



TENDENCIAS TRANSDISCIPLINARIAS EN LOS ESTUDIOS MÉTRICOS DE LA INFORMACIÓN Y SU RELACIÓN CON LA GESTIÓN DE LA INFORMACIÓN Y DEL CONOCIMIENTO

Salvador Gorbea-Portal

Doutor en Documentación por la Universidad Carlos III de Madrid, España.
Profesor de la Universidad Nacional Autónoma de México, México.

E-mail: portal@unam.mx

Resumen

Se perfila una nueva tendencia en los Estudios Métricos de la Información a partir del enfoque transdisciplinario propio de la naturaleza de este tipo de estudio, a través del cual se identifican y definen un conjunto de procesos y metodologías convergentes con este tipo de estudio, conocidas como: *KDD (Knowledge Discovery Database – Descubrimiento de Conocimiento en Bases de Datos)*, la Minería de Datos, de Texto y de la *WWW*, la *Bibliomining* (Bibliominería) y el *OLAP (Online Analytical Processing)* (Procesamiento Analítico en Línea), utilizadas para el descubrimiento de conocimiento en los repositorios de datos (*Data Warehouse*), así como la vinculación de esta convergencia con los Sistemas de Gestión de Información y de Conocimiento, finalmente se propone un modelo en el que se muestran estas relaciones.

Palabras claves: *Transdisciplinariedad. Estudios Métricos de la Información. KDD. Gestión de la Información. Gestión del Conocimiento.*

TRANSDISCIPLINARY TRENDS IN INFORMATION METRICS STUDIES AND ITS RELATION TO THE MANAGEMENT OF INFORMATION AND KNOWLEDGE

Abstract

Transdisciplinary relationships are presented in the Information Metrics Studies from the identification and definition of a set of processes and methodologies converged with this type of study, known as KDD (Knowledge Discovery Databases), Data Mining, Text and WWW, the Bibliomining and OLAP (Online Analytical Processing) used for knowledge discovery in data repositories (Data Warehouse) and linking this convergence with Information Management Systems knowledge and finally we propose a model that shows these relationships.

Keywords: *Transdisciplinarity. Information Metrics Studies. KDD. Information Management. Knowledge Management.*

1 INTRODUCCIÓN

La Metría de la Información, como campo de investigación emergente y de frontera de las Ciencias Bibliotecológica y de la Información, ha tenido un desarrollo acelerado en los últimos tiempos, ello se debe a su naturaleza multi, inter y transdisciplinaria. Esta característica vincula a este tipo de investigación con una muy variada gama de disciplinas en forma directa o indirecta, relación que ya ha sido estudiada en detalle en un trabajo anterior a éste (GORBEA-PORTAL, 2006).

Este carácter acentuado en lo fundamental por la interacción entre contenidos clásicos de la Matemática y la Estadística y las Ciencias Bibliotecológica y de la Información, se ha visto potenciado con el desarrollo de las Ciencias de Cómputo y otros métodos, procesos y tecnologías particulares de éstas, relacionados con el tratamiento, almacenamiento y análisis de los datos acumulados en las organizaciones.

La generación de soportes con formatos que permiten mayores niveles de integración de los datos ha acelerado su almacenamiento y conservación con vistas a su búsqueda y recuperación. Este escenario ha propiciado el interés por conocer, o más bien, por descubrir el conocimiento que subyace al interior de esos grandes cúmulos de datos.

Al principio, el análisis cuantitativo de las bases de datos bibliográficas, por ejemplo, estuvo más orientada a la identificación de regularidades presentes de forma explícita en las variables bibliográficas, y más recientemente su interés se ha centrado en descubrir conocimiento tácito, es decir implícito pero no visible en los datos, con el propósito de identificar nuevos patrones, nunca antes hallados y que aportan nuevo conocimiento, que sin la implementación de una serie de procesos, de técnicas y tecnologías novedosas, no hubiera sido posible descubrir.

Para llegar a este nivel en el procesamiento de los datos ha sido necesario un desarrollo paralelo de la teoría y organización de las bases de datos el cual ha concebido a los datos no sólo como objeto de almacenamiento, búsqueda y recuperación, sino como una fuente incalculable de conocimiento. Para lograr este propósito, estas nuevas teorías sobre la gestión de bases de datos han tenido que desheredar a los datos de su relación con el sistema operacional destinado a su almacenamiento, búsqueda y recuperación para crear otros repositorios de datos de mayor nivel de integración y limpieza, además de presentar una exquisita orientación para su análisis y para las necesidades de sus usuarios. Un ejemplo depurado de este nuevo enfoque en el tratamiento de los datos lo constituyen el surgimiento de los *Data Warehouse* o repositorio de datos integrados.

Esta nueva orientación centra su interés en una metodología conocida como Descubrimiento de Conocimiento en Bases de Datos KDD (sigla de su nombre en inglés) proveniente como ya se indicó del ámbito informático y computacional, pero convergente con los Estudios Métricos de la Información, lo que le ha permitido a este frente de investigación pasar de la identificación de regularidades estadísticas presentes en el Flujo de Información Documental, acumulado en bases de datos bibliográficas, al descubrimiento de conocimiento en los nuevos repositorios de datos

Es precisamente de esta nueva perspectiva de desarrollo que presentan los Estudios Métricos de la Información y de su aportación a la Gestión del Conocimiento el tema tratado en este trabajo, el cual tiene como objetivo: identificar y definir los procesos y metodologías utilizadas en el descubrimiento de conocimiento en los repositorios de datos, así como su relación con las especialidades métricas, con la finalidad de presentar un modelo de relaciones en el que se muestre su carácter transdisciplinario y su vinculación con los Sistemas de Gestión de Información y de Conocimiento.

2 EL ENFOQUE TRANSDISCIPLINARIO

La adopción de este enfoque en el análisis de las tendencias de investigación en los Estudios Métricos de la Información implica no sólo la aceptación del conocimiento que surge en la intercepción o entre las diferentes y especialidades métricas, sino que implica aquello que rebasa el límite de cada especialidad individual y que por lo general le viene dado por conocimiento importado por disciplinas y especialidades de otros sistemas de conocimiento. El objetivo que se persigue con este enfoque es la comprensión del mundo actual, para lo cual uno de los imperativos es la unidad del conocimiento global. En este sentido, Cobo-Suero

considera la transdisciplinariedad, término introducido en 1970 por Jean Piaget, como una forma superior de la interdisciplinariedad, citando a Piaget y Jantsch, señala que es la noción con la que se apunta principalmente a una superación de las fronteras entre las disciplinas (COBO-SUERO, 1996, p. 29).

Entre los autores que han sentado las bases teóricas de este tipo de enfoque se destacan Edgar Morin (2001) y Basarab Nicolescu (2006a; 2006b) este último incursiona en los fundamentos científicos transdisciplinarios al postular tres principios básicos que gravitan su razón de ser. Estos son la existencia de varios niveles de realidad, la lógica del tercero incluido y la complejidad (PÉREZ-MATOS; SETIÉN QUESADA, 2008a).

Uno de los aspectos que caracteriza el enfoque transdisciplinario y que lo distingue de los demás enfoques (multi e interdisciplinario) se asocia a que en este tipo de investigaciones los límites del campo disciplinar estudiado se ven rebasados por contenidos pertenecientes a otras disciplinas o contenidos que no están ni en una ni en otra sino que aparecen como intercepción entre ambas. En este sentido, Martínez-Miguel señaala que en la investigación transdisciplinaria los participantes trascienden las propias disciplinas (o las ven sólo como complementarias) logrando crear un nuevo mapa cognitivo común sobre el problema en cuestión, es decir, llegan a compartir un marco epistémico amplio y una cierta metametodología que les sirven para integrar conceptualmente las diferentes orientaciones de sus análisis: postulados o principios básicos, perspectivas o enfoques, procesos metodológicos, instrumentos conceptuales, etc. Este tipo de investigación es, sobre todo, un ideal muy escasamente alcanzado hasta el momento (MARTÍNEZ-MIGUELEZ, 2009).

Uno de los hitos más significativos que marca el desarrollo del enfoque transdisciplinario se refiere a la celebración, en noviembre de 1994 en el convento de la Arrábida, en Portugal, del Primer Congreso Mundial de Transdisciplinariedad que reunió a más de setenta participantes (en su mayoría franceses, francoparlantes y portugueses con una nutrida representación de psicoanalistas brasileños), entre estos participante se encontraba el propio Edgar Morin. El evento estuvo organizado bajo los auspicios del Centre de Recherches et d'Etudes Transdisciplinaires (CIRET) de París. Al congreso de la Arrábida le siguieron otro en Locarno en 1997 (Towards a Transdisciplinary Evolution of the University) y otro más en Zurich en el 2000 (Transdisciplinary Conference) (ALVARGONZALEZ, 2003). En la región latinoamericana Brasil y Costa Rica han sido escenario de la celebración de otros dos Congresos Internacionales sobre Transdisciplinariedad y Ecoformación el último y IV Congreso se celebra en la Universidad de Costa Rica del 22 al 26 de febrero de 2010.

En el marco del primer Congreso celebrado en Portugal algunos de sus asistentes firmaron la Carta de la transdisciplinariedad en la que se definen como una «comunidad de espíritus transdisciplinarios», además de dejar acotadas en su articulado las bases en las que se sustentan los principales principios de este tipo de enfoque, por ejemplo el carácter complementario de la transdisciplinariedad con los otros tipos de enfoques queda definido en el Artículo 3, que dice a la letra:

La transdisciplinariedad es complementaria al enfoque disciplinario; hace emerger de la confrontación de las disciplinas nuevos datos que las articulan entre sí, y nos ofrece una nueva visión de la naturaleza y de la realidad. La transdisciplinariedad no busca el dominio de muchas disciplinas, sino la apertura de todas las disciplinas a aquellos que las atraviesan y las trascienden. Mientras que en el Artículo 2 de esta carta se alude sobre el reconocimiento de varios niveles de la realidad al declarar que: el reconocimiento de la existencia de diferentes niveles de realidad, regidos por diferentes lógicas, es inherente a la actitud transdisciplinaria. Toda tentativa de reducir la realidad a un solo nivel, regido por una única lógica, no se sitúa en el campo de la transdisciplinariedad (CARTA de la

transdisciplinaria, 1994)

Las Ciencias Bibliotecológica y de la Información no han quedado fuera del análisis bajo este tipo de enfoque. Quizás ello se deba a su marcado carácter social y a que como disciplinas científicas han sido más abiertas a la inter y a la transdisciplinaria que otras ciencias sociales, además a la flexibilidad de su cuerpo teórico de conocimientos, a su fuerte componente teórico-práctico o a la complejidad y carácter de sus propios objetos de estudio, tales como la información, el libro, los documentos, los sistemas de información, los usuarios de la información, entre otros. Aspectos que han sido tratados a profundidad en otros trabajos de Pérez Matos y Setién Quesada (2008b) y Setién Quesada (2009). Estas características presentes en la Ciencias Bibliotecológica y de la Información también le son atribuibles a los Estudios Métricos de la Información, por derivarse de ellas y por compartir los mismos objetos de estudio.

3 LOS ESTUDIOS MÉTRICOS DE LA INFORMACIÓN

Los Estudios Métricos de la Información constituyen un campo multi, inter y transdisciplinario en el cual interactúan diversos métodos y modelos matemáticos y estadísticos con las disciplinas que integran el denominado Sistema de Conocimientos Bibliológico Informativo (Bibliotecología, Bibliografología, Archivología y Ciencia de la Información). A esta interacción se le atribuye el surgimiento de un conjunto de especialidades métricas (Bibliotecometría, Bibliometría, Archivometría e Informetría), las cuales han alcanzado cierta notoriedad, unas más que otras, en la literatura especializada en los últimos veinte años.

La historia y desarrollo de los Estudios Métricos de la Información ya ha sido esbozada en un modelo teórico sobre esta especialidad (GORBEA-PORTAL, 2005). De acuerdo con los antecedentes históricos identificados en ese trabajo, el desarrollo de este tipo de estudio, así como el de las especialidades métricas vinculadas a las disciplinas de las esferas Bibliológico-Informativa y de la Ciencia, se puede dividir en tres etapas, a saber:

- **Etapla pré-disciplinar (1743-1897):** Desarrollo de estudios, de forma aislada, sobre las relaciones de citas, principalmente en la esfera jurídica; análisis cuantitativo sobre producción literaria y compilaciones de datos estadísticos sobre el comportamiento de las bibliotecas y la composición de sus colecciones.
- **Etapla disciplinar (1917-1979):** Denominación y definición de especialidades métricas clásicas que agrupan este tipo de estudio asociado a las disciplinas que le dieron origen.
- **Etapla de desarrollo disciplinar (1979- hasta la fecha):** Consolidación de las especialidades métricas en un cuerpo de conocimiento, surgimiento de nuevas especialidades y denominaciones, desarrollo matemático, terminológico, curricular e investigativo de este campo del saber.

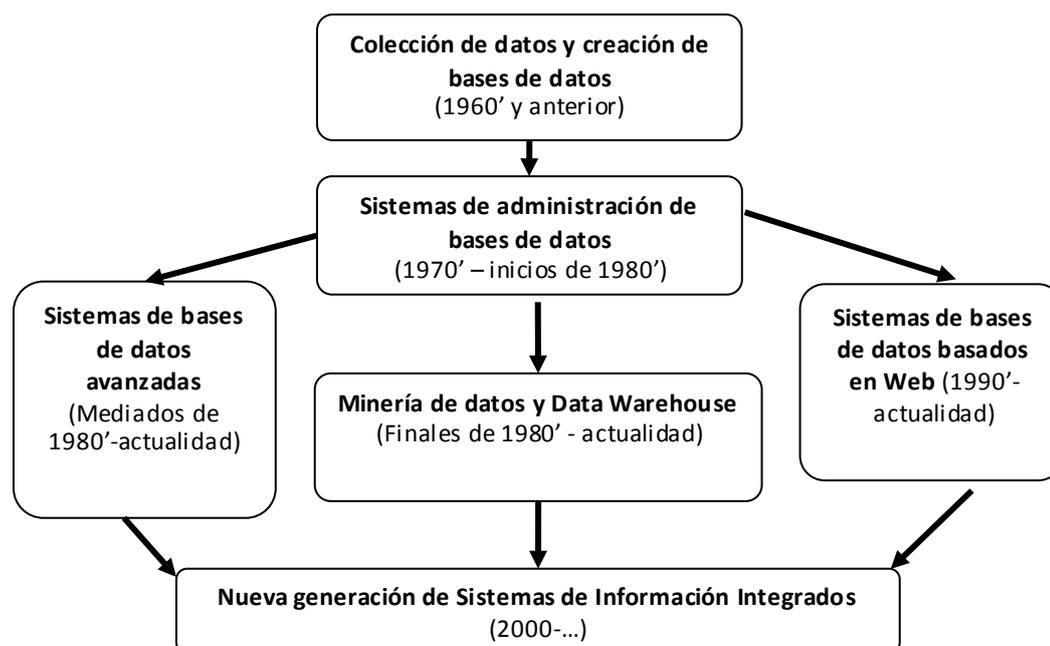
Durante estas tres etapas de desarrollo este tipo de estudio ha puesto a prueba la factibilidad del uso de teorías, métodos, modelos e indicadores cuantitativos en la identificación de las regularidades de la producción y comunicación científica en muy diversas áreas temáticas y regiones, así como las reveladas en los sistemas bibliotecarios y de la información. Estos resultados han aportado información muy valiosa para la toma de decisiones en los Sistemas de Gestión de Información, constituyendo en estos últimos una herramienta novedosa de evaluación en su organización y planificación.

Paralelamente al desarrollo y consolidación de los Estudios Métricos de la Información, en los Sistemas de Gestión de Información también se han experimentado importantes cambios, estos sistemas han transitado de la gestión de conocimiento explícito, registrado en las fuentes de información, a la del conocimiento tácito; este último como el producto de la interacción entre las personas, su inteligencia, sus habilidades y sus experiencias en su relación con el medio ambiente y demás miembros de las organizaciones a las que pertenecen.

4 PROCESOS Y TÉCNICAS CONVERGENTES CON EL ESTUDIO MÉTRICO DE LA INFORMACIÓN

En las últimas décadas, el creciente cúmulo de datos almacenados en las empresas, la industria y otras organizaciones, unido a la necesidad de resolver problemas asociados a la toma de decisiones en materia de mercadotecnia, financiera y de inteligencia competitiva en las mismas, ha propiciado la aparición de una serie de procesos y técnicas orientadas a identificar patrones de comportamiento ocultos en los reservorios de datos. Han y Lamber en su libro de referencia muestran la evolución y desarrollo alcanzado en la tecnología de bases de datos, como consecuencia de lo anterior, y presentan un modelo del cual se muestra a continuación una versión abreviada (HAN; LAMBER, 2001, p. 2).

Figura 1 - Evolución de la tecnología de Bases de Datos



Fuente: Versión abreviada sobre el modelo presentado por Han y Lamber (2001 p. 2)

Estas técnicas y procesos, no obstante a que han surgido en cuerpos de conocimientos tales como el Tratamiento Automatizado de Datos, la Gestión de Bases de Datos y otros campos de la Informática y la Computación, constituyen áreas muy relacionadas pero no incluyentes en las Ciencias Bibliotecológica y de la Información. Sin embargo, este tipo de técnicas y procesos convergen con estas disciplinas a través del Estudio Métrico de la Información, lo cual le ha permitido a este tipo de estudio extender sus líneas de investigación e identificar otras regularidades y patrones de comportamiento antes no estudiadas con las variables de descripción de forma y representación de contenido de los documento, empleando para ello los métodos y herramientas de especialidades métricas ya conocidas como la Bibliometría y la Informetría.

Esta nueva interacción de los Estudios Métricos de la Información Documental ha permitido fortalecer el tema de estudio de la Informetría - definida por Nacke (1979) como la aplicación de métodos y modelos matemáticos a los hechos y situaciones del campo de la información para describir sus leyes y servir de soporte a sus decisiones - ampliando su alcance e incorporando otras técnicas y procesos en el tratamiento cuantitativo de la información, con el propósito de identificar otros patrones de comportamiento antes no identificados con los métodos y modelos matemáticos utilizados tradicionalmente por la Infometría. Entre estas nuevas técnicas y procesos se pueden mencionar: El KDD (*Knowledge Discovery Database*), la Minería de Datos, de Textos, de la Web, la Bibliominería y el Procesamiento Analítico en Línea (*OLAP Online Analytical Processing*), los cuales se definen a continuación.

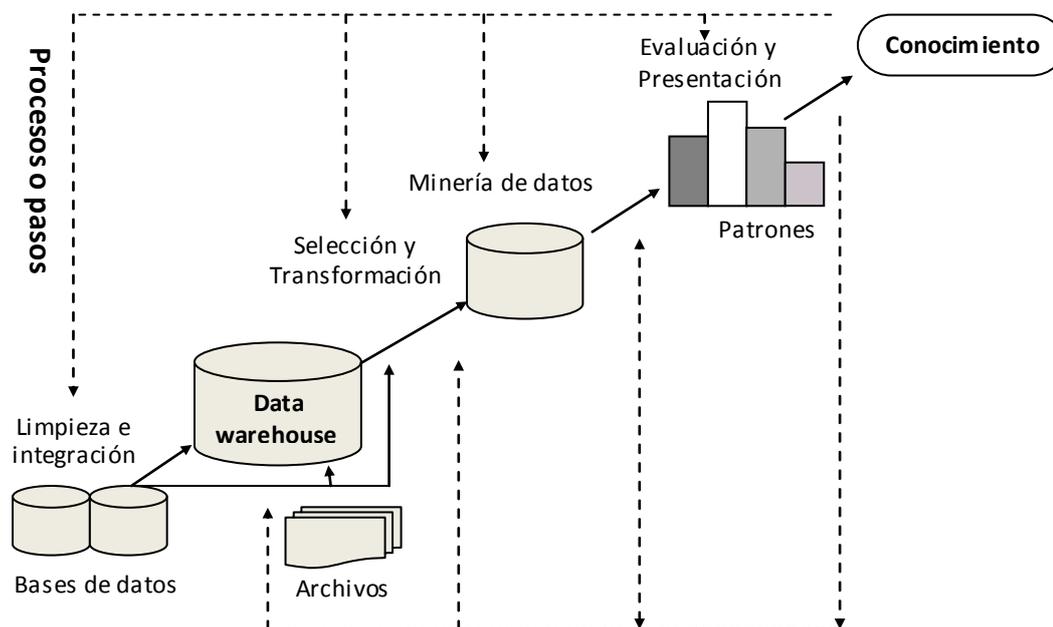
4.1 El KDD (Knowledge Discovery Database – Descubrimiento de Conocimiento en Bases de Datos)

El *Descubrimiento de Conocimiento en Bases de Datos* es una metodología en la que se incluyen un conjunto de procesos introducidos al análisis de las Bases de datos Bibliográficas a finales de los años 80 (SWANSON 1986, 1989 apud HUDOMALI; VIDMAR, 2003, p. 611). Estos autores señalan que dicho proceso ha sido definido por Fayyad et al. (1996) como procesos no comunes de identificación de patrones válidos, novedoso, potencialmente útil y finalmente comprensibles en los datos. Su propia definición como procesos (en plural) se debe a que esta metodología entraña un conjunto de pasos o procesos los cuales han sido definidos por Fayyad, Piatetsky-Shapiro y Smyth (1996), así como por Han y Lamber (2001) y que comprende las operaciones y tareas siguientes, representadas de forma gráfica por estos autores tal y como se muestra en la Figura 2.

- **Limpieza de los datos** (Para remover los datos inconsistentes que adicionan ruido)
- **Integración de los datos** (Combinación de múltiples datos fuentes, en datos de mayor nivel de integración)
- **Transformación de los datos** (Consolidar los datos en formas más apropiadas para ser utilizados en la Minería de Datos)
- **Minería de Datos** (Proceso esencial donde son aplicados métodos inteligentes para extractar patrones de los datos)
- **Evaluación de patrones** (Identificar los patrones que verdaderamente representan el conocimiento basado en algunas mediciones interesantes)
- **Representación del Conocimiento** (Presentar el conocimiento minado al usuario, mediante el uso de técnicas de visualización y representación de conocimiento)

En este conjunto de procesos interviene la creación de una estructura de datos denominado en la literatura especializada como *Data Warehouse*, el cual comprende un repositorio de datos que han sido separados de las bases de datos operacionales de un sistema y contiene una versión limpia, normalizada y anónima de los datos formateados específicamente para su análisis.

Figura 2 - La Minería de datos como uno de los procesos de la metodología de KDD



Fuente: Fayyad et al. (1996, p. 41) y Han y Lamber (2001,p.6)

4.2 La Minería de Datos

La Minería de Datos además de ser un proceso muy importante en la metodología arriba descrita, denominada como KDD, constituye un campo interdisciplinario que emerge de áreas tales como los Sistemas de Bases de Datos, los *Data Warehouse* (repositorios de datos), la Estadística, el Aprendizaje Automático, la Visualización de Datos, la Búsqueda y Recuperación de la Información y de la Computación de alta ejecución, además de otras contribuciones procedente de los Modelos de Redes Neuronales, el Reconocimiento de Patrones, el Análisis de datos Espaciales, entre otras. Constituye una tarea de descubrimiento de patrones interesantes a partir de grandes volúmenes de datos almacenados en Bases de Datos y otros repositorios de datos como los *Data Warehouse*, por ejemplo (HAN; LAMBER, 2001, p. 33).

Otros autores definen a la Minería de Datos como el proceso de descubrimiento de conocimiento sobre repositorios de datos complejos, mediante la extracción oculta y potencialmente útil en forma de patrones globales y relaciones estructurales implícitas entre los datos (KOPANMAKIS; THEODOLUIDIS, 2003 apud INFOVIS, 2006a).

Los métodos de la Minería de Datos permiten descubrir la información oculta en las bases de Datos y transformarlas en un valioso conocimiento tanto retrospectivo (histórico) como prospectivo (proyecciones) o comprensivos (entender lo que ocurre), lo cual resulta muy útil en la toma de decisiones en los gobiernos, empresas u organizaciones, entre éstas últimas las bibliotecarias y de la información (COLLE, 2001, p. 68)

Aunque desde sus inicios sus mayores aportaciones, como ya quedó señalado, estuvieron orientadas a la toma de decisiones en las grandes empresas e industria como la aplicación de algoritmos específicos para extraer patrones en grandes volúmenes de datos, en la actualidad ha encontrado una acertada valía en la actividad científica como parte del proceso de Descubrimiento de Conocimiento, así como en las actividades bibliotecaria y de la información, en la búsqueda de soluciones en el Desarrollo de Colecciones, las Bibliotecas Digitales y las relaciones encontradas en el uso de las colecciones y los servicios, a través de las

consultas realizadas a los servidores, entre otras. Lo anterior puede ser consultado en los trabajos de Guenther (2000), Nicholson (2003a, 2003b), Jonson (2004) y INFOVIS (2006b).

4.2 Minería de Texto

La Minería de Textos es otra tecnología emergente vinculada con el Estudio Métrico de la Información, cuyo objetivo es la búsqueda de conocimiento en grandes colecciones de documentos no estructurados, es decir: consiste en descubrir, a partir de cantidades de textos grandes, el conocimiento que no está literalmente escrito en cualquiera de los documentos (GELBUKH; SIDOROV, 2006, p. 45). Esta aplicación es considerada también por estos autores como un área emergente y muy interesante del Procesamiento de Textos y de la Minería de datos. Además está vinculada con la Informetría de igual forma que algunos Modelo Matemático orientados a identificar regularidades en el comportamiento de las palabras en textos de gran volumen, tales como el de Zipf y el de Heaps (ver un estudio particular de estos dos modelos matemáticos en Gelbukh y Sidorov, 2001)

Estos criterios de establecer una relación de pertenencia de la Minería de Texto con la Minería de Datos puede ser reforzada por una definición anterior de Colle, quien considera que uno de los métodos de la Minería de Datos es, precisamente, la manipulación de datos no estructurados como son los textos, sean archivos de formato TXT Bases de datos de texto completos o imágenes digitalizadas (COLLE, 2001, p. 69)

4.4 La Bibliominería (Bibliomining)

La Bibliominería es un proceso emergente y resultante de la combinación de la Minería de Datos, la Bibliometría, la Estadística y de las herramientas usadas para extraer patrones de comportamiento basados en los recursos digitales de un sistema de biblioteca. Este término (*Bibliomining*) fue usado por primera vez por Nicholson and Stanton en el 2003 como resultado de una discusión de la aplicación de la Minería de datos a las bibliotecas. Aunque en la literatura especializada muchos trabajos que hacen referencia en este contexto al término biblioteca no se refieren a una biblioteca tradicional, sino que están haciendo alusión a las bibliotecas implementadas en determinado *software*, como los sistemas de recursos que acumulan grandes volúmenes de información en las bibliotecas especializadas (NICHOLSON, 2003b, 2004).

De la definición anterior aportada por los propios creadores del término se concibe a la *Bibliominería* como un proceso en el cual se deben considerar los pasos o etapas siguientes:

- Determinación de áreas focales o problemas específicos que requieren de la toma de decisiones en la biblioteca.
- Identificación de las fuentes de datos apropiadas.
- Creación del *Data Warehouse* o repositorios de datos, confeccionado a partir de los datos seleccionados de las bases de datos del sistema y que, como ya se indicó anteriormente, contiene una versión de los datos limpios y reformateados para su análisis.
- Selección de las herramientas apropiadas para el análisis, en este sentido el uso de la estadística multivariada y la inteligencia artificial han resultado ser una buena herramienta en la identificación del comportamiento de patrones no visibles por otros métodos en los datos.

- Descubrimiento de patrones y creación de reportes a través de la Minería de Datos, de acuerdo con lo antes expuesto sobre este concepto anteriormente.
- Análisis e implementación de los resultados.

Estas etapas de la Bibliominería son consideradas por el propio Nicholson como un proceso cíclico por naturaleza, es decir el descubrimiento de patrones de comportamiento en los datos pueden suscitar nuevas interrogantes que echen a andar el proceso de nuevo. (NICHOLSON, 2003, p. 146). Además este autor en su sitio Web ha recalado que estos procesos de la Minería de Datos constituyen una porción de un proceso mucho más amplio, ya definido en líneas anteriores, denominado como el Descubrimiento de Conocimientos en Bases de Datos o KDD¹.

4.5 Minería de la WWW

El creciente desarrollo y diversificación de fuentes y recursos de la Web han convertido a esta importante Red en el mayor repositorio de datos y contenidos de muy variada índole. Lo anterior, unido al auge de herramientas y *software* de Minería de Datos, ha propiciado el interés de expertos por conocer el comportamiento de patrones que revelen la estructura, el uso y las regularidades de esos contenidos.

De esta aplicación de la Minería de datos a los contenidos de la Web ha surgido un nuevo y más reciente proceso denominado como *Webmining* (Minería de la Web) el cual incluye el análisis de los enlaces o link, de los contenidos y los patrones de uso en la Web para mejorar los servicios de los usuarios de este importante medio y que consiste en tres fases: La recolección de los datos, el análisis y el reporte de los resultados (PARTOW-NAVID; SLUSKY, 2004, p. 321). Este nuevo proceso debe ser entendido también como un método convergente con otro campo temático propuesto por Abraham (1996) y retomado por Almind e Igwersen en 1997 denominado como *Webometrics* (Webmetría), definido por estos últimos autores como el estudio de la WWW y de todas las comunicaciones basadas en redes, mediante los métodos informétricos (Almind e Igwersen, 1997). Mientras que Björneborn e Ingwersen, con el propósito de precisar más este término, proponen diferenciar los estudios de la Web de todas las aplicaciones realizadas a la Internet en su conjunto y la definen como: El estudio de los aspectos cuantitativos de la construcción y uso de los recursos de información, estructuras y tecnologías en la WWW, desde el enfoque bibliométrico e informétrico (BJÖRNEBORN; INGWERSEN, 2003 apud GORBEA-PORTAL, 2005, p. 117).

4.6 El OLAP (Online Analytical Processing – Procesamiento Analítico en Línea)

El Procesamiento Analítico en Línea es considerado por algunos autores como una metodología de la Minería de Datos, la cual se centra en obtener de forma rápida y en línea información sobre un conjunto de consultas relativas a combinaciones de variables (COLLE, 2001, p. 71).

Otros autores le atribuyen características comunes con el KDD, debido a que ambos procesos constituyen algoritmos que operan en grandes conjuntos de datos y sus resultados finales tienen como propósito agregar información, previamente desconocida, a la contenida en las bases de datos en línea (HUDOMALJ; VIDMAR, 2003, p. 610). Estos autores en un estudio sobre la aplicación del OLAP a bases de datos bibliográficas señalan que esta

¹ Disponible en: <http://www.bibliominig.com/bibproc.html>

metodología fue introducida por Codd y otros, en 1993 y que el KDD y en especial el OLAP pueden ayudar a las bibliotecas y centros de información a resolver importantes problemas a partir de la observación y análisis de las bases de datos bibliográficas.

De la misma forma que los Estudios Métricos de la Información han venido aportando, desde hace ya varias décadas, resultados valiosos para la toma de decisiones en materia de política científica y de información, así como para la Gestión de la Información (GI), su perspectiva inter y transdisciplinaria en la actualidad, permeable en gran medida por la rápida evolución de los procesos y técnicas emergentes antes definidas, han permitido que la Bibliometría y la Infometría se renueven y aporten resultados también a la Gestión del Conocimiento (GC) en las organizaciones. Es por ello que en el apartado siguiente conviene entonces definir la relación entre estos dos conceptos (GI y GC), para con ello esclarecido identificar las relaciones de los Estudios Métricos de la Información con la GC.

5 LA GESTIÓN DE LA INFORMACIÓN Y DEL CONOCIMIENTO

Diversos han sido los enfoques que en la literatura especializada han tratado las diferencias y semejanzas entre la Gestión de la Información y la del Conocimiento, los cuales han tenido como sustrato el propio debate que subyace entre los conceptos Información y Conocimiento. Un esclarecedor trabajo de Fernández-Marcial (2006, p. 48) analiza y contrapone cada uno de estos conceptos, y al abordar estos últimos denota la pacífica convivencia que se ha dado durante muchos años entre los conceptos Información y Conocimiento.

En este sentido, otro autor Al-hawamdeh (2002, p. 16), citado en el trabajo antes referido, señala al respecto que el conocimiento puede sólo residir en la mente de las personas y en el momento que abandona la mente humana se convierte en información, sin embargo, no todos los tipos de conocimientos pueden ser codificado y capturado, el conocimiento en forma de habilidades y competencias puede ser sólo transferido de persona a persona.

Este comportamiento le permite a este autor introducir en esta discusión dos conceptos:

- El conocimiento tácito, el que muchas personas creen que puede sólo existir en la mente humana [pero que en realidad incluye además las habilidades, competencias y experiencias de los individuos], que constituye un producto de la interacción entre las personas y su relación con su medio ambiente, además de resultar esencial para el desarrollo de una propuesta efectiva para la Gestión de Conocimiento, y
- El conocimiento explícito, es decir el *Know how* y la información codificada y registrada en cualquier tipo de soporte, el cual constituye el objeto de los Sistemas de Gestión de Información.

Las definiciones anteriores contribuyen a establecer, aunque de manera muy somera, la delgada línea que separa a la Gestión de la Información (GI) de la Gestión del Conocimiento (GC). Wilson (2002 apud FERNÁNDEZ-MARCIAL, 2006, p. 46) destaca que en la naturaleza de esta relación es posible identificar tres tendencias fundamentales:

- Una que concibe a la GC como un estadio superior de la GI
- Otra mantiene que la GC es una parte integrante de la GI
- Y una tercera afirma que ambas actividades forman aparte, entre otras, de la Gestión de Contenidos

En todos los casos se admite la diferenciación que existe entre estos dos conceptos, así como las relaciones dialécticas definidas como parte – todo, es decir la GC constituye una parte del proceso más amplio que se define como GI, concebida además como una fase o etapa superior de esta última.

6 LA METRÍA DE LA INFORMACIÓN Y SU APORTACIÓN A LA GESTIÓN DE CONOCIMIENTO

La identificación de regularidades presentes en el Flujo de Información Documental (FID), compilado en Bases de Datos Bibliográficas, tales como la Obsolescencia, la Concentración-dispersión, la Productividad de autores científicos y otras presentes en los sistemas bibliotecarios y de la información, le han permitido a los Estudios Métricos de la Información aportar información valiosa para la toma de decisiones a los Sistemas de Gestión de Información.

La renovación de este tipo de estudio mediante su convergencia con los procesos de Descubrimiento de Conocimiento, antes descrito, abre ante sí una nueva etapa de incalculables resultados orientados en esta ocasión a la toma de decisiones en los Sistemas de Gestión de la Información y otros sistemas, pero fundamentalmente, a los de Gestión de Conocimiento, como una fase superior de gestión en estos tipos de sistemas.

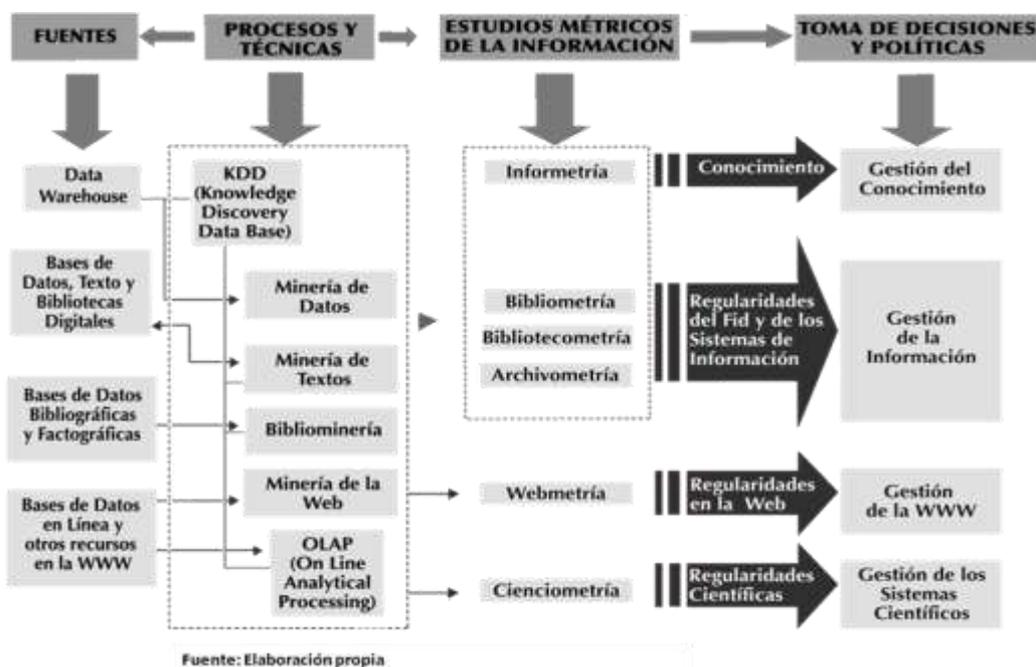
Lo anterior se debe a que los nuevos sistemas de Gestión de Conocimiento, aunque también utilizan información, en la toma de decisiones y sus actividades gerenciales, centran su funcionamiento en la inteligencia, la experiencia, las habilidades o las capacidades cognitivas de los miembros de la organización y de su entorno, así como en el nuevo conocimiento que se encuentra oculto en los datos o que no aparece de forma explícita en las bases de datos que se disponen.

Este nuevo requerimiento que diferencia a los sistemas de Gestión de Conocimiento de los de la Información, según Stern, exige de nuevas técnicas, procesos, tecnologías de Descubrimiento de Conocimiento y hasta de un nuevo cambio de rol entre los bibliotecarios y profesionales de la información (STERN, 2003).

En estas nuevas relaciones, las fuentes que tradicionalmente se venían utilizando en el Estudio Métrico de la Información, tales como bases de datos bibliográficas o factográficas se amplían y diversifican en otras de tipo textuales, Bibliotecas Digitales y en un estado superior de integración de los datos en los denominados *Data Warehouse* y otros recursos y contenidos en la *WWW*. Ello con el propósito de satisfacer las exigencias del proceso de Descubrimiento de Conocimiento y de las técnicas de la Minería de Datos, de texto, de la Web, de la Bibliominería y del OLAP.

En este tipo de estudio, de la selección de la fuente o del tipo de proceso utilizado, así como del fin último que se pretenda con los resultados obtenidos, dependerá su relación con las especialidades métricas en las que se alcance su explicación, además de la orientación que se le den a los resultados, es decir al tipo de Sistema de Gestión que los demande. Las relaciones entre los procesos, especialidades métricas y Sistemas de Gestión hacia el cual se orientan los resultados pueden ser representadas en el modelo de la Figura 3:

Figura 3 - Modelo de relaciones transdisciplinarias entre los Estudios Métricos de la Información y los procesos de Descubrimiento de Conocimiento



Fonte: Autoria propia

Las relaciones identificadas en el modelo anterior indican la afinidad existente entre determinadas fuentes, procesos y especialidades métricas en las que convergen, así como el tipo de resultado obtenido y la orientación de éste al sistema de gestión que lo utiliza en la toma de decisiones y la formulación de políticas.

En este sentido, se observa como fuente principal del KDD y la Minería de Datos a los *Data Warehouse* y su vínculo con la Informetría para la obtención de nuevo conocimiento utilizado en la toma de decisiones de los Sistemas de Gestión de Conocimiento.

La Minería de Texto utiliza las Bases de Datos de Texto y las Bibliotecas Digitales para adicionar valor a los documentos contenidos en estas fuentes, sus resultados pueden estar vinculados con la Informetría y con el resto de especialidades métricas, de igual forma sus resultados podrán estar orientados tanto en los sistemas de Gestión de Información como en los de Conocimiento (WITTEN et. al., 2004).

La Bibliominería por su parte utiliza la Minería de Datos para el análisis de Bases de Datos Bibliográficas y Factográficas y con ello descubre en este tipo de fuente inteligencia y conocimiento que puede ser utilizada en los Sistemas de Gestión de Conocimiento. De igual forma puede identificar regularidades y patrones de comportamiento nunca antes visto en los Flujos de Información Documental y en los Sistemas de Gestión Bibliotecarios y de la Información, como ejemplo de este tipo de estudio pueden ser consultados los trabajos de Nicholson (2003a, 2003b) y el de Quoniam y otros (2001).

Otra relación visible en el modelo es la que se observa entre el proceso de Minería de la Web con las fuentes y recursos disponibles en la *WWW*, con el propósito de identificar los enlaces o *Link* que se presentan entre contenidos, los patrones de comportamiento en el uso que se hace de esos contenidos, además del análisis de dominios que se deriva de los sitios de la Web en los que estén los contenidos, todos estos procesos estudiados bajo el marco

conceptual de la Webmetría, antes definida, permiten identificar regularidades en la WWW, con el propósito de mejorar su acceso y uso.

Por último, el proceso del OLAP asociado al descubrimiento de patrones en las bases de datos en línea permite mejorar el Aprendizaje Automático en este tipo de sistemas, y junto al resto de los procesos de Descubrimiento de Conocimiento aportan conocimiento útil a la Ciencimetría para la toma de decisiones en los Sistemas Científicos.

7 CONSIDERACIONES FINALES

La existencia de un considerable cúmulo de datos almacenados en las empresas, las industrias y las organizaciones ha propiciado un cambio de enfoque referente a los datos, es decir, los datos han dejado de ser sólo objeto de almacenamiento, búsqueda y recuperación de información para convertirse en una fuente de conocimiento e inteligencia, actitud que ha sido favorecida gracias al desarrollo de los *Data Warehouse*.

El Descubrimiento de Conocimiento como el proceso realizado en los grandes reservorios de datos y los pasos que incluye, así como los otros procesos asociados a éste, definidos en este trabajo, constituyen tecnologías emergentes surgidas como consecuencia del desarrollo alcanzado por las nuevas Teorías y Modelos de la Gestión de Bases de Datos.

La convergencia de los procesos de Descubrimiento de Conocimiento con el Estudio Métrico de la Información, propiciado por el enfoque transdisciplinario, ha potenciado el desarrollo de este tipo de estudio, al mismo tiempo que les ha creado nuevas perspectivas en sus relaciones entre diversas disciplinas y especialidades métricas, mediante la aportación de resultados útiles, orientados a la toma de decisiones y formulación de políticas en los Sistemas de Gestión de Conocimiento.

Este cambio de paradigma en el tratamiento métrico de la información, así como las nuevas exigencias que se derivan de los Sistemas de Gestión del Conocimiento, amplían y ponen a prueba el rol de bibliotecarios y especialistas de la información, al tiempo que potencian el uso del enfoque transdisciplinario en las investigaciones desarrolladas en las Ciencias Bibliotecológica y de la Información en general y con particular énfasis en las que se deriven del Estudio Métrico de la Información.

REFERENCIAS

AL-HAWAMDEH, S. Knowledge management: re-thinking information management and facing the challenge of managing tacit knowledge.– **Information research**, vol 8, No.1, p. 1-16, 2002. Disponible en: <http://informationr.net/ir/8-1/paper143.html>. Consultado el: 16 oct. 2012.

ALVARGONZALEZ, D. Transdisciplinariedad El Catoblepas. **Revista Crítica del Presente**, n. 11, p. 12, enero, 2003. Disponible en: <http://www.nodulo.org/ec/2003/n011p12.htm>. Consultado el: 16 oct. 2012.

CARTA de la Transdisciplinariedad. Convento de Arrábida, Portugal, 1994. Disponible en: <http://www.filosofia.org/cod/c1994tra.htm>. Consultado el: 16 oct. 2012.

COLLE, R. **Explotar la información noticiosa**: data mining aplicado a la documentación periodística, Madrid: Universidad Complutense de Madrid: Departamento de Biblioteconomía y Documentación, 2001.

COBO SUERO, J. M. **Interdisciplinariedad y universidad**. Madrid: UPCM, 1986.

FAYYARD, U. G; PIATETSKY-SHAPIO; P. SMYTH. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n.3, p. 37-54, 1996.

FERNÁNDEZ MARCIAL, V. Gestión del Conocimiento vs. Gestión de la Información. **Investigación Bibliotecológica**, v. 20, n. 41, p. 44-62, 2006.

GELBUKH, A.; G. SIDOROV. Zipf and Heaps Laws' Coefficients Depend on Language, Proc. CILing-2001, 2nd International Conference on Intelligent Text Processing and Computational Linguistics, February 18–24, Mexico City. In: **Lecture Notes in Computer Science**, 2004, Springer-Verlag, p. 332–335, 2001.

GELBUKH, A. Y SIDOROV, G. **Procesamiento automático del español con enfoque em recursos léxicos grandes**. México: Instituto Politécnico Nacional, Centro de Investigación en Computación. p. 240, 2006.

GORBEA-PORTAL, S. **Modelo Teórico para el Estudio Métrico de la Información Documental**. Gijón: TREA p. 171, 2005.

_____. Perspectivas interdisciplinarias de los Estudios Métricos de la Información. In: MARTÍNEZ ARELLANO, F. F.; CALVA GONZÁLEZ, J. J. (Comp.) **Memoria del XXIII Coloquio de Investigación Bibliotecológica y de la Información**. Problemas y métodos de investigación en Bibliotecología e Información. Una perspectiva interdisciplinaria. México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas. 2006.

GUENTHER, K. Applying data mining principles to library data collection. **Computers in libraries**, v. 20, n. 4, p. 60-63, 2000.

HAN, J.; M. LAMBER. **Data mining: concepts and techniques**- San Francisco: Morgan Kaufmann Publishers, 2001

HUDOMALJ, E.; G. VIDMAR. OLAP and bibliographic databases. **Scientometrics**, v. 58, n. 3, p. 609-622, 2003.

INFOVIS. Visualización de Información en Documentación. **¿Qué es la minería de datos?**, 2006a. Disponible en: <http://infovis.rivarela.com/index.php?q=meriadadedatos>. Consultado el: 16 oct. 2012.

_____. Visualización de Información en Documentación. **Bibliominig**: una oportunidad para extraer información oculta en nuestras bibliotecas, 2006b. Disponible en: <http://infovis.rivarela.com/index.php?q=bibliominig>. Consultado el: 16 oct. 2012.

JOHNSON, D. A data mining primer and implications for school library media specialists. **Knowledge Quest**, v. 32, n. 5, p. 32-35, 2004.

MARTÍNEZ-MIGUELEZ, M. Transdisciplinarietà y Lógica Dialéctica: Un Enfoque para la Complejidad del Mundo Actual. **Conciencia Activa** vol. 21. No., 1, p. 107-146, 2003 Disponible en: <http://prof.usb.ve/miguelm/transdiscylogica-dialectica.html>. Consultado el: 5 nov. 2012.

MORIN, E. **Los siete saberes necesarios para la educación del futuro**. México: UNESCO, 2001.

NACKE, O. Informetría: un nuevo nombre para una nueva disciplina. Definición, estado de la ciencia y principios de desarrollo. **Revista Española de Documentación Científica**, Madrid, v. 6, n. 3, p. 183-203, 1983.

NICOLESCU, B. Transdisciplinarietà: Pasado, presente y futuro. Primera Parte. **Visión Docente Con-ciencia**, n. 31, p. 15-31, 2006a.

_____. Transdisciplinarietà: Pasado, presente y futuro. Segunda Parte. **Visión Docente Con-**

sciencia, n.32, p. 14-33, 2006b.

NICHOLSON, S. Bibliomining for auntomated collection development in a digital library setting: using data mining to discover web-based scholarly research works. **Journal of the American Society for information science and technology**, v. 54, n. 12, p. 1081-1090, 2003a. Disponíble en: <http://www.scottnicholson.com>. Consultado el: 16 oct. 2012.

_____. The bibliomining process: data warehousing and data mining for library decision making. **Information technology and libraries**, v. 22, n. 4, p.146-151, 2003b.

_____. O processo da bibliomineracao: repositório de dados e mineração de dados para tomada de decisão em bibliotecas = the bibliomining process: data warehousing and data mining for library decision-making. **Transiformação**, v. 16, n. 3, p. 253-261, 2004.

_____. **Bibliomining: data mining for libraries**: What is bibliomining? 2006. Disponíble en: <http://www.bibliomining.com/>. Consultado el: 16 oct. 2012.

PARTOW-NAVID P.; L. SLUSKY. E-Commerce and Data Mining: Integration Issues and Challenges. In: NEMATI, H. R.; BARKO, Ch. D. **Organizational data mining**: leverangng enterprise data resources for optimal performance. Hershey: Idea Group Publishing, 2004.

QUONIAM, L. et al. Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. **Ciência da Informação**, v. 30, n. 2, p. 20-28, 2001.

PÉREZ-MATOS, N. E.; SETIÉN-QUESADA, E. La interdisciplinariedad y la transdisciplinariedad en las ciencias: una mirada a la teoría bibliológico-informativa. **Acimed**, v. 18, n. 4, 2008a. Disponíble en: http://bvs.sld.cu/revistas/aci/vol18_5_08/aci021108.htm. Consultado el: 16 oct. 2012.

_____; _____. Bibliotecología y Ciencia de la Información: enfoque interdisciplinario. **Acimed**. v. 18 n. 5, 2008b. Disponíble en: http://bvs.sld.cu/revistas/aci/vol18_5_08/aci021108.htm. Consultado el: 26 oct. 2012.

SETIÉN-QUESADA, E. Lo multi-inter y transdisciplinario en las ciencias bibliológico informativas. **Hélice**: Revista Venezolana de Ciencias de la Información, Maracaibo, Venezuela, enero-junio, 2009. Disponíble en: <http://www.cidtec.luz.edu.ve/page.php?static=true&module=helice>. Consultado el: 16 oct. 2012.

STERN, D. New knowledge management systems: the implications for data discovery, collection development, and the changing role of the librarian. **Journal of the American Society for Information Science and Technology**, v. 54. n. 12, p. 1138-1140, 2003.

WITTEN, I. H. et al. Text mining in a digital library. **International Journal Digital Libraries**, v. 4, n. 1, p. 56-59, 2004.

Artigo recebido em 19/11/2012 e aceito para publicação em 10/05/2013
