

## **Modelo de Gestión de Datos Compartidos de Producción Científica**

Reporte de proyecto

Dra. Ma. Auxilio Medina<sup>1</sup>, Dr. J. Alfredo Sánchez<sup>2</sup>, Dr. Antonio Benitez R.<sup>1</sup> y Dr. Jorge de la Calleja Mora<sup>1</sup>

<sup>1</sup> Universidad Politécnica de Puebla, Programa Académico de Ingeniería en Informática  
Tercer Carril del Ejido Serrano S/N, San Mateo Cuanalá, Juan C. Bonilla, Puebla, México  
Tel. 01 (222) 7 74 66 41

{mmedina, abenitez, jcalleja}@uppuebla.edu.mx

<sup>2</sup> Universidad de las Américas Puebla  
Laboratorio de Tecnologías Interactivas y Cooperativas  
Exhacienda Santa Catarina Mártir S/N, San Andrés Cholula, Puebla, México  
Tel. 01 (222) 2 29 26 66

j.alfredo.sanchez@uppuebla.edu.mx

### **Resumen**

En instituciones de educación superior (IES), es común que la evaluación de la investigación se realice con base en indicadores que de forma cuantitativa, estiman la calidad de productos científicos. Este documento describe un modelo de gestión de datos compartidos de publicaciones y proyectos, con diseño adaptable a otros productos, que permite la recolección de datos y su descripción con estándares de metadatos, con el propósito de agregar valor práctico a los datos de las IES y posibilitar la colaboración entre profesores al hacer público la producción de los usuarios. El modelo se ha implementado como una aplicación web diseñada desde la perspectiva de los sistemas de organización y sus diferentes niveles de conocimiento. Los usuarios de la aplicación web realizan tareas como inserciones, bajas, consultas, reportes y exportación de datos al formato Bibtex. Las consultas se llevan a cabo a nivel personal, general o bien, se puede acceder a la producción de cualquier otro usuario.

**Palabras clave:** Bibliotecas digitales, recursos bibliográficos, gestión del conocimiento.

### **Introducción**

En ámbitos académicos, la investigación aplicada orientada a resolver problemas o motivada por la curiosidad intelectual, es frecuente que se perciba como un ciclo de generación del conocimiento, formado por las etapas siguientes: información, conocimiento, comunicación y producción. La información asigna una o más interpretaciones a los datos, cuando la información se procesa por un individuo o una comunidad se genera conocimientos. El conocimiento se distribuye a través de varios canales de comunicación y se representa en productos [3].

De acuerdo a [9], la investigación se realiza conforme a políticas públicas y procesos administrativos, y tiene

influencia directa en los sistemas de enseñanza y en comunidades científicas, o bien, atiende a un mercado bien definido.

En la evaluación y asignación de recursos en las IES, es frecuente que además de la formación de recursos humanos, se tome en cuenta la cantidad y calidad de un conjunto de productos científicos que se concretan en documentos como artículos, tesis, reportes técnicos, prototipos, proyectos o patentes.

Los indicadores bibliométricos se utilizan frecuentemente en los análisis cuantitativos de los productos generados, están orientados a lograr un análisis del grado de desarrollo en investigación científica y tecnológica de los investigadores. A través de medidas estadísticas, se trata de estimar la calidad de los productos. Independientemente de la confiabilidad y validez de los indicadores, para utilizarlos es indispensable contar con un registro de los productos.

Este documento propone transformar el conocimiento de la producción científica de una comunidad de profesores investigadores en valor práctico, a través de un modelo de gestión que potencie la colaboración y que permita recolectar datos que faciliten la aplicación de indicadores. La transformación se percibe desde la perspectiva de los sistemas de organización del conocimiento y se implementa en una aplicación web denominada "Sistema de Gestión de Producción Científica" (GPC), disponible en:

<http://informatica.uppuebla.edu.mx/~produccioncientifica/index.php>

El documento está organizado como sigue: la sección 2 describe brevemente trabajos relacionados relevantes; la sección 3 presenta el modelo de gestión y la perspectiva de investigación; la sección 4 describe la implementación del modelo y presenta algunas de las interfaces principales. La sección 5 muestra los resultados experimentales. Las conclusiones y el trabajo a futuro se citan en la sección 6.

## Trabajo Relacionado

La investigación es una de las principales tareas de los profesores adscritos a las IES. Instituciones como el Programa de Mejoramiento del Profesorado (PROMEP) y el Consejo Nacional de Ciencia y Tecnología (CONACYT), emplean herramientas de software para registrar, consultar y actualizar los productos científicos. En la sección “producción académica” de la sección “datos curriculares” del sistema de solicitudes en línea del PROMEP [14], se puede consultar el título, año, número de citas y tipo de producto académico. En ese apartado, junto con la sección “proyectos de investigación aplicada y desarrollos tecnológicos” y “dirección individualizada”, se concentra la producción científica que reportan los investigadores.

Por otra parte, en el sistema de captura del CVU (Curriculum Vitae Único) utilizado por el CONACYT (Consejo Nacional de Ciencia y Tecnología) [14], la producción científica se introduce en secciones como “proyectos de investigación y consultoría”, “divulgación y difusión”, “libros”, “reportes técnicos”, “artículos publicados”, “reseñas”, “capítulos de libro”, “estancias de investigación”, “participación en congresos”, “tesis dirigidas”, “desarrollos tecnológicos” y “patentes”.

Los sistemas [14] y [15] son medios oficiales utilizados ampliamente en las IES. Desde el punto de vista de este proyecto, ambos sistemas presentan dos desventajas: 1) no es factible acceder a la producción de otros investigadores y 2) no hay interoperabilidad semántica en diferentes niveles de organización. En este contexto, la interoperabilidad semántica se refiere por un lado, a que un mismo producto científico se describe de diferente forma en ambos sistemas y por otro lado, a que la forma de organizar y priorizar los productos científicos es dependiente del sistema.

Los datos de los sistemas [14] y [15] pueden exportarse a Word para procesarse posteriormente, sin embargo, no existe una estructura común que facilite el empleo de indicadores de calidad. Aunque los indicadores están condicionados por factores políticos y sociales, independientemente de su validez, fidelidad, limitaciones y desviaciones, una necesidad de las IES es disponer en tiempo y forma de datos para estimar el grado de desarrollo de actividades de investigación, científicas o tecnológicas.

Algunas IES, como la Universidad de Zaragoza, España, han optado por el desarrollo de herramientas a la medida que apoyen la recolección de los datos de los productos de investigación. El sistema de gestión de producción científica SIDERAL [14], es un ejemplo de este tipo de herramientas. En este sistema, los

usuarios realizan tareas de registro y consulta de datos acorde a los intereses de su contexto académico. El procesamiento de datos permite la generación de un currículum vitae normalizado (CVU) para sus usuarios así como la generación automática de reportes como “proyectos en los que ha participado algún investigador” y “artículos en revistas con factor de impacto”.

Existen herramientas donde los usuarios pueden acceder a sus propios datos o a los de otros investigadores como “*The DBLP Computer Science Bibliography*” (DBLP) [2], Thomson Reuters EndNote [12], Bibexcel[10], ViblioSOM [1] y Scopus [16]. Este tipo de herramientas cuentan con características únicas que requieren que los usuarios se familiaricen con cada una de las interfaces para interpretar datos de interés. En este proyecto, se propone un modelo que reduzca detalles técnicos y ofrezca un nivel de abstracción mayor en los elementos clave. La siguiente sección describe el modelo propuesto desde la perspectiva de los sistemas de organización del conocimiento.

## Modelo de gestión de producción científica

El modelo de gestión de producción científica se concibe desde el área de bibliotecas digitales en la web semántica, con siglas BDS. Las BDSs se consideran un tipo especial de sistemas de organización del conocimiento que integran materiales, herramientas y significados [8]. Los materiales o datos forman colecciones organizadas en archivos o bases de datos; las herramientas implementan servicios sobre los materiales, por ejemplo, digitalización, preservación, consulta, actualización o respaldo; los significados apoyan la interpretación no ambigua de los términos incluidos en los materiales y en sus estructuras de almacenamiento. La gestión de significados se basa en vocabularios bien definidos o taxonomías temáticas denominadas *ontologías* [5].

Una BDS atiende diferentes niveles de organización. En el modelo propuesto, los niveles son los siguientes:

- *Organización de información en bases de datos.* Una base de datos relacional almacena los datos de los siguientes productos de investigación: artículo indexado (artículo en revista de prestigio), artículo con arbitraje internacional y nacional o artículo de difusión, capítulo de libro, libro, memorias en extenso arbitradas y no arbitradas, reporte técnico, tesis de doctorado o de maestría. Para describir estas publicaciones, se extrajeron los elementos comunes reportados en [13], [14], y [15]. Se agregaron

datos descriptivos para representar un proyecto de investigación, el cual está representado como una entidad formada por un responsable y un grupo de colaboradores, título, objetivo general y específicos, fecha inicial, fecha de término, lista de productos esperados, un estado (en proceso o terminado) y un atributo booleano que indica si el proyecto recibe o no financiamiento. El diccionario de la base de datos guarda la definición de los objetos anteriores, sus características, restricciones, privilegios y la relación entre ellos.

- Organización de información en documentos.*- Los datos de los productos científicos se pueden almacenar en un archivo de texto, exportar a otra base de datos, almacenarse en formatos con alta interoperabilidad como XML u otros formatos acorde al manejador. En particular, es de interés la exportación a un archivo con formato Bibtex, el cual representa una ontología del editor Latex [7] utilizada para producir una lista de referencias. Bibtex emplea pares de etiquetas/valor que pueden procesarse por algunas de las herramientas descritas en la sección 2.
- Organización de metadatos.*- Los metadatos (definidos como datos altamente estructurados que describen el contenido, la calidad, la condición de uso y otras características de los datos [6]), corresponden al diccionario de la base de datos y a los elementos de entrada del archivo con formato Bibtex. Las características principales de este formato son: manejo de referencias independiente del documento que se edita, uniformidad de formato en función del estilo que indique el usuario y organización de referencias en más de un archivo, lo cual facilita que éstas se clasifiquen en varios criterios como área de investigación o fuente de la referencia.

El nivel de *organización de temas, así como el de organización de conceptos, términos y relaciones*, hace uso de una ontología construida a partir de las siete áreas del conocimiento propuestas por CONACYT. El usuario es el encargado de asociar una clase de la ontología como tema a cada producto de investigación. El modelo de gestión de producción científica de la Figura 1 se concibe como parte de una BDS que apoya el proceso de transformación del conocimiento de la producción científica en valor práctico, consta de 9 bloques que representan la lógica de un negocio o aplicación [4].

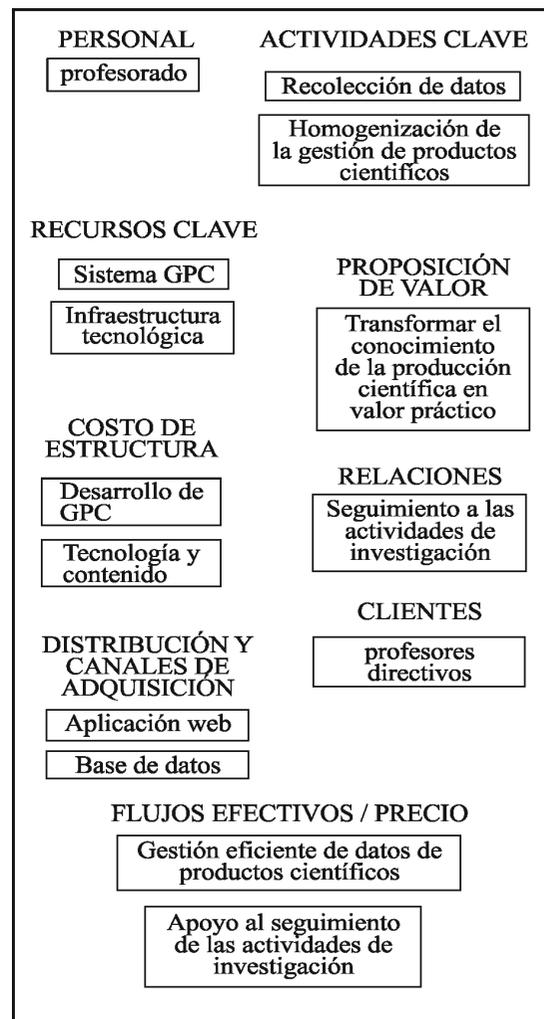


Fig. 1. Modelo de gestión de producción científica

Como se ilustra en la Figura 1, el personal se refiere a los profesores. Para alcanzar el objetivo del modelo, se debe dar seguimiento a las actividades de investigación, lo cual requiere de la recolección de datos y del uso de estándares de metadatos para describir productos científicos. El hecho de que los datos puedan exportarse al formato BibTex facilita su procesamiento por herramientas especializadas en la extracción de indicadores bibliométricos. Como infraestructura tecnológica, se propone un sistema de gestión de producción científica con una base de datos accesible vía web. Los usuarios pueden ser profesores o directivos de las diferentes IES.

En el modelo plantea que un usuario de la comunidad de profesores investigadores pueda consultar la producción de cualquier otro usuario. Esta característica potencia el trabajo colaborativo e interdisciplinario y se ajusta a una política de transparencia y apertura de información. La sección siguiente describe la implementación del modelo.

### GPC: Sistema de Gestión de Producción Científica

El sistema de Gestión de Producción Científica (GPC) se encarga de almacenar, gestionar, preservar y publicar los datos del modelo de gestión. La Figura 2 muestra la pantalla inicial. Del lado derecho se encuentra el objetivo del sistema y las instrucciones de uso. Se requiere que cada usuario ingrese con un identificador y una contraseña.

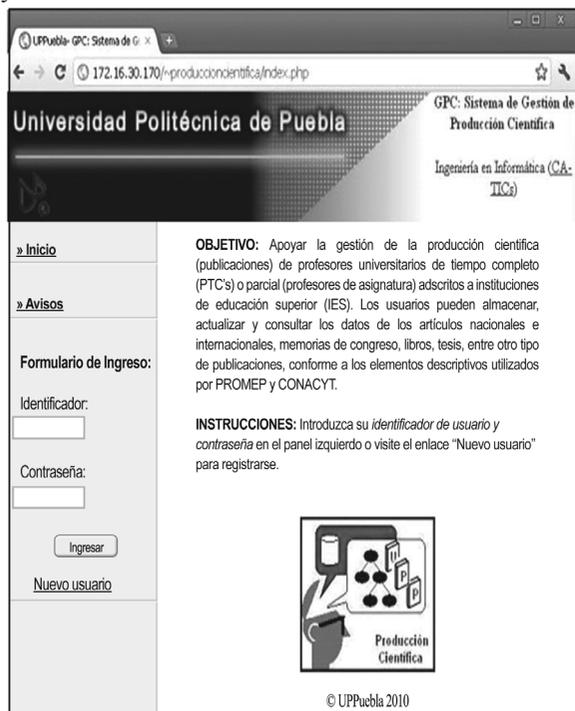


Figura 2. Página inicial del sistema GPC

El menú del GPC está dividido en tres áreas principales: gestión de producción científica, proyectos de investigación e información cuantitativa. (Ver la Figura 2). En las primeras dos áreas se realizan altas, bajas y modificaciones. En la tercer área, se utiliza la base de datos para construir gráficas que faciliten el acceso a información de tipo cuantitativa de interés general.

El diseño de GPC permite que en el futuro puedan incorporarse otro tipo de productos. Los usuarios tienen asignado un rol: administrador, profesor, directivo o público. A través de estos roles se implementa una política de seguridad para acceder a las diferentes páginas de la aplicación, así como a la base de datos.

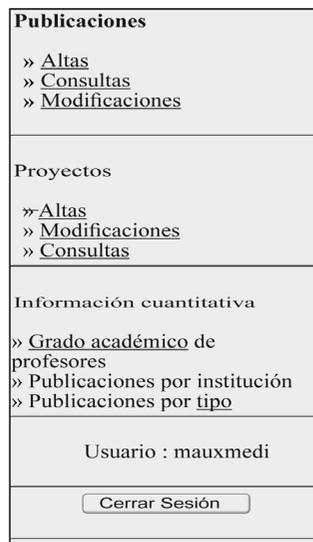


Figura 3. Página inicial del sistema GPC

La Figura 4 muestra el tipo de publicaciones que pueden agregarse en el sistema. Observe que el nombre de las publicaciones aparece en español e inglés, esto con el propósito de que el usuario se familiarice con las entradas de los objetos del formato Bibtex.

#### Alta de publicaciones

Selección de la lista el tipo de publicación que desea dar de alta



Figura 4. Tipos de publicaciones del sistema GPC

Las consultas de la sección de producción científica para un usuario con el rol de profesor son de tres tipos:

- 1) *Publicaciones de un profesor:* Después de seleccionar el nombre de un profesor, se solicita que el usuario especifique los criterios para ver los datos, por ejemplo, por año, tipo, ISSN o ISBN.
- 2) *Publicaciones de todos los profesores:* Lista la producción de todos los profesores ordenada por año, tipo, ISSN e ISBN.
- 3) *Estadísticas.-* Muestra información cuantitativa de la producción de todos los profesores o de uno sólo ordenada por año.

En la sección de gestión de proyectos se realiza el registro, la modificación y las consultas de los proyectos

de investigación. A manera de ejemplo, considere las Figuras 5 y 6 que muestran parte del resultado de la consulta que visualiza la publicación de todos los profesores por tipo y el resultado de una consulta estadística. De la Figura 5, el título de la publicación se muestra en negrita, seguido de la lista de autores. Posteriormente se muestran los datos descriptivos de la publicación. Observe que se incluye el rol de los profesores en cada producto científico, es decir, se precisa quién es el primer autor, el segundo u otro.

Opciones:	<b>Consulta: Publicaciones de todos los profesores por Tipo</b>	
» <a href="#">Menú anterior</a>	<b>Profesor:</b>	
Usuario: mauxmedi	Tipo	Publicación
<input type="button" value="Cerrar sesión"/>	Artículo con arbitraje internacional	<b>Inverse Heat Transfer using Levenberg-Marquardt and Particle Swarm Optimization Methods for Heat Source Estimation</b> , Autor: Cortes Aburto, Obed, ISSN: 1662-7482, Applied Mechanics and Materials, 15,35-40, Número: 1 Primer autor  <b>Realization of multiscroll chaotic attractors by using current-feedback operational amplifiers</b> , Autor: Tlelo Cuautle, Esteban, ISSN: 0035-001X,-----, Revista Mexicana de Física, 0, 0, Número: 0 Segundo autor

Figura 5. Publicaciones de profesores por tipo

Opciones:	<b>Consulta: Estadística por año</b>		
» <a href="#">Menú anterior</a>	<b>Profesor:</b>		
Usuario: mauxmedi	Año:	Tipo de Publicación	Total
<input type="button" value="Cerrar sesión"/>	2010	Memorias en extenso no arbitradas	8
		Artículo con arbitraje nacional	5
		Memorias en extenso arbitradas	3
		Artículo con arbitraje internacional	2
		Reporte técnico	2
		Capítulo de libro	1
		Revista	1
	2009	Artículo con arbitraje nacional	7

Figura 6. Número y tipo de publicaciones por año

La Figura 6 muestra el tipo y número de productos científicos por año. Esta información pertenece a las diferentes IES, pero puede accederse también a la información de una en particular. Otros datos de interés se recuperan a través medidas estadísticas y de la implementación de las consultas desde el manejador de la base de datos. Información de interés general de tipo cuantitativo se muestra al establecer una conexión entre la base de datos y la versión 5.4 de la librería PHPlot, como se aprecia en las Figuras 7 y 8.

GPC está desarrollado con herramientas de software libre. Los datos se almacenan en el manejador MySQL versión 5.0.77. La interfaz de usuario está formada por páginas web y scripts en lenguaje PHP versión 5.3.2. GPC se ejecuta sobre un servidor bajo el sistema operativo Red Hat 5.3

### Resultados Experimentales

GPC es una herramienta que puede apoyar la toma de decisiones académicas y administrativas de las IES, habilita la generación de reportes y facilita la evaluación de actividades científicas del profesorado. Los profesores adscritos a la Universidad Politécnica de Puebla (UPPuebla) han colaborado con pares de las siguientes IES: Benemérita Universidad Autónoma de Puebla (BUAP), el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOEP), Instituto Politécnico Nacional (IPN), Centro de Investigación en Ingeniería y Ciencias Aplicadas de la Universidad Autónoma del Estado de Morelos (UAEM), Universidad Nacional Autónoma de México (UNAM), Universidad de las Américas – Puebla (UDLAP), el “Laboratoire d’Analyse et d’Architecture des Systemès” y del “Computer Science Department University of Texas El paso”. La Figura 7 muestra el porcentaje de productos científicos procedentes de las instituciones que aparecen en la leyenda de la derecha.



Figura 7. Porcentaje de publicaciones por institución

En la Figura 8 se muestra una gráfica que cuenta el número de productos de investigación por tipo. En el eje X se muestra un número asociado al tipo de publicación de acuerdo con la leyenda de la gráfica. El número de publicaciones de un tipo, se representa con el valor del eje Y. Esta información es útil principalmente para usuarios administrativos, podría utilizarse para enfatizar la calidad y no la cantidad de los productos científicos. La información de la Figura 8 pertenece a todos los profesores, sin embargo, existe otra opción del menú en donde después de seleccionar el nombre de un profesor y de introducir un periodo de tiempo, se recupera su producción científica solicitada.

De la base de datos se pueden extraer indicadores como el índice de producción, definido como la cantidad de autores responsables del 50% de las publicaciones, o el índice de transitoriedad, que es la cantidad de autores responsables de un solo trabajo.

Actualmente, se han recopilado datos de la producción de los 2 años anteriores de algunos de los profesores investigadores adscritos a la UPPuebla, El número de publicaciones que actualmente se encuentran registradas es de 58. Se espera que este número se incremente en un corto plazo. La captura de éstas se realizó por una sola persona, pero se planea a corto plazo que cada usuario realice esta tarea. Sin embargo, no se ha realizado una evaluación de la efectividad del modelo ni a su capacidad para modelar el trabajo colaborativo de los usuarios.

Publicaciones por tipo

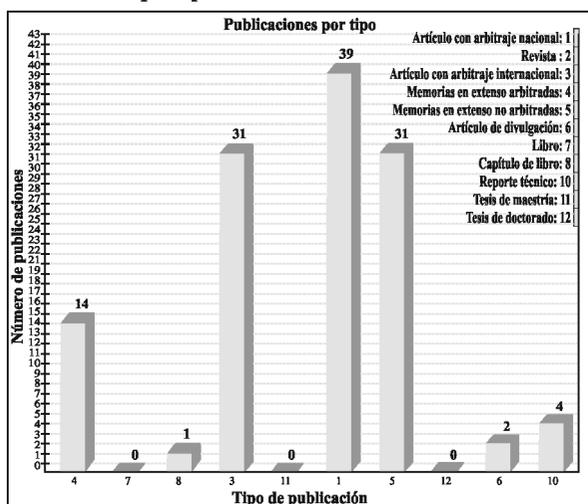


Figura 8. Número de publicaciones por tipo de todos los profesores

## Conclusiones y Trabajo a Futuro

Comúnmente, las IES recolectan datos de productos científicos. La selección de una herramienta de apoyo o el desarrollo de una hecha a la medida es difícil en términos de la ausencia de un instrumento que represente las características de los contextos de aplicación que podría resultar en un caso de éxito. Este documento propone un modelo de gestión que se implementa en una aplicación web desarrollada con software libre. La aplicación está disponible para los usuarios que así lo soliciten.

El modelo propuesto identifica elementos esenciales y descarta detalles técnicos que promueva una cultura de mejora continua en la calidad de la gestión de los productos científicos. Propone transformar el conocimiento de la producción a valor práctico que represente una ventaja competitiva para las IES. Algunos de los beneficios que puede ofrecer el modelo son: 1) seguimiento a los proyectos de investigación, 2) homogenización de criterios de descripción y evaluación de productos científicos a través de estándares de metadatos, 3) integración de datos para mostrar tendencias de investigación en las IES, 4) agrupación y comparación de productos, 5) fomento al trabajo colaborativo e interdisciplinario y 6) exportación de datos al formato Bibtext.

El uso del modelo facilitará la comparación entre IES con similitudes socio - demográficas y económicas. Como trabajo a futuro, se planea una evaluación sistemática del modelo y su interoperabilidad con herramientas de extracción de indicadores bibliométricos, así como la incorporación de ontologías representadas en lenguajes de la web semántica que apoyen la clasificación temática de los productos de investigación.

## Agradecimientos

Los autores desean expresar su agradecimiento al personal del Departamento de Posgrado de la UPPuebla por su valiosa colaboración y retroalimentación. Este trabajo forma parte del proyecto Biblioteca Digital Semántica de Recursos Educativos. Registro PROMEP /103-5/09/4023.

## Referencias

[1] Carrillo H., Guzmán Sanchez M. V., Sotolongo A. G. (2002), ViBLIOSOM: Visualización de información bibliométrica mediante el mapeo autoorganizado. *Revista española de documentación científica*, ISSN 0210-0614, Vol. 25, N° 4, págs. 477-483.

- [2] De Virgilio R., Giunchiglia F., Tanca L. (2010). *Semantic web information management, a model-based perspective*. Springer.
- [3] Hamilton W. M., Pezo P.A. (2005). Instrumentos de gestión de la ciencia, la tecnología y la innovación. Serie de gestores tecnológicos. CAB, *Ciencia y Tecnología* No. 144
- [4] Osterwalder A., Pigneur Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley and Sons.
- [5] Ryszard S. K., McDaniel B. (2010). *Semantic digital libraries*. Springer. Knowledge Organization, 2004, vol. 31, nº 4, p. 231-244.
- [6] Silverschatz B. (2002). *Fundamentos de bases de datos*. McGraw- Hill. Cuarta edición.
- [7] Valiente F.G. (1997). *Composición de textos científicos con Latex*. Edicions UPC, Universitat Politècnica de Catalunya.
- [8] Witten I.H., Bainbridge D. (2009). , Nichols D.M.. *How to build a digital library*. Morgan Kauffmann. Segunda Edición. [9] Callon, M., Courtial, J. P., Penan, Hervé. *Cienciometría: la medición de la actividad científica, de la bibliometría a la vigilancia tecnológica*. 1995. Editores Trea.

### Páginas web

- [10] Bibexcel. A tool-box developed by Olle Persson, Inforsk, Umeå univ, Sweden. Disponible en: <http://www8.umu.se/inforsk/Bibexcel/>
- [11] DBLP. The DBLP Computer Science Bibliography. Copyright © 1993-2009 Michael Ley (Universität Trier, Informatik, ley@uni-trier.de).
- [12] EndNote (Thomson Reuters EndNote). Advance your research an publish instantly. Copyright © 2011 Thomson Reuters Disponible en: <http://www.endnote.com>
- [13] Sistema de captura del Curriculum Vitae Único del Consejo Nacional de Ciencia y Tecnología. Disponible en: <http://epmwserv.main.conacyt.mx/psp/REGCYT/?cmd=login>
- [14] Sideral. Sistema de gestión científica de la Universidad de Zaragoza está disponible en: <http://www.unizar.es/sideral>
- [15] Sistema de solicitudes en línea de PROMEP. Secretaría de Educación Pública. Disponible en: <http://promep.sep.gob.mx/solicitudesv3/comun/login.php>
- [16] Scopus. Copyright©2011 Elsevier B.V. Disponible en: <http://www.scopus.com/home.url>