La fusión de las ontologías se consigue extrayendo de los documentos instancias que pertenecen a los conceptos de ambas ontologías. Si el concepto C1 de la ontología O1 tiene instancias en los mismos documentos que el concepto C2 de la ontología O2, entonces C1 y C2 son candidatos a ser considerados el mismo concepto. Así la ontología fusionada se va “podando” para que resulten sólo los términos existentes en los documentos, y los necesarios para unir esos términos. FCA-Merge es un método adecuado para ontologías con un nivel semántico similar a los tesauros.

Un diseño más próximo a nuestro trabajo es el que realiza Sensus para adaptar su ontología a las instancias de un nuevo dominio (Gómez-Pérez, 2004), el proceso es el siguiente:

1. Identificar “términos semilla”. Un término semilla es un término que resulta especialmente relevante en el dominio que se quiere modelar. Normalmente, se trata de conceptos específicos de la ontología Sensus, aunque no siempre aparecen en esta ontología.



2. Los términos semilla se deben relacionar semánticamente a conceptos que si estén presentes en la ontología Sensus.

3. Añadir todos los conceptos presentes entre el concepto raíz de Sensus y los identificados en dicha ontología en el punto anterior. Los conceptos no identificados en el anterior punto y que no se encuentran en la jerarquía que une los conceptos con la raíz son eliminados. Dicho de otro modo, se “podan” o eliminan aquellos nodos no existentes en el nuevo dominio y no necesarios para unir los conceptos identificados. El resultado es una subontología de la ontología original.

4. Por último, la subontología se puede extender con nuevos conceptos próximos que no hayan aparecido anteriormente, adoptando la hipótesis de que si muchos términos del subárbol son relevantes todo el subárbol lo es.

De acuerdo con estos postulados, Rueda (2008) creó una aplicación capaz de especializar un tesoro genérico preexistente a partir de una colección de documentos representativos de una temática inmersa en el tesoro genérico. El software empleado en este proyecto se ha desarrollado bajo un planteamiento similar a la propuesta en dicho trabajo.

METODOLOGÍA

El método propuesto persigue comparar dos dominios solapados mediante la creación de un único tesoro que contenga la terminología presente en ambos dominios. Para conseguir este fin son necesarios los siguientes elementos (Figura 1):

A. Creación de un corpus documental representativo de ambos dominios:

1. Recopilación de documentos para la creación del tesoro:

- Libro blanco del título de grado en información y documentación (ANECA, 2004)
- Oferta de prácticas y empleo informe gestión SOPP 2007 de la Universidad Carlos III [<http://www.fundacion.uc3m.es/Sopp/>]
- Planes de estudios del grado en Biblioteconomía y Documentación [www.uc3m.es]

2. Recopilación de un corpus de ofertas de trabajo: se recopilieron 1087 ofertas entre septiembre de 2005 a marzo de 2008 del Web de Documentación [www.documentación.com]. La elección del site, elaborado por Javier Leiva y María José Solá, se debió a ser el la mayor recopilación de demandas de trabajo en biblioteconomía y documentación en España. Un valor añadido para la elección de este recurso es la capacidad de poder acotar por fechas para la realización de estudios acotados en el tiempo.

B. Creación de un tesoro genérico:

1. Desarrollo de estructuras de almacenamiento para el material base: dada la necesidad de almacenar de forma eficiente el tesoro y poder consultarlo de forma



rápida se almacenó mediante la base de datos referencial MySQL. La importación del tesoro se realiza mediante una aplicación JAVA de forma automática a partir de un tesoro en formato alfabético.

2. El tesoro genérico se creó de forma manual, con ayuda del editor Domain Reuser, y utilizando como fuente de conocimiento el vocabulario procedente de los corpus descritos previamente. La extracción de términos se hizo con ayuda de herramientas de Procesamiento del Lenguaje, que determinaban los sustantivos presentes en los textos. La estructuración de dichos sustantivos mediante relaciones semánticas se realizó manualmente.

C. Recopilación de los términos que compondrán el vocabulario del subtesoro:

1. Identificación de aquellos términos del tesoro que están presentes en los documentos. En la comparación entre las palabras de los documentos y los términos del tesoro hay que prever posibles variaciones debidas a:

- Normalización por presencia de diferente forma flexionada o derivada entre los dos recursos.
- En los términos compuestos por varias palabras hay que neutralizar la presencia de palabras vacías. En principio se intentó disminuir este error mediante la proximidad de las palabras principales del concepto del tesoro en los documentos. El orden de los términos no se tuvo en consideración.
- Otras características como acentuación o uso de mayúsculas no fue considerado. Con las soluciones previas, términos como “Técnicas analíticas e instrumentales”, y “Técnica instrumental y analítica” pueden ser identificados como el mismo término.

2. Establecer relaciones entre los términos del vocabulario: Una vez identificados los términos y para evitar “conceptos huérfanos” se pretende unir los conceptos jerárquicamente por el camino más corto a un mismo término “padre” del tesoro genérico. Esta acción implica añadir nuevos nodos que permitan unir estos términos. Se pregunta al usuario un número máximo de saltos, para evitar introducir, mediante un número excesivo de nodos, conceptos lejanos semánticamente.

3. Las relaciones con términos no descriptores de términos presentes en el subtesoro se importan directamente del tesoro genérico.

4. Se “podan” o eliminan los descriptores que o bien no hayan aparecido en los documentos, o bien no sirvan para unir los descriptores jerárquicamente en n saltos

5. Se comprueba si los términos que no han podido ser unidos a través de la jerarquía pueden ser unidos mediante asociaciones de relación del tesoro genérico.

6. Por último, se genera un tesoro expresado en un formato alfabético y con la estructura que permite importar y exportar por la aplicación *Domain Reuser*.

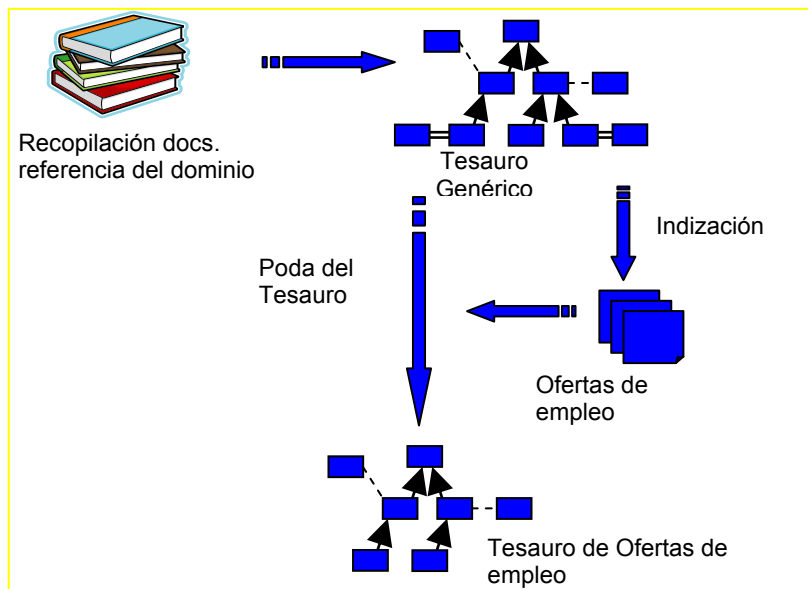


Figura 1. Metodología aplicada en el proyecto

RESULTADOS

El tesauro original tiene 812 términos con las siguientes familias: Búsqueda de empleo (Demandas y ofertas de empleo), Condiciones Laborales (contratos, salario, tipo de jornada), Empresas (área de actividad, sector de actividad, tamaño de empresa, tipo de empresa), Entidades e Instancias de Organismos, Formación (conocimientos, contenidos, formación reglada), Instrumentos (encuestas y entrevistas, foros, etc.), Lugares (topónimos), Perfil académico profesional, Perfil del solicitante (competencias, disponibilidad, edad, experiencia, habilidades, perfil ocupacional, perfil prácticas), Personas y Procesos (acreditación, gestión de contenido, etc.).

Del total de términos del tesauro genérico se identificaron 269 términos, a los que hubo que añadir 27 términos para unir los distintos subárboles, con un total de 296 términos. Estos 296 descriptores tenían un total de 76 sinónimos. Los siguientes descriptores han sido identificados en el tesauro específico, si bien no se trata de un listado completo.

- Áreas de Actividad: Administración, Agencia de Comunicación, Alimentación, Bancos, Comercio, Construcción, Consultora, Editoriales, Empresa Informática, Empresas De Contenidos, Financiero, Hostelería, Industria, Institución Educativa, Investigación, Legales, Prensa, Recursos Humanos, Salud, Seguros, Servicios, Transportes

- Conocimientos Demandados: Archivística, Biblioteconomía, Conocimiento de Idiomas, Documentación, Empresariales, Información, Informática, Lingüística, Tratamiento de Imágenes
- Administración de Bibliotecas: Análisis de Fuentes Documentales, Atención al Usuario de la Biblioteca, Catalogación del Fondo Bibliográfico, Gestión Bibliotecaria, Inventario, Organización de Bibliotecas,
- Tipos de Bibliotecas que más Demandan: Bibliotecas en Centros de Enseñanza, Bibliotecas Especializadas, Bibliotecas Musicales, Bibliotecas Publicas, Bibliotecas Escolares, Bibliotecas Universitarias
- Competencias demandadas: Capacidad de Comunicación, Capacidad de Trabajo, Competencias Especificas de la Titulación, Competencias Transversales (Instrumentales (Capacidad de Análisis y Síntesis, Resolución de Problemas, Toma de Decisiones), Personales (Capacidad de Trabajar En Equipo, Habilidades en las Relaciones Interpersonales) y Sistémicas (Creatividad, Iniciativa y Espíritu emprendedor, Liderazgo), Responsabilidad, Competencias en Formación, Competencias en Técnicas Comerciales
- Idiomas: Alemán, Castellano, Catalán, Francés, Inglés, Portugués
- Informática: Lenguajes de Programación (Java, Java Script, Perl, Php), Ofimática, Sistemas Operativos (Linux, Windows), Web (CMS, Diseño Web, Lenguajes de Etiquetado, Posicionamiento Web, Sindicación De Contenidos), Gestión de Bases de Datos
- Representación y Recuperación de Información (Catalogación Descriptiva, Clasificación Indización, Lenguajes Documentales, Recuperación de Información, Resúmenes), Sistemas de Organización del Conocimiento (Taxonomías, Thesaurus, Vocabularios de Metadatos), Tecnologías de La Información (Pruebas de Aplicaciones Informáticas, Postventa), Tratamiento de la Información (Búsquedas En Internet, Escaneado, Gestión De Bases de Datos, Tratamiento Informático de la Información, Elaboración de Informes Sobre La Competencia), Gestión de la Información (Archivo Documental, Búsqueda de Documentación, Mantenimiento de Archivos, Realización de Búsquedas de Información)
- Perfil Académico (Diplomado, Licenciado, Postgraduado), Perfil Ocupacional (Administración de Bibliotecas, Comercial, Formación de Personal, Gestion de la Información, Tecnologías de la Información, Tratamiento de la Información, Asistente de Proyectos), Perfil Profesional (Ayudante de Documentalista, Bibliotecarios Escolares, Control de Calidad de Fondo Antigo, Digitalización de Fondo Antigo, Documentalista, Gestores de Contenido, Grabador de Datos, Linkbuilder, Personal de Archivos, Webmaster).

PRIMEROS ESTUDIOS

Como consecuencia de la metodología descrita se han obtenido los siguientes resultados parciales:

- **Áreas de Actividad:** los conceptos con mayor relevancia son Medios de Comunicación (20%), Educación (18%) y Banca (14%) y Editoriales (13%). Los datos contrastan con los resultados obtenidos en el análisis manual realizado por el COIE de la Universidad Complutense en el 2004, donde los principales grupos fueron Medios de comunicación (32%), Consultoría (13%), Servicios a empresas y a la comunidad (16%), Telecomunicaciones (12%), Editoriales (5,4%). Moreiro (1998), que basó su estudio en encuestas, obtuvo los siguientes datos: Administración pública (43%), Consultoría (9%) y Medios de Comunicación (7,6%). Todos los estudios destacan la importancia de los Medios de Comunicación.
- **Conocimientos y Tareas:** En el análisis desarrollado se ha constatado la importancia que ha adquirido la informática. Su aparición se observaba en diferentes áreas. De este modo, el conocimiento de lenguajes de programación aparece con el 13%, la ofimática aparece con un 11%, las bases de datos con un 2%. Dentro de los conocimientos informáticos, tiene especial importancia el conocimiento de relacionados con la Web, que aparece en el 30% de las ofertas, el conocimiento de las técnicas de posicionamiento (7%) y los lenguajes de programación más asociados con la Web (JavaScript -7,5%-, XML -5%-, PHP -4%-, Java -2%-). Otros conocimientos demandados son el diseño Web y la usabilidad (6 %). La importancia de la informática anteriormente había sido corroborada por el informe del COIE de la UCM en el 2004, donde la ofimática aparecía en el 16% de las ofertas, las bases de datos en el 12,7 % (para Access y para otras Bases de Datos) y los lenguajes de programación en el 7.5%.

Los Conocimientos más demandados, excluida la informática, son: Conocimientos de documentación 56%, bibliotecas 12,5%, idiomas 5%, publicidad, marketing e investigación comercial 5%, relación con proveedores y soporte clientes 5%, representación y recuperación de información 5%, escaneado 2%, etc. Concretamente, en este apartado se detectan las mayores diferencias con trabajos previos, así en Moreiro (2008) se manifiestan como tareas más demandadas la catalogación 8%, el mantenimiento de bases de datos 7%, la organización de archivos 7%, los servicios de búsqueda 6,5%, etc. Estas carencias en los planes formativos se descubren también en las facetas señaladas de publicidad y marketing y/o de soporte a clientes.

DIFICULTADES

Algunas dificultades que se han identificado en esta metodología se resumen en:

- La utilización de diferente terminología dificulta la comparación. En el estudio de Moreiro (2008) no aparece la banca como grupo destacado. En el trabajo del COIE aparece tan solo con un 3% y en el estudio efectuado aquí sube al tercer grupo en importancia. Sin embargo, la educación no aparece como grupo diferenciado en los estudios previos.
- Otro problema es la dificultad de comparar diferente granularidad. En los estudios previos se desconoce si cuando se habla de ofimática se está incluyendo temas que aparecen posteriormente desglosados como Word, Excel y Access.



- El tesoro debe mantenerse y ampliarse para ser una herramienta útil. De este modo, las ofertas que demandaban conocimiento de “lenguas cooficiales del Estado” del tesoro no fueron detectadas hasta el análisis manual, por ser un término que aparecía en pocas demandas y no haber sido localizado en la documentación empleada en la generación del tesoro.

CONCLUSIONES Y TRABAJOS FUTUROS

Se ha constatado la idoneidad de la respuesta, su menor coste y el gran potencial que puede suponer este enfoque para el futuro. La propuesta es aplicable para otros entornos sin apenas cambios en sus planteamientos.

Al evitar realizar encuestas se evita depender de las visiones parciales de los encuestados, y su acotación a un conjunto limitado de consultas, o su limitación a perfiles que normalmente no se asocian a los estudios de Biblioteconomía y Documentación. Todas estas consecuencias pueden aumentar la credibilidad de los resultados, si bien por el momento es necesario contrastar los datos del estudio con las metodologías tradicionales para confirmar su idoneidad.

Aunque los resultados son provisionales, el estudio realizado permite comprobar la validez de la propuesta. No obstante, sería necesario un mayor desarrollo de los recursos de análisis para dar un potencial equivalente a los trabajos referenciados. El tesoro realizado tiene actualmente casi un millar de términos, tras el procesamiento de los corpus se ha constatado que la calidad del análisis tiene aparejada la actualización y expansión del tesoro actual.

La mayor ventaja reside en su capacidad de reutilización y automatización que permitirá la repetición del estudio por estratos (por ejemplo a conjuntos de la población, a determinado periodo temporal o geográfico). De esta manera, este método basado en tesoros permite agrupar los conceptos más específicos de la jerarquía, con lo que los informes pueden graduarse según distintos niveles de abstracción.

Del análisis de los resultados se debe enfatizar la ausencia en los planes formativos de competencias como son la publicidad y marketing o el soporte a clientes, que ha resultado ser muy demandadas por las empresas.

Como trabajo futuro, se depurará, ampliará y complementará el tesoro y el corpus de análisis, y se automatizará la elaboración de informes anuales en Web.

Por otra parte, está en desarrollo realizar comparativas entre las situaciones entre Brasil-España, traduciendo el tesoro e implementando las herramientas lingüísticas de forma bilingüe.

AGRADECIMIENTOS



El siguiente estudio ha sido financiado por los programas de formación de la Fundación Carolina durante el año 2008.

REFERENCIAS BIBLIOGRÁFICAS

ANECA. Libro blanco del título de grado en información y documentación. Madrid, España : Agencia Nacional de Evaluación de la Calidad y Acreditación, 2004. 238p.

BORREGO, A.; COMALAT, M.; ESTIVIL, A. “Inserció laboral dels titulats en Biblioteconomia i Documentació per la Universitat de Barcelona”. En: *BiD, textos universitaris de biblioteconomia i documentació*, n. 12, juny 2004. Disponible en: http://www2.ub.es/bid/consulta_articulos.php?fichero=12borreg.htm (consultado el 2 de julio de 2008)

ECIA Euroreferencial en Información y Documentación. Competencias y aptitudes de los profesionales europeos de información y documentación. 2 ed. Madrid, España : SEDIC, European Council of Information Associations, 2004, 127p. <http://www.certidoc.net> (consultado el 2 de julio de 2008)

FUJINO, A. ; LARA, M. ; AMARO, R. ; VERGUEIRO, W.; TEJADA ARTIGAS, C. ; MOREIRO, JA. Perfil profissional e competências de informação em Biblioteconomia e Documentação: pesquisa com egressos da ECA/USP. In: XXII CBBB, 2007, Brasília. Anais XXII CBBB. Brasilia: FEBAB, 2007.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M.; CORCHO, O. Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web, London, Springer, 2004: 403 p. ISBN 1852335513.

MOREIRO GONZÁLEZ, J. A. "Ajuste de los licenciados en documentación a la oferta de trabajo. Observaciones hechas en la Universidad Carlos III de Madrid". En: *Revista general de información y documentación*, 1999, v. 9, n. 2, pp. 13-28.

MOREIRO GONZÁLEZ, J. A. “Figures on employability of spanish library and information science graduates”. En: *Libri*, vol. 51 2001, pp. 27-37

MOREIRO-GONZÁLEZ, JA; AZCÁRATE-AGUILAR-AMAT, P.; MARZAL-GARCÍA-QUISMONDO, MA. “Desarrollo profesional y opinión sobre la formación recibida de los titulados universitarios en información y documentación de las universidades públicas de Madrid (2000-2005)”. *El Profesional de la Información*, 2008, vol. 17, núm. 3

RUEDA, Ana. Construcción de ontologías a partir de la especialización de fuentes estructuradas. Proyecto Fin de Carrera dirigido por J. Morato y V. Moreno. Madrid: Universidad Carlos III, 2008.



SÁNCHEZ-CUADRADO, S; MORATO, J; MOREIRO, JA; MARRERO, M.
“Definición de una metodología para la construcción de sistemas de organización del conocimiento a partir de un corpus documental en lenguaje natural”. Revista de la SEPLN, 2007, v. 39, p. 213-220. XXIII Congreso de la SEPLN. Universidad de Sevilla 10, 11, 12 de septiembre de 2007

TEJADA ARTIGAS, C. Adecuación de los planes de estudio de la diplomatura en biblioteconomía y documentación a las demandas del mercado de trabajo. El caso concreto del nuevo plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de la Universidad Complutense de Madrid. Tesis doctoral. Madrid: Universidad Carlos III, 2002.

TEJADA ARTIGAS, C.; MOREIRO GONZÁLEZ, J. A. “Mercado de trabajo en Biblioteconomía y Documentación. Estudios sobre la inserción laboral de los titulados universitarios”. En: *El profesional de la información*, v. 12, nº 1, enero-febrero 2003, pp. 4-9.