Minería tecnológica para el análisis de oportunidades de publicaciones en la universidad

Tech mining for analysis of publishing opportunities in university environment

Marta Infante[1], Yoel Abreu[2], Mercedes Delgado[3], Olga Infante[4]

[1]miabreu@ind.cujae.edu.cu,
Instituto Superior Politécnico "José Antonio Echeverría", Calle 114 No. 11901 e/ 119 y 127, Marianao (Cuba)

Minería tecnológica para el análisis de oportunidades de publicaciones en la universidad

Abstract
El conocimiento agregado que se puede obtener de la información contenida en la Web hoy es limitado para la universidad cubana, incluso los procesos para transferir estos conocimientos a la investigación, desarrollo y difusión de la actividad científica de la universidad. Existe una determinada porción del contenido total que hoy está disponible en la Web, del cual se puede obtener información valiosa, a través de técnicas de Minería Tecnológica (tech mining). Se extraen oportunidades de publicación de las revistas de corriente principal, las cuales ponen a disposición resúmenes, autores, años de publicación del artículo, nombre de la revista, utilizando técnicas de minería de tecnológica y para ello se recupera información de la web, específicamente de Science Direct, a través de canales RSS. Se comienza a utilizar el software YALE RapidMiner. La aplicación de la minería tecnológica se ha limitado sólo a una revista científica de la WoS, obteniéndose todos los artículos de la revista Information Systems desde el año 2005 hasta la actualidad, recuperando un total de 220 registros. De esa forma se pueden identificar los tópicos más abordados, los investigadores líderes y las tendencias en las investigaciones. La implementación de la minería tecnológica a través de la búsqueda, captura y análisis de la información disponible sobre revistas de la WoS permite la identificación de oportunidades de investigación en el ámbito universitario y con ello un mayor impacto en la I+D en las universidades cubanas.

Keywords: tech mining, oportunidades de publicaciones, minería de texto, vigilancia tecnológica.

Abstract
The added knowledge that can be obtained from the information contained on the Web today is limited to Cuban universities, including the processes for transferring these skills to research, development and dissemination of scientific activity of the university. There is a certain portion of the total content now available on the Web, which you can obtain valuable information, through techniques of Mining Technology (tech mining). We extracted publishing opportunities for mainstream magazines, which make available summaries, authors, year of publication of the article, journal name, using mining techniques and technology to recover this information web, specifically Science Direct, from RSS feeds. The autor start to use the software RapidMiner YALE. Application of mining technology has been limited only to a scientific journal of the WoS, obtaining all the articles in the journal Information Systems from 2005 until today, recovering a total of 220 records. That way you can identify research themes, researchers and leaders in research trends. The implementation of the mining technology using the search, capture and analysis of available information on the WoS journal allows the identification of research opportunities in the university and thus a greater impact on R & D in Cuban universities.

Keywords: tech mining, publishing opportunities, text mining, technology surveillance.

INTRODUCCIÓN

Los resultados de muchos estudios de investigación empírica y teórica (Batista, Sánchez et al. 2003; Lichtenthaler 2003; Morcillo 2003; Ocáriz and Arruza 2004; Cañizares and Vergara 2006; Comai, Tena et al. 2006; Velasco and García 2006; Alpízar Terrero 2007; Bucheli and González 2007; Osorio Rodríguez and Almagro Peñalve 2007; MALVIDO 2008; Pérez-Salmerón 2009; Rey Vázquez 2009) destacan la necesidad que tienen las organizaciones de realizar vigilancia tecnológica (VT) para: desarrollar las tecnologías y la observación de las prácticas para defenderse contra las amenazas y aprovechar las oportunidades derivadas de sus entorno tecnológico. También se reconoce que esta actividad de VT puede desempeñar un papel fundamental en el proceso de planificación de las organizaciones.

En la literatura, muchos términos diferentes se utilizan para describir la adquisición, evaluación y la comunicación de información sobre tecnología, por ejemplo, inteligencia tecnológica, vigilancia tecnológica, previsión tecnológica y la evaluación tecnológica. En este artículo utilizaremos VT. Elegimos este término, ya que permite considerar todos los aspectos de los demás (inteligencia tecnológica, la evaluación de la tecnología y la previsión de la tecnología).

Adicionalmente las fuentes de información son extensas, destacándose que en la WEB, el conocimiento técnico de la humanidad registrado en patentes se incrementa anualmente en aproximadamente unas 600.000 y que en el mundo existen más de 24.000 revistas científicas que anualmente recogen unos dos millones y medio de artículos científicos. La web se ha convertido en un medio para recopilar, procesar, presentar, almacenar y usar información. Conceptualmente, la WEB puede ser vista como una amplia e investigable biblioteca virtual (Yao and Yao 2003; León 2006) dejando de ser objeto de estudio, solo de profesionales de la bibliotecología o la documentación, para convertirse en un componente esencial para las investigaciones. (Cook 2000; O'Hanlon 2002; Kuhlemeier and Hemker 2007; Torricella Morales, Lee Tenorio et al. 2008; Thompson, Lewis et al. 2009; van Deursen and van Dijk 2009; Vasileiadou and Vliegenthart 2009)

Con el advenimiento de la Web y su uso creciente en la ciencia, se ha renovado el interés en si la comunicación a través de la Web está vinculada a una mayor productividad. El uso de la Web se considera una ventaja, que (explícita o implícita) conducen a una mayor productividad: porque facilita el acceso a los recursos y a la información; facilita el intercambio de archivos, datos e ideas creativas; proporciona los medios para procesar, almacenar e intercambiar información –por ejemplo: ofrece una variedad de nuevos mapas, modelos y herramientas para generar una mayor variedad de ideas y conceptos –, lo que aumenta el ritmo de la investigación. Sin embargo, la paradoja está en que, al mismo tiempo, puede reducir la productividad científica. Las estadísticas iniciales mostraron la sobrecarga de información en la Web, lo que, unido a los costos de la llamada curva de aprendizaje para la utilización de la tecnología, o el efecto de distracción por la manera en que encontramos la información, puede dar lugar a la pérdida de productividad (Vasileiadou and Vliegenthart 2009).

Los artículos científicos, serán la fuente de información a utilizar en el presente trabajo. El conocimiento agregado que se puede obtener de la información contenida en esta fuente de información hoy es limitado para la universidad cubana, incluso los procesos para transferir estos conocimientos directamente a la investigación, desarrollo y difusión de la actividad científica de la universidad. La universidad cubana posee limitaciones de capacidades tecnológicas y económicas para la descarga de este tipo de fuente de información. Sin embargo, existe una determinada porción del contenido total que hoy está disponible en la Web, del cual se puede obtener información valiosa, a través de técnicas de Minería Tecnológica (tech mining). En el presente artículo realizaremos un acercamiento a cómo realizar un análisis de oportunidades de publicación para el ámbito universitario cubano desde una perspectiva de Vigilancia Tecnológica.

Antecedentes Teóricos

1. Vigilancia Tecnológica

La importancia del entorno tecnológico de una empresa y la necesidad de un enfoque sistemático de observación del mismo y evaluación de las tecnologías, sin dudas han sido reconocidos en los modelos de gestión moderna (Nosella, Petroni et al. 2008; Penteado and Boutin 2008).

El aumento de la complejidad que ha caracterizado las últimas dos décadas como resultado de los cambios tecnológicos, ha conllevado a que las empresas tengan una mayor dificultad en la interpretación y la gestión de la tecnología como un activo estratégico. En los últimos años, se ha incrementado la atención en la observación y evaluación de tecnologías, tanto a nivel académico como industrial. A pesar de este creciente interés y reconocimiento de la importancia del tema, un enfoque integrado parece necesitarse, sobre todo en lo relacionado con la forma en que deben ser gestionarse los recursos humanos en estos procesos.

Muchos términos se han utilizado en la literatura para indicar la observación de la tecnología y su evaluación: vigilancia tecnológica(Porter and Cunningham 2005; Porter, Alencar et al. 2006; Porter 2009), inteligencia tecnológica(Lichtenthaler 2003; Savioz 2004; WATTS and PORTER 2007; Yoon 2008), la previsión tecnológica(Garud and Ahlstrom 1997; George 2006; Anderson, Daim et al. 2008; Zenobia, Weber et al. 2009), y la evaluación tecnológica(Garud and Ahlstrom 1997; Jolly 2008).

La tabla 1 muestra, que no parece existir consenso en relación con las definiciones o en la elección de una terminología apropiada para indicar el proceso de observación y evaluación de tecnologías. Diferentes términos se han utilizado para indicar similares conceptos, así como diferentes significados han sido adscritos al mismo término, por ejemplo, la definición de EIRMA sobre Vigilancia Tecnológica es muy diferente a la de Porter.(Nosella, Petroni et al. 2008)

Tabla 1. Principales definiciones de Vigilancia tecnológica

Principales definiciones		
Autores	**Términos**	**Definiciones**
Porter, A.L., Rossini, F., Mason, T.W., Banks, J., Roper, T., 1991. Forecasting and Management of Technology. Wiley, USA.	Technology monitoring (Vigilancia tecnológica)	"Explorar el entorno adecuado para la información pertinente" para obtener: "información histórica sobre el desarrollo de la tecnología, la información del estado del arte actual, y/o información que apunta directamente a las perspectivas de futuro"
EIRMA (European Industrial Research Management Association), 1999. Working group 55. Technology monitoring for business success, Edizioni EIRMA.	Technology monitoring (Vigilancia tecnológica)	"Identificación y evaluación de los avances tecnológicos fundamentales para la posición competitiva de la empresa"
AIRI (Associazione Italiana per la Ricerca Industriale), 2002. Il monitoraggio tecnologico, Edizioni AIRI.	Technology monitoring (Vigilancia tecnológica)	"proceso de gestión de la investigación dirigida hacia la identificación y la evaluación de los avances tecnológicos (amenazas y oportunidades) críticos para el posicionamiento competitivo de la empresa"
Ashton, W.B., Bryan, A., Richardson, A., et al., 1997. Keeping Abreast of Science and Technology: Technical Intelligence for Business. Battelle Press, Columbus, Ohio.	Technology intelligence (Inteligencia tecnológica)	"'la información comercial sensible sobre las amenazas u oportunidades científicas o tecnológicas externas, o los acontecimientos que tienen el potencial de afectar la situación competitiva de la empresa"
Lichtenthaler, E., 2003. Third generation management of technology intelligence processes. R&D Management 33 (4), 361–375.	Technology intelligence (Inteligencia tecnológica)	"tarea que es independiente de la forma en que se lleva a cabo, y cuyo objetivo es explotar las oportunidades potenciales y defenderse contra amenazas potenciales, a través de la entrega oportuna de información relevante acerca de las tendencias tecnológicas en el entorno de la empresa"
Vanston, J.H., 2003. Better forecast, better plan, better results. Research Technology Management 47–58 gen-feb.	Technology forecasting (Previsión tecnológica)	"Serie de prácticas técnicas probadas, que proyecta con razonable exactitud la naturaleza, frecuencia, magnitud y consecuencias de los avances futuros en la tecnología"
Bright, J.R., 1978. Practical Technology Forecasting: Concepts and Exercises, third ed. The Industrial management center, Austin, TX.	Technology forecasting (Previsión tecnológica)	"Una predicción cuantificada en el tiempo y del carácter de la medida de cambio en los parámetros técnicos"
Twiss, B., 1992. Forecasting for Technologists and Engineer. Peter Peregrinus Ltd., Stevenage.	Technology forecasting (Previsión tecnológica)	"el medio por el cual un enfoque sistemático puede ser aplicado para obtener una mejor visión del futuro, que es lo suficientemente sólida para dar una adecuada base para la toma de decisiones"
ESTO (European Science and Technology Observatory), 2001. Monitoring of Technology Forecasting Activities. JRC-IPTS, European Commission, Sevilla. ESTO (European Science and technology observatory), 2001. Strategic policy intelligence: current trends, the state of play and perspective. JRC-IPTS, European Commission, Sevilla.	Technology forecasting (Previsión tecnológica)	"seguimiento continuo de los avances tecnológicos que llevan a una identificación temprana de las prometedoras aplicaciones futuras y de una evaluación y validación de su potencial"
Loveridge, D., 1996. Special publication on technology assessment. International Journal of Technology Management 11, 5–6.	Technology assessment (Evaluación tecnológica)	"puede ser descrito como la previsión de los efectos y la retroalimentación con el fin de reducir los costos humanos y sociales de aprender a manejar la tecnología en la sociedad por ensayo y error"
Norma Experimental Francesa: XPX50-053. 1998. Servicios de Vigilancia y servicios de establecimiento de un sistema de Vigilancia. AFNOR.	Inteligencia Económica / Vigilancia tecnológica	"Conjunto de acciones coordinadas de búsqueda, procesamiento y distribución con miras a su explotación, información útil a los actores económicos."
Norma española UNE 166006:2006 EX. Sistema de Vigilancia Tecnológica. (Citada por: Cañozares, J. (2006). "Vigilacia Tecnológica. La última novedad de AENOR en I+D+I." PUZZLE Revista de Inteligencia competitiva. Año 5(22): 32-41.)	Vigilancia Tecnológica	"Proceso organizado, selectivo y sistemático, para captar información del exterior y de la propia organización sobre ciencia y tecnología, seleccionarla, analizarla, difundirla y comunicarla, para convertirla en conocimiento con el fin de tomar decisiones con menor riesgo y poder anticiparse a los cambios"
León López, A. M., O. F. Castellanos Domínguez, et al. (2005). Tendencias actuales en el entendimiento de la vigilancia tecnológica como instrumento de inteligencia en la organización: 15.)	Vigilancia Tecnológica	"actividad de identificar las evoluciones y novedades de la tecnología, tanto en proceso como en producto, con el fin de determinar las oportunidades y amenazas, provenientes del entorno que puedan incidir en el futuro de una organización y sus procesos productivos."
Palop, F. and J. M. Vicente (1999). Vigilancia tecnológica e inteligencia competitiva.	Inteligencia Competitiva / Vigilancia tecnológica	"filtra, interpreta y valoriza la información para permitir a sus usuarios decidir y actuar más eficazmente."

Fuente: Technological change and technology monitoring process: Evidence from four Italian case studies. Anna Nosella, Giorgio Petroni, Rossella Salandra. J. Eng. Technol. Manage. 25 (2008) 321–337. Ampliado por los autores.

Un análisis de las definiciones presentadas en la Tabla 1 pone de relieve algunas diferencias en los diversos términos utilizados y, en particular:

- El concepto de inteligencia tecnológica indica que el proceso de recopilación y utilización de información sobre las tendencias de las tecnologías que tienen el

potencial de afectar la situación competitiva de la empresa, a veces por inteligencia tecnológica se entiende, no sólo como el proceso, sino también como su resultado, es decir, la información valiosa;

- La previsión tecnológica, cuyo objetivo es identificar las predicciones de los cambios tecnológicos, parece tener por objeto la recolección de información tecnológica y procesamiento de la información (Nosella, Petroni et al. 2008).
- La evaluación de la tecnología a largo plazo tiene diferentes significados dependiendo de los objetivos del proceso de evaluación y el contexto en el que se está llevando a cabo. Se utiliza a menudo para referirse al análisis de los objetivos (los costos y beneficios) relacionados con la adquisición de una nueva tecnología, ya que se utiliza en la práctica de gestión, por ejemplo: Hacer una evaluación del patrimonio tecnológico de una empresa;

La vigilancia tecnológica abarca los tres aspectos que se acaban de mencionar, es decir, "puede ser entendida como un proceso, proporcionando información sobre la tecnología (inteligencia), la predicción de la las direcciones que tendrá el cambio tecnológico (previsión) o la evaluación y el potencial de exploración las tecnologías que una empresa debe adoptar (evaluación), pero es más frecuentemente utilizado para indicar el proceso de identificación y evaluación de los avances tecnológicos, fundamentales los que pueden tener un importante impacto en la posición competitiva de una empresa (inteligencia).

2. Tecnologías para obtener información de la Web.

La tecnología RSS (“rich site summary”) es la utilizada por los autores del trabajo para obtener de manera automática, por ejemplo, el historial de artículos de una revista. Esta tecnología provee metadatos (nombre de la revista, base de datos donde se encuentra indizada, autores, resumen, año de publicación) acerca de los artículos de una base de datos que posea este tipo de servicio. Realmente se ha generalizado la utilización de este tipo de servicio en la web y en las bases de datos académicas donde se almacenan los artículos científicos. El costo es completamente gratuito y ahorra tiempo tanto al investigador como a los que prestan servicio de minería tecnológica.

Los canales RSS te permiten agregar noticias y distribuir información de acuerdo con necesidades específicas de un tipo de cliente en particular. Adicionalmente se señala que en nuestros días se han convertido esta tecnología en una proveedora eficiente de noticias, las cuales podrían ser buenas para el intercambio generalizado de información en ciencia, publicaciones tecnológicas y patentes. (Porter and Cunningham 2005)

3. Técnicas de Descubrimiento de Conocimientos (KDD)

3.1 Minería de Datos (Data Mining)

Los proyectos de minería de datos tienen por objetivo extraer información útil a partir de grandes cantidades de datos y se aplican a todos los sectores y en todos los campos. Así existen proyectos de este tipo en sectores tan dispares como el comercio electrónico, la banca, las empresas industriales o la exploración petrolífera.

En (Witten and Frank 2005) se define la minería de datos como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Es decir, la tarea fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos. Para que este proceso sea efectivo debería ser automático o semi-automático (asistido) y el uso de los patrones descubiertos debería ayudar a tomar decisiones más seguras que reporten, por tanto, algún beneficio a la organización.

3.1.1 Bases de datos relacionales

Una base de datos relacional es una colección de relaciones (tablas). Cada tabla consta de un conjunto de atributos (columnas o campos) y puede contener un gran número de tuplas (registros o filas). Cada tupla representa un objeto, el cual se describe a través de los valores de sus atributos y se caracteriza por poseer una clave única o primaria que lo identifica.

Aunque las bases de datos relacionales son la fuente de datos para la mayoría de aplicaciones de minería de datos, muchas técnicas de minería de datos no son capaces de trabajar con toda la base de datos, sino que sólo son capaces de tratar con una sola tabla a la vez. Lógicamente, mediante una consulta (por ejemplo en SQL, en una base de datos relacional tradicional), podemos combinar en una sola tabla o vista minable aquella información de varias tablas que requiramos para cada tarea concreta de minería de datos. Por tanto, la presentación tabular, también llamada atributo-valor, es la más utilizada por las técnicas de minería de datos.

3.1.2 Bases de datos documentales

Las bases de datos documentales contienen descripciones para los objetos (documentos de texto) que pueden ir desde las simples palabras clave a los resúmenes. Estas bases de datos pueden contener documentos no estructurados (como una biblioteca digital de novelas), semi-estructurados (si se puede extraer la información por partes, con índices, etc.) o estructurados (como una base de datos de fichas bibliográficas). Las técnicas de minería de datos pueden utilizarse para obtener asociaciones entre los contenidos, agrupar o clasificar objetos textuales.

3.1.3 Tipos de modelos de Minería de Datos

Modelos predictivos: pretenden estimar valores futuros o desconocidos de variables de interés, que denominamos variables objetivo o dependientes, usando otras variables o campos de la base de datos, a las que nos referiremos como variables independientes o predictivas. Por ejemplo, un modelo predictivo sería aquel que permite estimar la demanda de un nuevo producto en función del gasto en publicidad.

Modelos descriptivos: identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos. Por ejemplo, una agencia de viaje desea identificar grupos de personas con unos mismos gustos, con el objeto de organizar diferentes ofertas para cada grupo y poder así remitirles esta información; para ello analiza los viajes que han realizado sus clientes e infiere un modelo descriptivo que caracteriza estos grupos.

¿A qué tipo de datos puede aplicarse la minería de datos? En principio, ésta puede aplicarse a cualquier tipo de información, siendo las técnicas de minería diferentes para cada una de ellas. Cuando el dato es textual se ha investigado en una nueva corriente que se le ha dado en llamar Minería de Texto.

3.2 Minería de Texto (Text Mining)

El interés y la investigación sobre la Minería de Texto han aumentado, definiéndose como el proceso de extracción de información y conocimiento de los textos. La Minería de Texto analiza documentos. De modo más formal puede definirse del siguiente modo: “La Minería de Texto es el proceso consistente en reunir, organizar y analizar gran cantidad de documentos para proporcionar a los analistas y directivos de la empresa informaciones

sobre temas concretos que sean útiles para la toma de decisiones, descubriendo relaciones entre distintos hechos.

La minería de texto requiere también la previa preparación y almacenaje de los documentos o texto seleccionado. Se propone tareas tales como identificar los temas dominantes en un documento, elaborar índices de documentos, resumir textos de forma automática, clasificar los documentos, etc. Para realizarlas se han desarrollado distintas herramientas.(KOU and Gardarin 2000; Choochart Haruechaiyasak, Prapass Srichaivattana et al. 2002; Ingrid Renz, Andrea Ficzay et al. 2002)

3.3 Minería Tecnológica (Tech Mining)

La minería tecnológica es la observación de la tecnología para detectar y analizar los cambios tecnológicos. Esto es, analizar las direcciones de lo que está ocurriendo ahora y basado en esto que ocurrirá en el futuro considerando el desarrollo de una tecnología en particular. Para realizar esto habrá que compilar y analizar información desde múltiples recursos.(Porter and Cunningham 2005) La tabla 2 muestra varios tipos de análisis tecnológicos que pueden ser realizados utilizando técnicas de Minería tecnológica.

Tabla 2. Posibles análisis tecnológicos a realizar utilizando Minería Tecnológica.

A	Vigilancia Tecnológica	Cataloga, caracteriza e interpreta las actividades de desarrollo tecnológico
B	Inteligencia Tecnológica Competitiva	Encontrar del ambiente externo ¿Quién está haciendo qué?
C	Previsión Tecnológica	Anticiparse a posibles desarrollos futuros en tecnologías particulares
D	Mapeo Tecnológico	Seguir los pasos de la evolución dentro de tecnologías relacionadas y familias de productos.
E	Evaluación Tecnológica	Anticiparse a posibles inentendidos, indirectas y consecuencias fuera de tiempo de un cambio tecnológico en particular.
F	Gestión de los procesos tecnológicos	Brindar información acerca de las tecnologías a los que toman decisiones

Fuente: Porter, A. L. and S. W. Cunningham (2005). Tech mining. Exploiting New Technologies for Competitive Advantage. New Jersey, Wiley-Interscience.

La minería tecnológica es la aplicación de herramientas de minería de texto sobre información científica y tecnológica, manteniendo actualizados los procesos de innovación tecnológica. La minería tecnológica se distingue de la minería de datos y de texto por su énfasis en el conocimiento del dominio de la ciencia y la tecnología para informar sus prácticas.

A continuación se consideran las áreas más significativas a través de las cuales se puede llevar un análisis de Minería Tecnológica.

- Inteligencia: Se rastrea la información externa y se interpreta para servir a un objetivo particular de la organización. Es central en A y B y contribuye al resto.
- Invetigaciones Futuras: Se anticipa a desarrollos futuros, enfatizando en desarrollos tecnológicos. Es central en C,D,E y F.
- Concerniente a los factores contextuales socioeconómicos, cuáles son sus influencias y su afectación en cambios de la tecnología. Es central en E y F e importante en B y G.

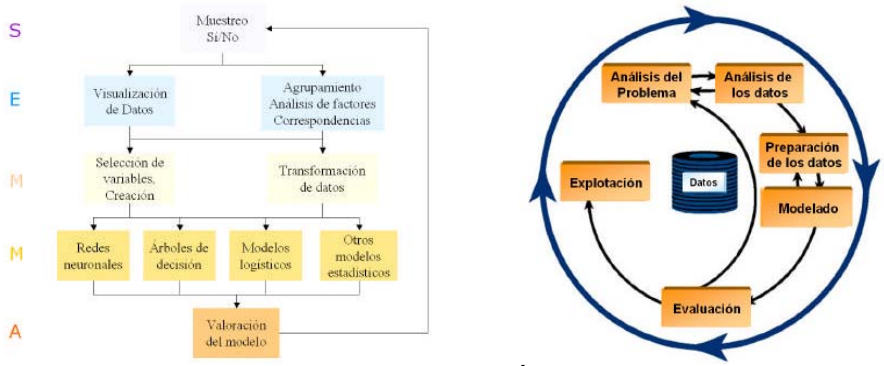
- Análisis de oportunidades: Interpretación de cambios tecnológicos relacionado con las amenazas y oportunidades para nuestra organización. Es vital en A, D y F.
- Consideraciones de procesos: Involucra a los proveedores en acciones determinadas, en este caso en particular concerniente a la tecnología es especialmente saliente en E.

4. Metodologías para el descubrimiento de conocimiento. Un análisis comparativo.

La utilización de las técnicas explicadas en el epígrafe 3 para la extracción de información útil es un proceso complejo, que requiere la aplicación de una metodología estructurada para la utilización ordenada y eficiente de las técnicas y herramientas disponibles. Las principales metodologías utilizadas por los analistas para la realización de proyectos de Minería de Datos son: CRISP-DM y SEMMA. Así SAS Institute propone la utilización de la metodología SEMMA (Sample, Explore, Modify, Model, Assess) y en 1999 un importante consorcio de empresas europeas, NCR (Dinamarca), AG(Alemania), SPSS (Inglaterra) y OHRA (Holanda), unieron sus recursos para el desarrollo de la metodología de libre distribución CRISP-DM (Cross-
Industry Standard Process for *Data Mining*).

Las metodologías *SEMMA* y *CRISP-DM* comparten la misma esencia, estructurando el proyecto de Minería de datos en fases que se encuentran interrelacionadas entre sí, convirtiendo el proceso de Minería de datos en un proceso iterativo e interactivo. La metodología *SEMMA* se centra más en las características técnicas del desarrollo del proceso, mientras que la metodología *CRISP-DM*, mantiene una perspectiva más amplia respecto a los objetivos empresariales del proyecto. (Ver figura 1.) Esta diferencia se establece ya desde la primera fase del proyecto de Minería de datos donde la metodología *SEMMA* comienza realizando un muestreo de datos, mientras que la metodología *CRISP-DM* comienza realizando un análisis del problema empresarial para su transformación en un problema técnico (Rodríguez Montequín, Álvarez Cabal et al. 2002). Desde ese punto de vista más global se puede considerar que la metodología *CRISP-DM* está más cercana al concepto real de proyecto de investigación, por tanto será la utilizada en este trabajo.

Figura 1. Representación gráfica de la metodología SEMMA (a la izquierda del gráfico) y la metodología CRISP (a la derecha del gráfico)

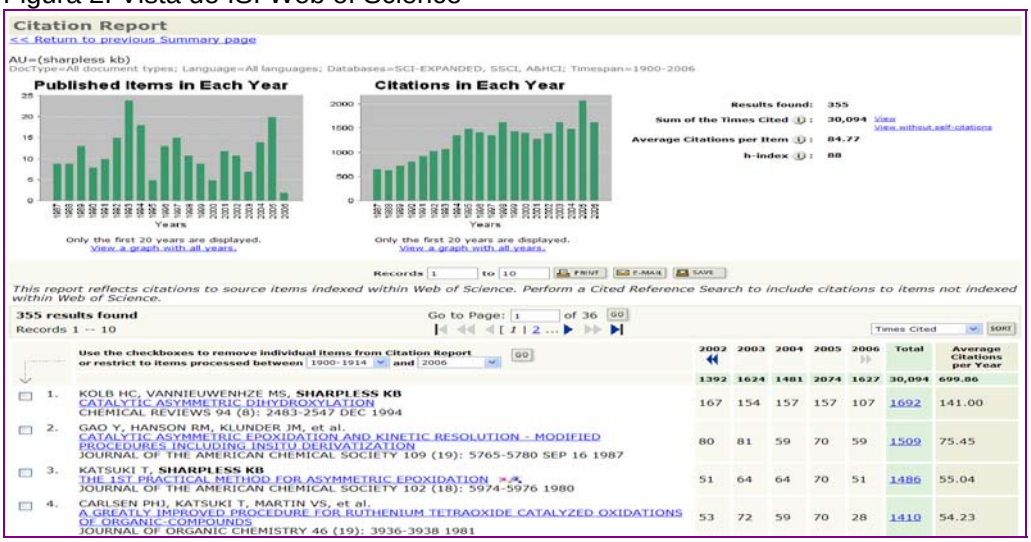


Fuente: Rodríguez Montequín, M. T., J. V. Álvarez Cabal, et al. (2002). Metodologías para la realización de proyectos de Data Mining**:** 12.

Breve descripción de la problemática. Problema a resolver.
Nuestro país posee un ancho de banda limitada debido a las imposiciones del bloqueo económico de los EU hacia Cuba, restricciones a las cuales no está exenta la Educación Superior Cubana. Otra de las afectaciones del bloqueo es que no podemos subscribirnos a

ninguna base de datos (editoriales) de las revistas de corriente principal, su costo es enorme y la mayoría de ellas se encuentran ubicadas en los Estados Unidos. Por lo cual los resultados de las investigaciones que parecen en las revistas de corriente principal son restringidas para nuestro país, igualmente en una buena parte de ellas hay que pagar para publicar un artículo. Las propias editoriales ya han desarrollado técnicas de minería tecnológica, de datos y de texto, para cobrar a los subscriptores la gran cantidad de dinero que exigen. (Ver figura 2)

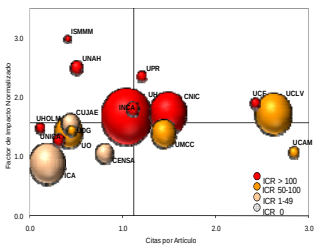
Figura 2. Vista de ISI Web of Science

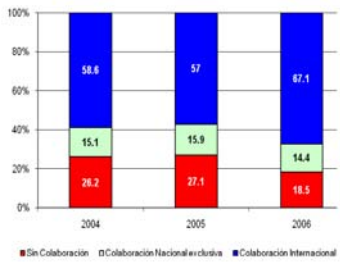


Fuente: ISI Web of Knowledge. La plataforma principal para la investigación académica y científica. Fundación Española para la Ciencia y la Tecnología. Jornada anual de presentación de resultados del Web of Knowledge. 2007.

Algunos estudios caracterizan la producción, impacto y colaboración científica de las instituciones de la Educación Superior Cubana (Arencibia Jorge and de Moya Anegón 2008) analizando dentro de las revistas de corriente principal el comportamiento de estos indicadores. Las revistas de corriente principal sólo muestran un fragmento del Iceberg que constituye la Ciencia nacional; aunque el hecho de resultar el fragmento más visible a nivel internacional, hace de su análisis un importante instrumento para determinar el impacto que la estrategia de internacionalización desarrollada por el MES está teniendo en sus instituciones. (Ver Figura 3.)

Figura 3. Producción, impacto y colaboración científica de las instituciones de la Educación Superior Cubana. (a,b y c respectivamente)





a) b) c)

Fuente: Ricardo Arencibia Jorge y Félix de Moya Anegón. Visibilidad internacional de la Educación Superior cubana en el siglo XXI. Análisis relacional de indicadores de producción, impacto y colaboración científica en el Web de la Ciencia. Universidad 2008.

La producción científica en corriente principal de las instituciones adscritas al MES evidenció un crecimiento gradual durante los primeros seis años del Siglo XXI, y significó durante el período 2004-2006 el 53,3 % de la producción total del país. El 92 % de los trabajos publicados fueron artículos de investigación, y el 96 % fue publicado en idioma inglés. El 37,8 % de los artículos fue citado, y el 13,6 % se dio a conocer en las principales revistas de las correspondientes categorías temáticas del ISI. Nuestra universidad la Cujae como se muestra en la figura 2b) sobre impacto de las instituciones de la Educación Superior Cubana, no posee un privilegiado cuadrante, además en los últimos años la publicación de artículos en revistas de corriente principal ha alcanzado indicadores críticos (20 artículos en el año 2009 en revistas de alto impacto, los años anteriores fue incluso más baja), teniendo en cuenta la masa de doctores que posee nuestro claustro.

Objetivo fundamental del trabajo.

De ahí que nuestra investigación tenga como objetivo "Extraer oportunidades de publicación de las revistas de corriente principal, las cuales ponen a nuestra disposición resúmenes, autores, años de publicación del artículo, nombre de la revista, utilizando técnicas de minería de tecnológica".

Alcance actual

Consideraremos como premisa de esta investigación que las oportunidades de publicación en una revista estarán dadas por las tendencias crecientes en el tiempo de la utilización de un concepto. Por ejemplo:

Si yo realizo una consulta a una base de datos en un año con respecto a un concepto obtengo una frecuencia de aparición de ese concepto. De esta misma manera, si yo obtengo por cada año la frecuencia de ocurrencia de este mismo concepto, las estadísticas de bases de datos prestigiosas como la ISI consideran que posiblemente este concepto ya no sea novedoso y que sus tecnologías hayan llegado al máximo de desarrollo. Por tanto si obtenemos que el concepto ha disminuido en el tiempo nos da una idea de que el concepto o tecnología ha llegado a su límite de desarrollo.(Frey 2007; Horky, Dougherty et al. 2008)

En esta primera aproximación al tema si vemos que el tema o conjunto de conceptos que estamos investigando siguen una tendencia creciente en el tiempo entonces posiblemente tenga éxito en la publicación. Adicionalmente puedo obtener del concepto o grupo de conceptos en los que estoy trabajando los autores que más publican y puedo establecer contacto con ellos.

Preparación de los datos

Recopilación: El primer paso en el proceso de extracción de conocimiento a partir de datos es precisamente reconocer y reunir los datos con los que se va a trabajar.

Las fuentes de información externa con que cuenta el observatorio actualmente son las bases de datos de publicaciones y de patentes fundamentalmente. Muchas de estas bases de datos proveen canales de sindicación de contenidos RSS mediante los cuales se puede obtener la información contenida en ellas.

Las publicaciones se obtienen mediante los canales RSS de ScienceDirect que da acceso libre al título, el resumen, los autores, el año de publicación, la revista, el volumen y el número de los artículos que contiene.

La información que proveen estos canales RSS son recuperados en un repositorio en MySQL, que cuenta con diferentes tablas para almacenar los diferentes tipos de fuentes de información. (Ver figura 4)

Para realizar una primera aproximación experimental se obtienen todos los artículos de la revista Information Systems desde el 2005 hasta la actualidad, recuperando un total de 220 registros.

Figura 4: Tabla de artículos recuperados de ScienceDirect.

id	titulo	autor1	autor2	autoresn	anno	resumen	revista	volumen	numero
1	When security meets software engineering: a case of modelling secure information	Haralambos Mouratidis	Paolo Giorgini	Gordon Manson	2005	(BLOB)	Information Systems	30	8
2	Incremental mining of information interest for personalized web scanning	Rey-Long Liu	Wan-Jung Lin	NULL	2005	(BLOB)	Information Systems	30	8
3	Modelling hypermedia and web applications: the Ariadne Development Method	Paloma DÃ-az	Susana Montero	Ignacio Aedo	2005	(BLOB)	Information Systems	30	8
4	A generic and customizable framework for the design of ETL scenarios	Panos Vassiliadis	Alkis Simitsis	Panos Georgantas ; Manolis	2005	(BLOB)	Information Systems	30	7
5	Information server for highly-connected cross-media publishing	Moira C. Norrie	Beat Signer	NULL	2005	(BLOB)	Information Systems	30	7
6	Management applications of the Web Service Offerings Language (WSOL)	Vladimir Tosic	Bernard Pagurek	Kruti Patel ; Babak Esfandia	2005	(BLOB)	Information Systems	30	7
7	Deliberation in a metadata-based modeling and simulation environment for inter-or	GÃ¼nter Gans	Matthias Jarke	Gerhard Lakemeyer ; Domini	2005	(BLOB)	Information Systems	30	7
8	A methodology and tool support for managing business rules in organisations	Marko Bajec	Marjan Krisper	NULL	2005	(BLOB)	Information Systems	30	6
9	Towards a theory of multimedia metacomputing	Ulrich Marder	Theo HÃ¤rder	NULL	2005	(BLOB)	Information Systems	30	6
10	An adaptive path index for XML data using the query workload	Jun-Ki Min	Chin-Wan Chung	Kyuseok Shim	2005	(BLOB)	Information Systems	30	6
11	DDR: an index method for large time-series datasets	Jiyuan An	Yi-Ping Phoebe Chen	Hanxiong Chen	2005	(BLOB)	Information Systems	30	5
12	Specification and validation of process constraints for flexible workflows	Shazia W. Sadiq	Maria E. Orlowska	Wasim Sadiq	2005	(BLOB)	Information Systems	30	5
13	Using BWW model to evaluate building ontologies in CGs formalism	Ahmad Kayed	Robert M. Colomb	NULL	2005	(BLOB)	Information Systems	30	5
14	An architecture for workflow scheduling under resource allocation constraints	Pinar Senkul	Ismail H. Toroslu	NULL	2005	(BLOB)	Information Systems	30	5
15	YAWL: yet another workflow language	W. M. P. van der Aalst	A. H. M. ter Hofstede	NULL	2005	(BLOB)	Information Systems	30	4
16	Exploiting locality for scalable information retrieval in peer-to-peer networks	D. Zeinalipour-Yazti	Vana Kalogeraki	Dimitrios Gunopulos	2005	(BLOB)	Information Systems	30	4
17	Generating page clippings from web search results using a dynamically terminated	Lin-Chih Chen	Cheng-Jye Luh	Chichang Jou	2005	(BLOB)	Information Systems	30	4
18	Towards practical meta-querying	Jan Van den Bussche	Stijn Vansummeren	Gottfried Vossen	2005	(BLOB)	Information Systems	30	4
19	Neighbor-finding based on space-filling curves	Hue-Ling Chen	Ye-In Chang	NULL	2005	(BLOB)	Information Systems	30	3
20	Sliding window filtering: an efficient method for incremental mining on a time-varia	Chang-Hung Lee	Cheng-Ru Lin	Ming-Syan Chen	2005	(BLOB)	Information Systems	30	3
21	A comprehensive framework for modeling set-based business rules during concept	Sudha Ram	Vijay Khatri	NULL	2005	(BLOB)	Information Systems	30	2
22	Entity identification for heterogeneous database integrationâ€"a multiple classifier	Huimin Zhao	Sudha Ram	NULL	2005	(BLOB)	Information Systems	30	2
23	Decision support queries on a tape-resident data warehouse	Damianos Chatziantoniou	Theodore Johnson	NULL	2005	(BLOB)	Information Systems	30	2
24	A formal approach to lock escalation	Ji-Woong Chang	Kyu-Young Whang	Young-Koo Lee ; Jae-Heon	2005	(BLOB)	Information Systems	30	2
25	Generating domain representations using a relationship model	Irene DÃ-az	Juan Llorens	Gonzalo Genova ; J. Miguel	2005	(BLOB)	Information Systems	30	1
26	Finding the most interesting correlations in a database: how hard can it be?	Christopher Jermaine	NULL	NULL	2005	(BLOB)	Information Systems	30	1
27	Iterative-improvement-based declustering heuristics for multi-disk databases	Mehmet KoyutÃ¼rk	Cevdet Aykanat	NULL	2005	(BLOB)	Information Systems	30	1
28	Database classification for multi-database mining	Xindong Wu	Chengqi Zhang	Shichao Zhang	2005	(BLOB)	Information Systems	30	1
29	A composite approach to automating direct and indirect schema mappingsâ€"†	Li Xu	David W. Embley	NULL	2006	(BLOB)	Information Systems	31	8
30	The MurMur project: Modeling and querying multi-representation spatio-temporal (	C. Parent	S. Spaccapietra	E. ZimÃ¡nyi	2006	(BLOB)	Information Systems	31	8
31	Processing approximate aggregate queries in wireless sensor networksâ€"†	Antonios Deligiannakis	Yannis Kotidis	Nick Roussopoulos	2006	(BLOB)	Information Systems	31	8
32	Spatio-temporal join selectivityâ€"†	Jimeng Sun	Yufei Tao	Dimitris Papadias ; George K	2006	(BLOB)	Information Systems	31	8
33	Evaluating chi-squared automatic interaction detectionâ€"†	Merel van Diepen	Philip Hans Franses	NULL	2006	(BLOB)	Information Systems	31	8
34	Adaptive cycle management in soft real-time disk retrievalâ€"†â€"†â€"†	Youjip Won	Il-Hoon Shin	Kern Koh	2006	(BLOB)	Information Systems	31	8

Fuente: Elaboración propia.

Limpieza y transformación: Se obtienen las palabras clave a partir del resumen del artículo, a través de éstas se aplicarán técnicas de minería tecnológica.

Exploración y selección: Se obtiene la siguiente vista minable la cual posee los campos de título, 1er autor, 2do autor, año y 5 palabras claves o conceptos relevantes. La extracción de palabras claves del resumen del artículo no se obtiene a través de un algoritmo eficiente, careciendo de enfoque semántico avanzado. Se prevee utilizar para perfeccionar esta primera aproximación los algoritmos descritos en el artículo. (KOU and Gardarin 2000; Choochart Haruechaiyasak, Prapass Srichaivattana et al. 2002; Ingrid Renz, Andrea Ficzay et al. 2002)

Figura 2: Vista minable de artículos.

id	titulo	autor1	autor2	anno	palabra1	palabra2	palabra3	palabra4	palabra5
1	When security meets software engineering: a case of modelli	Haralambos Mouratidis	Paolo Giorgini	2005	security	approach	problems	considers	system
2	Incremental mining of information interest for personalized w	Rey-Long Liu	Wan-Jung Lin	2005	interest	comprehensible	specification	scanning	precise
3	Modelling hypermedia and web applications: the Ariadne Dev	Paloma DÃ-az	Susana Montero	2005	hypermedia	mechanisms	software	applications	need
4	A generic and customizable framework for the design of ETL s	Panos Vassiliadis	Alkis Simitsis	2005	etl	data	activity	activities	workflowlike
5	Information server for highly-connected cross-media publishir	Moira C. Norrie	Beat Signer	2005	information	integration	different	media	interactive
6	Management applications of the Web Service Offerings Langu	Vladimir Tosic	Bernard Pagurek	2005	web	service	management	language	constraints
7	Deliberation in a metadata-based modeling and simulation en	GÃ¼nter Gans	Matthias Jarke	2005	networks	dynamic	interorganizational	goals	snet
8	A methodology and tool support for managing business rules	Marko Bajec	Marjan Krisper	2005	business	rules	level	methodology	system
9	Towards a theory of multimedia metacomputing	Ulrich Marder	Theo HÃ¤rder	2005	multimedia	transformation	processing	information	independence
10	An adaptive path index for XML data using the query workloa	Jun-Ki Min	Chin-Wan Chung	2005	xml	indexes	data	paths	label
11	DDR: an index method for large time-series datasets	Jiyuan An	Yi-Ping Phoebe Chen	2005	timeseries	datasets	method	searching	dimensionality
12	Specification and validation of process constraints for flexible	Shazia W. Sadiq	Maria E. Orlowska	2005	workflow	specification	control	constraints	flexibility
13	Using BWW model to evaluate building ontologies in CGs form	Ahmad Kayed	Robert M. Colomb	2005	cgs	construct	ontology	finds	constructs
14	An architecture for workflow scheduling under resource alloca	Pinar Senkul	Ismail H. Toroslu	2005	resource	constraints	allocation	task	resources
15	YAWL: yet another workflow language	W. M. P. van der Aalst	A. H. M. ter Hofstede	2005	workflow	patterns	languages	based	differences
16	Exploiting locality for scalable information retrieval in peer-to-	D. Zeinalipour-Yazti	Vana Kalogeraki	2005	mechanism	efficient	global	retrieval	problem
17	Generating page clippings from web search results using a dy	Lin-Chih Chen	Cheng-Jye Luh	2005	search	genetic	pcs	page	algorithm
18	Towards practical meta-querying	Jan Van den Bussche	Stijn Vansummeren	2005	queries	language	xml	databases	metaquery
19	Neighbor-finding based on space-filling curves	Hue-Ling Chen	Ye-In Chang	2005	data	spatial	curve	curves	proximity
20	Sliding window filtering: an efficient method for incremental m	Chang-Hung Lee	Cheng-Ru Lin	2005	swf	partitions	mining	algorithm	rules
21	A comprehensive framework for modeling set-based business	Sudha Ram	Vijay Khatri	2005	rules	business	constraints	database	framework
22	Entity identification for heterogeneous database integrationâ	Huimin Zhao	Sudha Ram	2005	data	entity	sources	identification	classification
23	Decision support queries on a tape-resident data warehouse	Damianos Chatziantoniou	Theodore Johnson	2005	data	storage	access	tapebased	tertiary
24	A formal approach to lock escalation	Ji-Woong Chang	Kyu-Young Whang	2005	lock	escalation	methods	existing	solution
25	Generatng domain representations using a relationship mode	Irene DÃ-az	Juan Llorens	2005	information	domain	technology	based	applied
26	Finding the most interesting correlations in a database: how h	Christopher Jermaine	NULL	2005	definitions	correlation	goal	problems	paper
27	Iterative-improvement-based declustering heuristics for multi	Mehmet KoyutÃ¼rk	Cevdet Aykanat	2005	distribution	data	query	declustering	method
28	Database classification for multi-database mining	Xindong Wu	Chengqi Zhang	2005	databases	mining	multiple	data	identify
29	A composite approach to automating direct and indirect scher	Li Xu	David W. Embley	2006	schema	matches	mapping	indirect	automating
30	The MurMur project: Modeling and querying multi-representat	C. Parent	S. Spaccapietra	2006	database	diversity	world	management	applications
31	Processing approximate aggregate queries in wireless sensor	Antonios Deligiannakis	Yannis Kotidis	2006	nodes	aggregation	exchanged	sensor	tree
32	Spatio-temporal join selectivityâ€"†	Jimeng Sun	Yufei Tao	2006	join	objects	selectivity	pairs	distance
33	Evaluating chi-squared automatic interaction detectionâ€"†	Merel van Diepen	Philip Hans Franses	2006	chaid	useful	method	bounds	evaluation
34	Adaptive cycle management in soft real-time disk retrievalâ€"†	Youjip Won	Il-Hoon Shin	2006	disk	cycle	realtime	bandwidth	i/o
35	Finding recently frequent itemsets adaptively over online trar	Joong Hyuk Chang	Won Suk Lee	2006	data	stream	elements	information	algorithms

Fuente: Elaboración propia.

Breve descripción de los métodos previstos para la solución de problemas:
A través de esta información de la vista minable, se puede obtener en un análisis inicial utilizando técnicas de visualización:

- ¿Cuáles son los tópicos más abordados?
- ¿Cuáles son los investigadores líderes en un área de investigación?
- ¿Cuáles son las tendencias en la investigación?

De esta manera obtengo oportunidades de investigación para el ámbito universitario.

Trabajo Futuro. Valoración de posibles métodos a ser aplicados alternativamente.

El análisis realizado se limita a una revista científica pero en un futuro cercano se ampliará a todas las revistas de esta base de datos en el dominio de conocimientos específicos de la Facultad de Ingeniería Industrial, como parte de este trabajo se han censado todas las revistas que aparecen en esta base de datos dentro del dominio de conocimiento.

Se conformará una base de datos con la información interna de las investigaciones realizadas por los profesores de la facultad, que nos permita caracterizar las fortalezas internas en cuanto a la investigación científica y poder apuntar a las revistas donde las fortalezas tribute directamente a las oportunidades del ambiente externo.

Ya hemos visualizado una posible dificultad en el futuro cercano y es que la información interna está en idioma español y la externa en su mayoría se encuentra en idioma inglés.

Habrá que considerar de alguna manera las competencias de nuestro claustro sobre todo en cuanto al dominio del idioma inglés para futuras publicaciones en las revistas de alto impacto.

El costo por acceder a estas revistas es elevado, por tanto realizar algún proceso de minería que me pueda ubicar a través de un proceso de clusterización las revistas que más se corresponden con un perfil determinado, disminuirá la posibilidad de una errada elección, esto justifica el estudio.

Se pretende utilizar YALE RapidMiner, para realizar minería. El cual constituye un software potente, de los más utilizados a nivel mundial, con interfaz gráfica muy amigable.

CONCLUSIONES

En este artículo se analiza cómo se recupera información de la web, específicamente de Science Direct, a través de canales RSS.

Los artículos científicos son de especial interés como indicador de investigaciones líderes en un dominio específico.

Los canales o tecnología RSS se tornan muy eficientes para la obtención de la información externa, para el marco de nuestra investigación.

Si en un momento en el mundo el cuello de botella de las investigaciones lo constituía la ausencia de información, hoy la web suple en creces cualquier tipo de necesidad. Sin embargo los volúmenes de información presentes son abismales, su síntesis y presentación constituyen el cuello de botella de nuestros días.

La minería tecnológica se diferencia de la minería de texto y minería de datos por utilizar como fuente de información: bases de datos científicas y tecnológicas.

Las técnicas utilizadas en esta primera aproximación para detectar oportunidades de publicación son limitadas en cuanto alcance, pero serán robustecidas en base a los antecedentes estudiados.

Referencias Bibliográficas

Alpízar Terrero, M. A. (2007). La Vigilancia tecnológica para la actividad de investigación y desarrollo. B. médica. Santiago de Cuba, Centro de Biofísica Médica. Universidad de Oriente.

Anderson, T. R., T. U. Daim, et al. (2008). "Technology forecasting for wireless communication." Technovation **28**(9): 602-614.

Arencibia Jorge, R. and F. de Moya Anegón (2008). Visibilidad internacional de la Educación Superior cubana en el siglo XXI. Análisis relacional de indicadores de producción, impacto y colaboración científica en el Web de la Ciencia. . Universidad 2010. M. d. E. Superior. La Habana, Cuba.

Batista, D. S., M. V. G. Sánchez, et al. (2003). "Establecimiento de un sistema de vigilancia científico-tecnológica." Acimed.

Bucheli, V. A. and F. A. González (2007). "Herramienta informática para vigilancia tecnológica VIGTECH." Revista Avances en Sistemas e Informática **4**(1).

Cañizares, J. and J. C. Vergara (2006). "Vigilancia Tecnológica: La última novedad de AENOR en I+D+I.

La vigilancia Tecnológica antes y después de UNE 166006:2006 Ex." PUZZLE Revista de Inteligencia Competitiva **Año 5**(22): 32-41.

Comai, A., J. Tena, et al. (2006). "Software para la vigilancia tecnológica de patentes: evaluación desde la perspectiva de los usuarios." El profesional de la información **15**(6): 452-458.

Cook, D. (2000). "Collaboration to teach the critical thinking skills needed to become a successful Internet searcher: The planning of a WWW search engine workshop." Research Strategies **17**(2-3): 195-199.

Choochart Haruechaiyasak, Prapass Srichaivattana, et al. (2002). Automatic Thai Keyword Extraction from Categorized Text Corpus**:** 4.

Frey, A. (2007). ISI Web of Knowledge: Contenido de primera calidad, herramientas administrativas con más aplicaciones innovadores de primera clase para ayudarle recuperar los mejores resultados de investigación. Un Nuevo Modo en el campo de la investigación. Jornada anual de presentación de resultados del Web of Knowledge, Madrid, España, Fundación Española para la Ciencia y la Tecnología

Garud, R. and D. Ahlstrom (1997). "Technology assessment: a socio-cognitive perspective." Journal of Engineering and Technology Management **14**(1997): 23.

George, R. P. (2006). SCALING THE TECHNOLOGY OPPORTUNITY ANALYSIS TEXT DATA MINING METHODOLOGY: DATA EXTRACTION, CLEANING, ONLINE ANALYTICAL PROCESSING ANALYSIS, AND REPORTING OF LARGE MULTI-SOURCE DATASETS, Capella University.

Horky, D., J. Dougherty, et al. (2008). WEB Of science and the ISI Web of Knowledge Platform. Creating and Supporting Research Pathways for Student.**:** 44.

Ingrid Renz, Andrea Ficzay, et al. (2002). Keyword Extraction for Text Characterization**:** 7.

Jolly, D. R. (2008). "Chinese vs. European views regarding technology assessment: Convergent or divergent?" Technovation **28**(12): 818-830.

KOU, H. and G. Gardarin (2000). Keywords Extraction, Document Similarity and Categorization**:** 9.

Kuhlemeier, H. and B. Hemker (2007). "The impact of computer use at home on students' Internet skills." Computers & Education **49**(2): 460-480.

León, A., Castellanos, O. y Vargas, F. (2006). "Valoración, selección y pertinencia de herramientas de software utilizadas en vigilancia tecnológica." Revista de Ingeniería e investigación **26**(1): 92-102.

Lichtenthaler, E. (2003). "Third generation management of technology intelligence processes." R and D Management **33**(4): 361-375.

MALVIDO, G. (2008). "La Norma UNE 166006:2006. Vigilancia Tecnológica."

Morcillo, P. (2003). "Vigilancia e inteligencia competitiva: fundamentos e implicaciones." madri+d **17**.

Nosella, A., G. Petroni, et al. (2008). "Technological change and technology monitoring process: Evidence from four Italian case studies." Journal of Engineering and Technology Management **25**(4): 321-337.
O'Hanlon, N. (2002). "Net knowledge: Performance of new college students on an Internet skills proficiency test." The Internet and Higher Education **5**(1): 55-66.
Ocáriz, S. S. d. L. S. d. and M. B. Arruza (2004). "Integración de agentes regionales de innovación y prestación de servicios avanzados de vigilancia tecnológica de inteligencia competitiva para PYMEs: el caso Zaintek." Scire **10**(2): 167-172.
Osorio Rodríguez, M. d. l. A. and O. Almagro Peñalve (2007). "Sistema de Vigilancia Científico Tecnológica en el sector agrícola bajo riego en la República de Cuba." PUZZLE Revista de Inteligencia Competitiva **6**: 5.
Penteado, R. and E. Boutin (2008). Creating Strategic Information for organization with Structured text. Emerging Technologies of TEXT MINING. Techniques and applications. **I:** 34-53.
Pérez-Salmerón, G. (2009). Interinformación. XI JORNADAS ESPAÑOLAS DE DOCUMENTACIÓN, Zaragoza, España.
Porter, A. L. (2009). TECH MINING FOR FUTURE-ORIENTED TECHNOLOGY ANALYSES. Text Mining. A. U. M. Project.
Porter, A. L., M. S. M. Alencar, et al. (2006). "Tech Mining: Multiple Ways to Exploit Science, Technology & Information Resources."
Porter, A. L. and S. W. Cunningham (2005). Tech mining. Exploiting New Technologies for Competitive Advantage. New Jersey, Wiley-Interscience.
Rey Vázquez, L. (2009). Informe APEI sobre vigilancia tecnológica. Informe APEI 4. APEI. Gijón, España**:** 64.
Rodríguez Montequín, M. T., J. V. Álvarez Cabal, et al. (2002). Metodologías para la realización de proyectos de Data Mining**:** 12.
Savioz, P. (2004). Technology Intelligence. Concept Design and Implementation in Technology-based SMEs. Zurich.
Thompson, N., S. Lewis, et al. (2009). "Information literacy skills: Medical radiation science students and the internet." European Journal of Radiography **1**(2): 43-47.
Torricella Morales, R. G., F. Lee Tenorio, et al. (2008). "Infotecnología : la cultura Informacional para el trabajo en la web." from http://revistas.mes.edu.cu.
van Deursen, A. J. A. M. and J. A. G. M. van Dijk (2009). "Using the Internet: Skill related problems in users' online behavior." Interacting with Computers **21**(5-6): 393-402.
Vasileiadou, E. and R. Vliegenthart (2009). "Research productivity in the era of the internet revisited." Research Policy **38**(8): 1260-1268.
Velasco, C. A. B. and C. Q. García. (2006). "Inteligencia competitiva, prospectiva e innovación. La norma UNE-166006 EX sobre el sistema de vigilancia tecnológica."
WATTS, R. J. and A. L. PORTER (2007). "Mining Conference Proceedings for Corporate Technology Knowledge Management." International Journal of Innovation and Technology Management **4**(2): 103-119.
Witten, I. H. and E. Frank (2005). Data Mining: Practical Machine Learning Tools and Techniques., ELSEVIER.
Yao, J. T. and Y. Y. Yao (2003). Web-based Information Retrieval Support System: building research tools for scientists in the new information age. Proceedings of the IEEE/WIC International Conference on Web Intelligence.
Yoon, B. (2008). "On the development of a technology intelligence tool for identifying technology opportunity." Expert Systems with Applications **35**(2008): 11.
Zenobia, B., C. Weber, et al. (2009). "Artificial markets: A review and assessment of a new venue for innovation research." Technovation **29**(5): 338-350.