

# Hacia una metodología de gestión del conocimiento basada en minería de datos

Claudia L. Hernández G\*., María Ximena Dueñas R.\*  
clhernandez@gmail.com, ximenadue@gmail.com

## Resumen

Con el paso del tiempo el conocimiento se ha convertido en el bien que mayor valor tiene para la sociedad y las organizaciones, sin importar el contexto en el cual se desarrolle su actividad económica. Por lo mismo, no sólo se quiere hacer uso de ese conocimiento sino buscar estrategias que permitan gestionarlo, utilizando la tecnología que se tiene a disposición o dando origen a nuevas tendencias tecnológicas. Este artículo pretende establecer la relación que existe actualmente entre la minería de datos y la gestión del conocimiento, haciendo énfasis en la evolución de cada una de estas áreas y cómo al final pueden confluir en un propósito común en el cual el descubrimiento y la gestión del conocimiento se complementan, dando la posibilidad de generar alternativas tecnológicas que mejoren el proceso de toma de decisiones en las empresas.

## Palabras clave:

Descubrimiento de conocimiento (KD), Gestión de conocimiento (KM), Minería de datos (DM), SEMMA, CRISP-DM.

## Abstract

Over time the knowledge has become the greatest good that has value to society and organizations, regardless of the context in which to develop his business. Therefore, not only wants to use that knowledge but to find strategies to manage and much better if it uses technology that is available or if it gives rise to new technological trends. This article seeks to establish the relationship that exists between data mining and knowledge management, emphasizing the evolution of each of these areas and how the end can come together in a common purpose in which the discovery and management knowledge are complementary, giving the ability to generate technological alternatives to improve the decision making process in companies.

## Keywords:

Knowledge discovering (KD), Knowledge management (KM), Data Mining (DM), SEMMA, CRISP-DM

## 1. Introducción

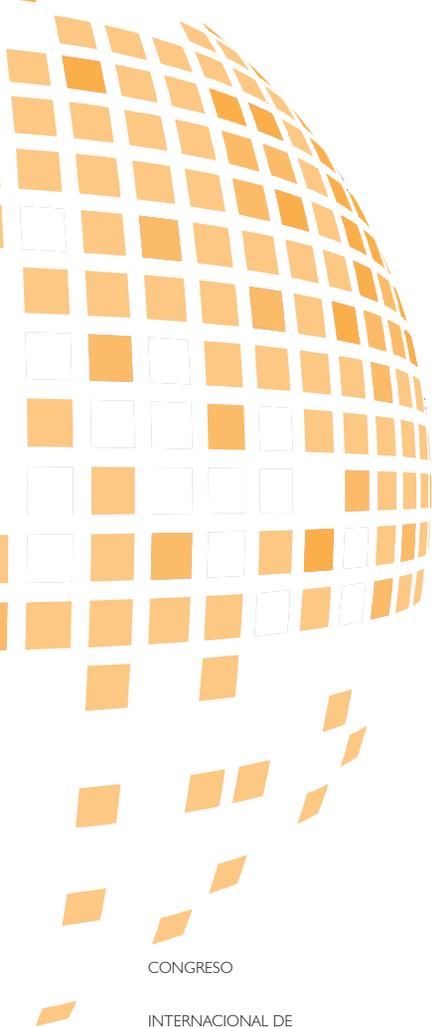
Las tecnologías modernas permiten recolectar información acerca de todo lo que ocurre en una empresa y que pueda ser procesado por un sistema de información. Actualmente, las organizaciones y en especial las que se dedican a la prestación de servicios se están interesando en almacenar datos relaciona-



"ESTRATEGIAS DE  
LA TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"

C. L. Hernández es Ingeniera de Sistemas egresada de la Universidad Nacional de Colombia y candidata a Magister en Ciencias de la Información y las Telecomunicaciones - Universidad Distrital "Francisco José de Caldas". Integrante del Grupo de Investigación en Inteligencia Artificial (IAFT) - Universidad Distrital "Francisco José de Caldas".

M. X. Dueñas es Reyes Ingeniera de Sistemas egresada de la Fundación Universitaria Juan de Castellanos (Boyacá, Colombia) y candidata a Magister en Ciencias de la Información y las Telecomunicaciones - Universidad Distrital "Francisco José de Caldas". Integrante del Grupo Internacional de Investigación en Informática, Comunicaciones y Gestión del Conocimiento (GICOGUE) - Universidad Distrital "Francisco José de Caldas".



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

dos con los movimientos que se realizan en ellas, pero en realidad ¿que se está haciendo con dicha información?, ¿qué beneficios obtiene la empresa con el almacenamiento de la información? En efecto, el propósito principal es analizar los datos, sólo que debido al volumen tan alto las empresas no alcanzarían a cubrir los gastos para sostener un grupo de analistas, sumado esto a que no se podría garantizar con exactitud el tiempo en el cual se obtendrían los resultados ni el grado de certeza de los mismos. Allí es donde se vuelve relevante la minería de datos, que de acuerdo con el Grupo Gartner "La minería de datos es el proceso de descubrir significativas correlaciones, patrones y tendencias a través del análisis de grandes volúmenes de datos almacenados en repositorios de datos usando tecnologías de reconocimiento de patrones tales como técnicas estadísticas o matemáticas" [1], de esta forma es posible hacer uso de técnicas de aprendizaje computacional que permiten establecer un nivel de respuesta y un grado de certidumbre más concreto. Pero ¿qué aprende la organización del análisis de los datos?, ¿cómo enriquece a la organización el resultado de los análisis de datos? Es a través de la gestión del conocimiento como las empresas pueden generar alternativas de solución a situaciones que se evidencien con los resultados arrojados por los procesos de minería de datos.

Este artículo se encuentra organizado en cuatro secciones con las cuales se busca presentar los enfoques asociados con la minería de datos, el descubrimiento y gestión de conocimiento y cómo se complementan con el fin de apoyar el proceso de toma de decisiones en las organizaciones. La primera sección hace referencia a las generalidades y a algunas metodologías utilizadas en la minería de datos, la segunda sección relaciona un marco general de la gestión del conocimiento en la cual se involucran algunos modelos y arquitecturas utilizados, en la tercera sección se realiza la disertación referente a la forma en que la gestión del conocimiento se apoya en la minería de datos para enriquecer la organización y en la última sección se presentan alternativas de trabajo y las conclusiones del análisis en este campo.

## 2. Minería de Datos

### a. Generalidades

Hoy la información es un recurso vital para el desarrollo y evolución de cualquier empresa, en donde la competitividad hace necesaria la obtención de información de una manera rápida y eficiente frente al crecimiento diario de la misma. A raíz de esto, en los últimos años la tecnología ha tenido un crecimiento acelerado, considerándose como una pieza útil y necesaria para el mejoramiento de los procesos y por ende el desarrollo organizacional.

Entre dichas herramientas se encuentra la minería de datos, la cual se ha convertido en un elemento clave para la adquisición de conocimiento dentro de las organizaciones, permitiendo el manejo de altos niveles de efectividad y competitividad en los procesos. Según Fayyad, la minería de datos es "Un proceso no trivial de identificación válida, novedosa, poten-

cialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos” (Fayyad. 1996, p.5). Esta herramienta hace parte de un proceso conocido como KDD (Discovery Knowledge in Databases), el cual es un proceso iterativo e interactivo organizado en torno a cinco fases: Integración y recopilación, preparación de los datos, minería, evaluación, difusión y uso de modelos [2].

El conocimiento adquirido a partir de la minería de datos permite definir asociaciones, patrones o reglas que habían sido desconocidas para los analistas y que al ser descubiertas, permiten realizar modelos de tipo predictivo o descriptivo de acuerdo a las necesidades del negocio [3].

En gran parte el éxito de un proyecto depende de la forma, el cómo y el cuándo se realizan los procesos y es ahí donde las metodologías dependiendo del contexto, brindan la estructura necesaria para un desarrollo secuencial y óptimo de cada uno de ellos. Debido al auge que hoy tiene la minería de datos dentro de las organizaciones, ha sido necesaria la creación de nuevas metodologías que brinden apoyo a las organizaciones que optan por este tipo de herramientas. A raíz de esto, distintos proveedores de este campo se dieron a la tarea de unir esfuerzos y crear una esquematización completa del proceso, surgiendo así dos de las metodologías bandera que hoy existen en el mercado: la metodología CRISP-DM y SEMMA.

### b. Metodología Crisp-Dm

La metodología CRISP-DM (CRoss-Industry Standard Process for Data Mining) fue diseñada en 1999 por un consorcio de empresas europeas (NCR (Dinamarca), AG (Alemania), SPSS (Inglaterra) y OHRA (Holanda)), los cuales unieron sus esfuerzos para crear un esquema que permitiera mostrar el ciclo de vida de un proyecto de minería de datos [4].

CRISP-DM estructura el ciclo de vida en seis fases cada una conformada por una serie de tareas que interactúan entre sí para el desarrollo óptimo del proyecto, como se aprecia en Fig. 1 [5].

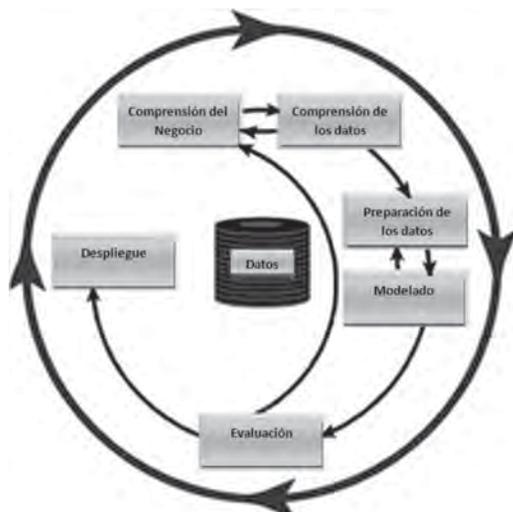
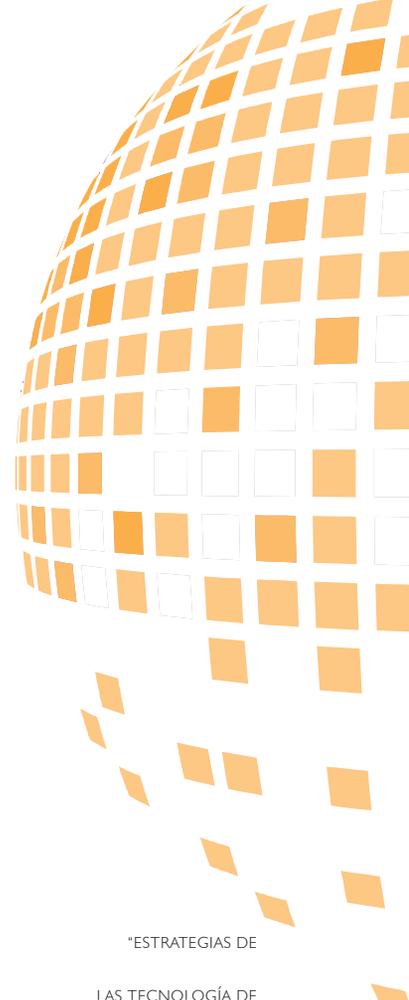
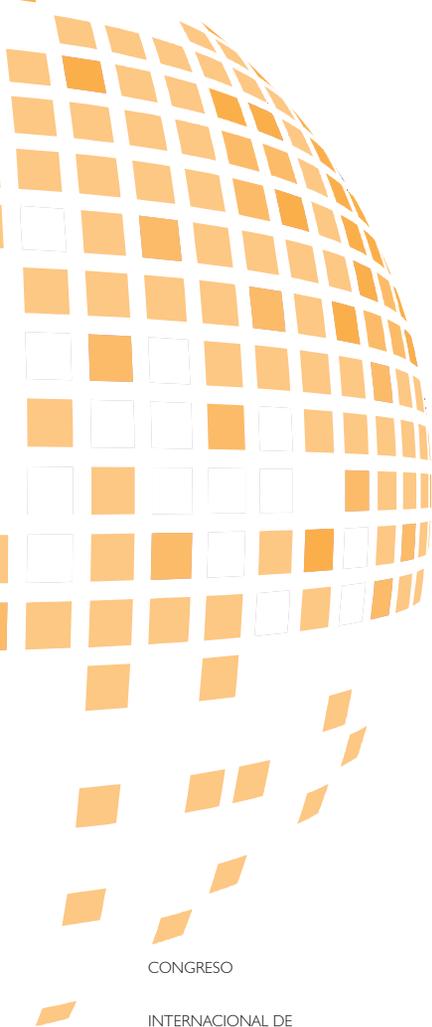


Fig. 1 Fases de la Metodología CRISP-DM.



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

## **Análisis del Problema**

En esta fase se debe conocer a fondo la organización, los objetivos y requerimientos del proyecto de acuerdo con las necesidades del negocio, para esto se debe contar con un plan diseñado preliminarmente para alcanzar los objetivos.

Una de las formas más eficaces de comprender el negocio es involucrando a la gente que hace parte del mismo para así tener una idea general del problema desde distintas áreas de la organización y reformularlo en términos de minería de datos.

## **Análisis de Datos**

En esta etapa se realiza la recolección de datos en donde se inicia la exploración de los mismos, logrando entender el fenómeno e identificando las primeras asociaciones y patrones.

Para realizar una óptima exploración de datos no es suficiente con la visualización es necesario aplicar técnicas estadísticas que permitan la descripción en detalle de las mismas.

## **Preparación de los Datos**

La preparación de datos permite entender el comportamiento de los datos, a través de conceptos estadísticos que permiten describir la variable en estudio. De igual forma, en esta etapa se realizan tareas de preprocesamiento como la reducción de inconsistencias en los datos o datos faltantes.

En esta etapa se seleccionan datos que no pertenecen al mismo espacio de tiempo logrando así modelos más estables de minería en la etapa de diseño.

## **Modelado**

En esta fase se seleccionan las técnicas adecuadas de minería de datos de acuerdo con el contexto del fenómeno. Para la creación de modelos se deben tener en cuenta que los parámetros se basan en la caracterización de los datos.

## **Evaluación**

Esta etapa es la más crítica del proceso, ya que permite verificar su funcionalidad y establecer el funcionamiento de las etapas anteriores. Si el proyecto es evaluado con éxito se realiza la explotación del modelo de lo contrario se verificarán las anteriores etapas.

## **Implantación del modelo**

Esta fase indica la puesta en marcha del proyecto dentro de la organiza-

ción, allí se realizará la debida documentación, presentación de resultados y mantenimiento de la aplicación.

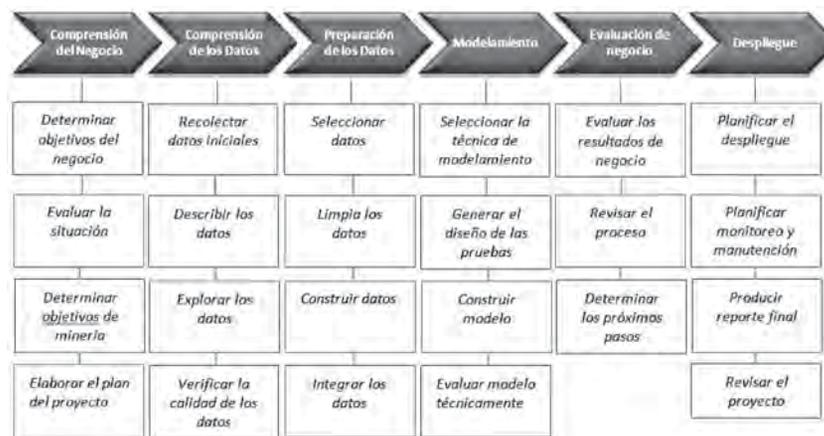


Fig. 2 Metodología CRISP-DM.

### c. Metodología SEMMA

La metodología SEMMA (Sample, Explore, Modify, Model, Assess) fue propuesta por SAS Institute, el cual la define como el proceso de selección, exploración y modelado aplicado a cantidades significativas de datos almacenados que permitan el descubrimiento de patrones como herramientas de apoyo para el negocio. Según SAS más que una metodología para minería de datos, SEMMA es un conjunto de herramientas funcionales enfocadas hacia los aspectos de desarrollo propio de un modelo de minería.

SEMMA se compone de cinco fases representando las etapas de un proyecto de minería de datos.

#### Muestreo

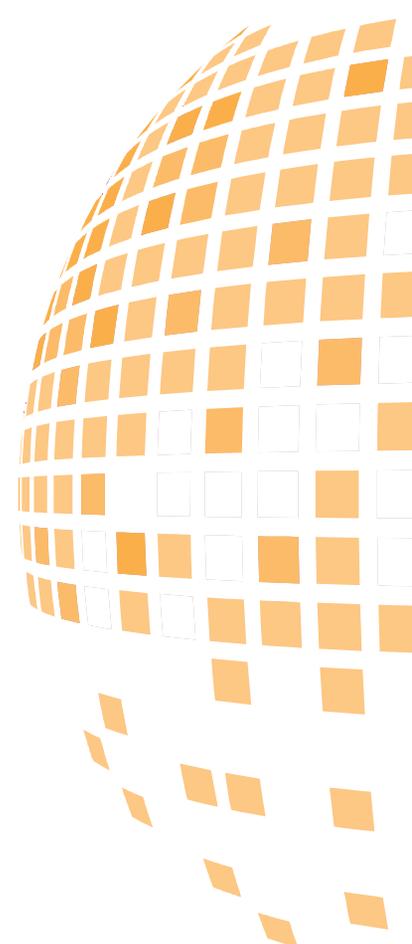
En esta etapa se realiza la extracción de una muestra de los datos que permita representar características comunes de la población para posteriormente comenzar el análisis de los mismos. Con esta fase se logra facilitar los procesos de minado sobre los datos, reduciendo costes y tiempo para la organización.

#### Exploración

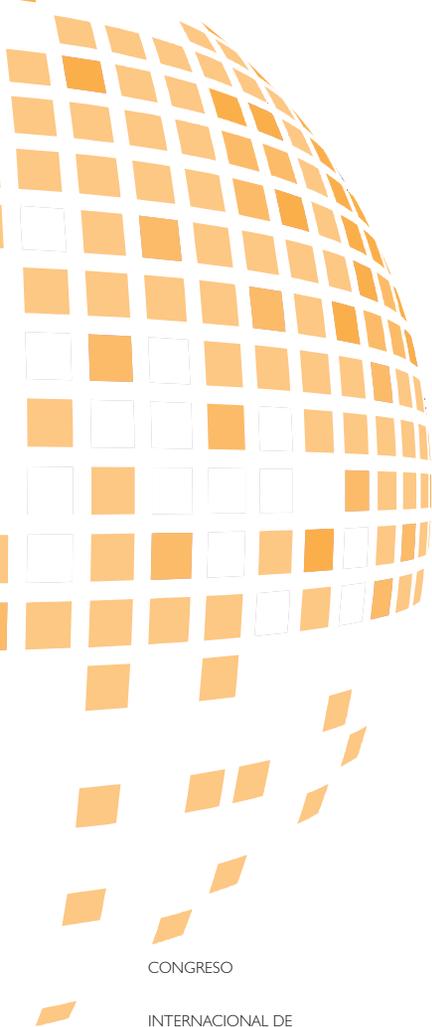
La exploración de datos a través de técnicas estadísticas permite realizar un seguimiento a los mismos logrando detectar, identificar y posteriormente eliminar datos que representen anomalías o deficiencias en las fases siguientes hacia el descubrimiento de información.

#### Modificación

En esta fase se realiza una selección y transformación de los datos de



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

acuerdo con las variables seleccionadas para el proceso de minado, la cual permitirá de acuerdo con éstas adaptar el enfoque de selección y diseño del modelo.

### Modelado

En este punto de la metodología se hace uso de herramientas de software que permitan la utilización de técnicas y métodos propios de la minería de datos, las cuales tiendan hacia el descubrimiento de asociaciones o combinaciones entre los datos, logrando así la predicción de resultados con un alto nivel de confianza.

Entre las técnicas más utilizadas para el modelado de datos, se encuentran: métodos estadísticos, de agrupamiento, redes neuronales, árboles de decisión, lógica difusa, reglas de asociación, entre otros.

### Evaluación

Uno de los pasos principales dentro de una metodología es la valoración de la solución. A partir del modelo obtenido en la fase anterior se realiza una evaluación de resultados para verificar el éxito del proyecto. Una buena práctica para comprobar la validez del modelo es seleccionar otra muestra de datos y aplicarlo para verificación de resultados, si este resulta óptimo se procede con el proceso de producción, en caso contrario se desarrollará otro modelo.



Fig. 3 Metodología SEMMA.

## 3. Gestión de conocimiento

Para iniciar a hablar de la gestión del conocimiento es necesario ahondar en este último término, que a la vez resulta muy complejo de definir sin hacer referencia a términos como datos e información. Estos tres términos pueden definirse de forma independiente sin dejar de lado que están estrechamente relacionados. Adoptando los conceptos que se manejan en la teoría de la información se pueden definir de manera informal. Entonces, se dice que los datos son la materia prima para producir información y que de

acuerdo con la definición en Collins Reference Dictionary, 1992, los datos pueden ser cualitativos o cuantitativos y pueden ser definidos como series de observaciones, medidas o hechos como base para la información. Así mismo, se puede decir que la información puede expresarse a partir de la experiencia adquirida y en cuyo caso una colección de datos adquiere significado. Entonces, es válido también argumentar que la información siempre debe estar asociada a un fenómeno. El conocimiento es, entonces, el concepto que involucra datos e información y habilita a las personas para darle un significado a su entorno. En seguida, se realizará una breve recopilación y representación de manera esquemática a través de los conceptos generales, modelos y arquitecturas que algunos autores han desarrollado con relación a la gestión del conocimiento.

#### a. Generalidades

### Tipos de conocimiento

La clasificación más aceptada en cuanto al conocimiento es la realizada por Nonaka y Takeuchi (1995), quienes lo clasifican en dos tipos:

- Conocimiento tácito. El que poseen las personas, obtenido de la experiencia personal y de un contexto, el cual es de difícil transmisión.
- Conocimiento explícito. Aquel que posee una persona u organización y que es posible codificarlo en alguna manera, comúnmente expresado mediante palabras en escritos, manuales, planos.

Con base en esta clasificación también establecen cuatro maneras de establecer su transmisión:

**Socialización.** Es la comprensión tácita del conocimiento tácito. Un ejemplo sería la transmisión de valores de una organización a otras personas a través de la convivencia.

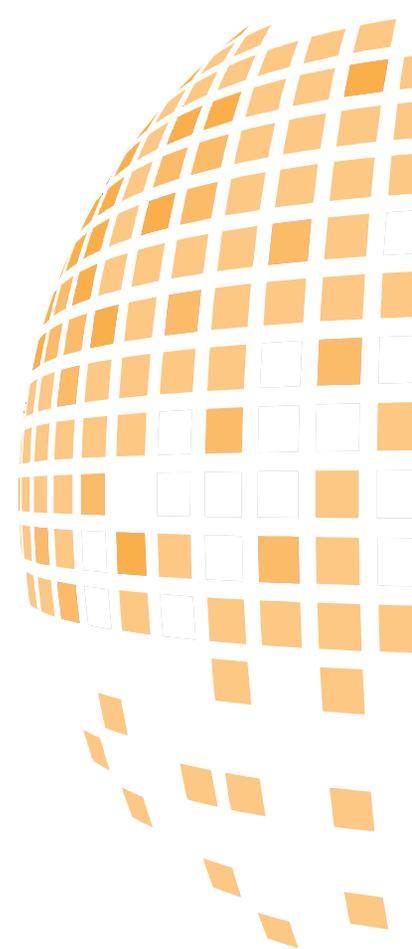
**Exteriorización.** Hacer explícito algún conocimiento tácito a través de algún soporte que permita a los demás conocerlo, como documentos.

**Interiorización.** Pasar de un conocimiento explícito a uno tácito. Es decir, cuando se interioriza algo que se lee u observa.

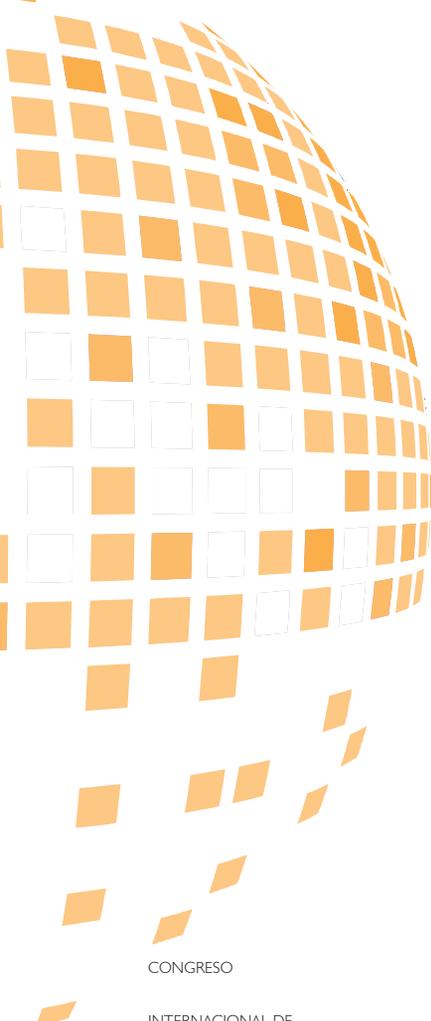
**Combinación.** Paso de un conocimiento explícito a otro explícito. Por ejemplo al complementar algún documento como manuales de procedimientos.

### Capital intelectual

Este concepto nace de la importancia que han adquirido dentro de las organizaciones las personas y su interacción con el entorno. El capital intelectual está formado por las ideas y no es más sino el conocimiento que se encuentra a disposición de la empresa convirtiéndose en un bien de un valor comercial específico. Se pueden diferenciar tres tipos de capital



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

intelectual: capital humano, capital estructural y capital cliente cuya composición se esquematiza en la Fig. 4.

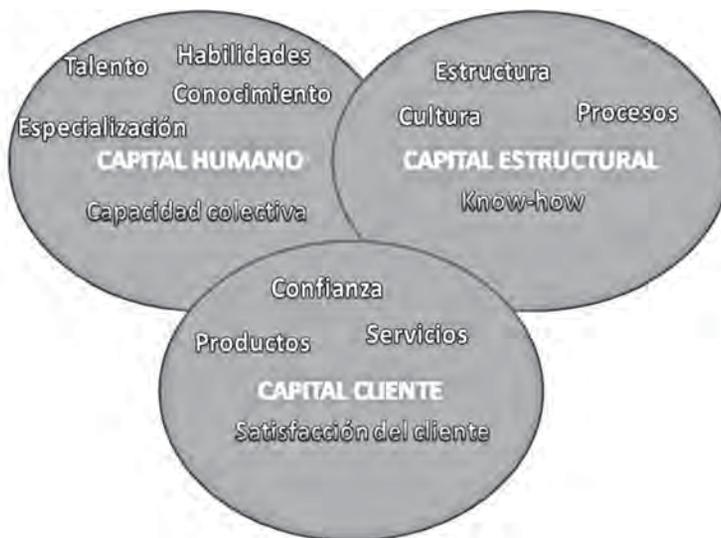


Fig. 4. Capital Intelectual.

### ¿Qué es la gestión del conocimiento?

Diferentes autores han hecho referencia a este concepto con base en la importancia que ha adquirido para las organizaciones el hecho de que guarden historia de todos los procesos y hacer del conocimiento un bien que adquiera valor comercial. Según Malhotra “la gestión del conocimiento es la combinación de sinergia entre datos, información, sistemas de información y la capacidad creativa e innovadora de seres humanos”. Otro de los autores con autoridad en esta temática, Karl Sveiby, definió la gestión del conocimiento como “el arte de crear valor a partir de los activos intangibles de una organización”.

En conclusión, la gestión del conocimiento es un proceso complejo compuesto de una serie de grupos de tareas divididas en patrones de tareas que son ejecutadas por agentes a través de ciclos de decisión compuestos por planeación, acción, monitoreo y evaluación de actividades. Así es como la gestión del conocimiento se ha convertido en un nuevo enfoque en el cual se enfrentan problemas relacionados con la forma de organización y distribución del conocimiento.

#### b. Modelos y arquitecturas

##### **Modelo tácito-explicito (Nonaka, I. and Takeuchi, H., 1995)**

Un sistema muy completo para la gestión del conocimiento debería soportar todas las transformaciones y componentes del ciclo de conversión. Un modelo de esta naturaleza es el expresado en la Fig. 5.



Fig. 5 Modelo tácito-explicito.

En la 0 se describe el modelo de los cinco estados en la gestión del conocimiento que incluyen algunas consideraciones críticas para garantizar la efectividad.

La captura a partir de los empleados y los aliados de la organización es una etapa crítica para generar la base del conocimiento. Para que en el almacenamiento se pueda establecer un orden para los datos es necesario definir un modelo de datos. La interpretación de la información es una de las tareas más sofisticadas porque se requiere que la información se pueda relacionar de manera coherente y esto se logra aplicando diferentes técnicas para esto. La difusión del conocimiento es una etapa que basa su importancia en las personas que forman parte de la organización y se encuentran comprometidas con ella. Una de las mejoras que puede realizarse con el período de auditoría es someter a evaluación la información recolectada y las relaciones existentes en el modelo y la red en la cual está almacenada.

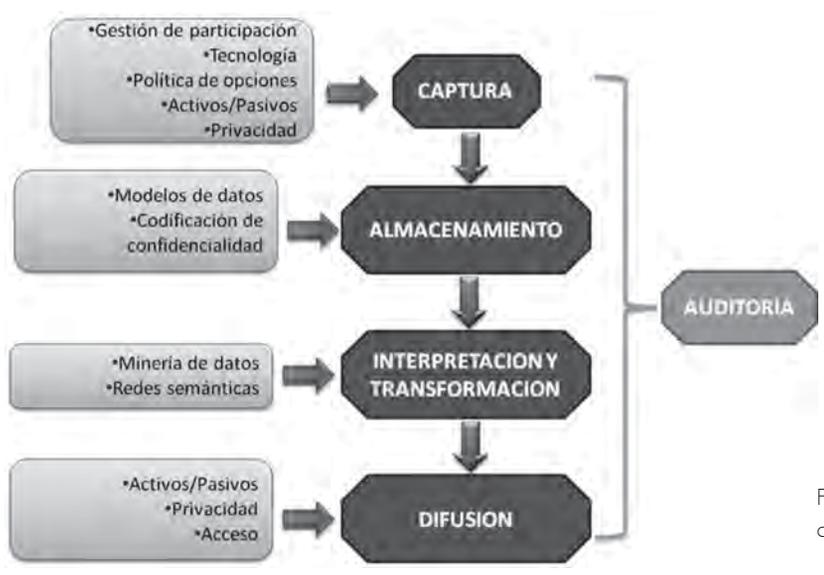
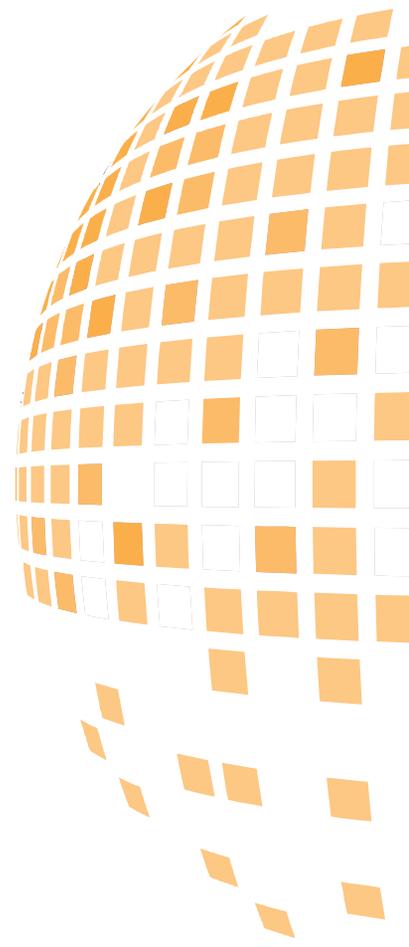
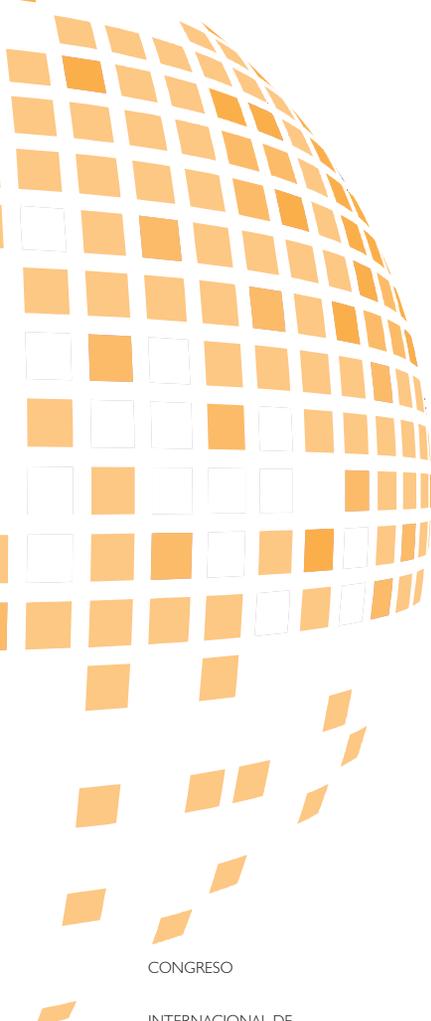


Fig. 6 Modelo de los cinco estados.



"ESTRATEGIAS DE LAS TECNOLOGÍA DE LA INFORMACIÓN Y COMUNICACIÓN EN EL CONTEXTO DE LA CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

## Arquitectura modelo para gestión del conocimiento

Una de las arquitecturas que pueden soportar la tecnología necesaria para la gestión del conocimiento está representada en la Fig. 7, en la cual se puede apreciar la combinación de tecnologías que permiten realizar autorización, indexación, clasificación, almacenamiento, contextualización y recuperación de la información. Las capas inferiores son las que hacen referencia al conocimiento explícito, el cual reside en repositorios como documentos y otros tipos de registros. A través de los portales de conocimiento pueden distribuirse diferentes aplicaciones a diferentes usuarios.



Fig. 7. Modelo de Arquitectura KM. Fuente: [17]

### c. Descubrimiento y gestión del conocimiento

La tecnología y la gestión del conocimiento

En la actualidad, las organizaciones giran en torno a una revolución de información en donde cada día los datos crecen rápidamente y los procesos deben ser cada vez más ágiles y eficientes capaces de manipularlos y adaptarse a las necesidades propias de la empresa. La obtención, manejo de información y posteriormente el descubrimiento de conocimiento son sinónimo de un alto nivel de competitividad en el mercado [20], como lo ha mencionado Ruben Sinisterra “quien tiene el conocimiento tiene el poder”, durante la conferencia sobre patentes, propiedad intelectual y transferencia de tecnología.

Debido a la gran relevancia del conocimiento, surgió una disciplina capaz de administrarlo, distribuirlo y usarlo; la gestión del conocimiento. En párrafos anteriores se ha mostrado su relevancia dentro de los contextos organizacionales, pero ¿cómo este campo puede ser tan ágil y eficiente?

La gestión debe comenzar por la creación de una cultura organizacional en donde el conocimiento debe ser de uso compartido al igual que un alto sentido de lealtad y compromiso con la empresa. Cuando todas las personas trabajan por un mismo objetivo, los procesos se tornan más eficientes mejorando los niveles de productividad y competitividad. De la mano a la gestión, la tecnología desempeña un papel importante ya que permite almacenar, manipular y distribuir la información que a diario se maneja dentro de la organización mediante herramientas automatizadas.

Las tecnologías de información (TI) apoyan el proceso de gestión, brindando la posibilidad de mejorar los procesos y por ende la información y el conocimiento fluyen en forma eficientemente [21]. Estas herramientas han sido conceptualizadas como la integración entre la computación, las comunicaciones y el procesamiento de datos, permitiendo una revolución informática en los diversos campos de la sociedad.

Las TI permiten ayudar a mejorar la productividad de las diversas funciones de la empresa, así como mejorar el flujo interno y externo entre las unidades de negocio. Una óptima utilización de estas brinda un valor agregado a las organizaciones [21].

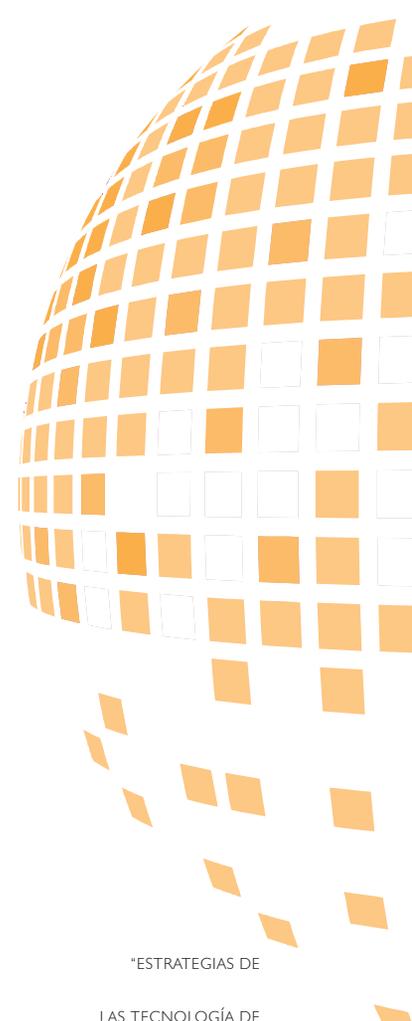
¿Cómo se complementan el descubrimiento y la gestión del conocimiento?

De acuerdo con la experiencia que existe en las diferentes organizaciones y el enfoque de razonamiento basado en casos, ampliamente utilizado en las empresas, se comparten metas con la gestión del conocimiento en el uso y el desarrollo del conocimiento [22]. La idea básica del razonamiento basado en casos es solucionar un nuevo problema con base en una solución de un problema particular que ha sido contextualizada y reutilizada [22].

La gestión del conocimiento es un elemento clave en la extracción del valor, cuando puede darse un uso efectivo del conocimiento [22]. En diversos enfoques la adquisición de la información resulta ser una de las tareas más significativas y haciendo una buena administración en esta etapa puede ser posible reducir tiempo y coste en todo el proceso de minería de datos. La gestión de conocimiento de manera inherente puede enfrentar el cuello de botella en la adquisición de la información, la dificultad para realizar el modelamiento y formalizar el conocimiento relevante para un dominio específico a lo cual la respuesta que actualmente se está dando, involucra una posible solución con tecnología de extracción de la información [24]. Esta tecnología presenta inconvenientes en el manejo de la calidad de los datos ya que dependiendo del problema que se quiera resolver.

#### 4. Metodología propuesta de gestión de conocimiento apoyado en la Minería de Datos

Una relación entre la minería de datos y la gestión del conocimiento hace referencia a que las dos disciplinas apoyan el proceso de toma de



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

decisiones dentro de las organizaciones de tal forma que se busca “llevar el conocimiento correcto a la persona correcta en el momento correcto y en la forma correcta, con el fin de que ésta pueda comprenderla y así tomar las mejores decisiones” [13].

Dentro de la investigación de la relación entre la minería de datos y la gestión del conocimiento, ha sido importante considerar los diversos enfoques teóricos y prácticos, de tal forma que se pueda argumentar la integración y complemento entre ellas. La ventaja de la integración es que el conocimiento está cambiando continuamente y puede ser la base para la reconstrucción del proceso de toma de decisiones [22].

Entonces, se puede decir que la minería de datos es un campo de tipo interdisciplinario que tiene mucha base en la estadística, el aprendizaje computacional y las bases de datos. La motivación principal es el fenómeno del crecimiento de los datos en todos los campos del conocimiento y la necesidad de extraer, validar y usar la información desde los datos recolectados [25].

Teniendo en cuenta la importancia de la minería de datos como herramienta en la búsqueda de información y la gestión de conocimiento como pieza clave para mejorar este recurso [2], se propone una metodología basada en la combinación CRISP-DM, el modelo de cinco estados y el modelo tácito-explicito dentro del contexto de gestión de conocimiento (ver Fig. 8).



Fig. 8 Metodología propuesta GC-CRISP. Fuente: Autores investigadores

La metodología propuesta GC-CRISP se basa en cinco fases iterativas, cuyas actividades y procesos son supervisadas en el período continuo de auditoría, que dependiendo del proyecto se puede tener en cuenta diversos tipos de auditoría como: auditoría de Fraunhofer, liebowitz, Pfeifer, entre otros [27].

### a. Análisis organizacional

En esta fase se debe comprender el contexto de la organización (misión, visión, objetivos y estrategias) al igual que las metas del problema que desea resolverse. Para esta etapa es recomendable contar con personas de distintas áreas de conocimiento dentro de la organización ya que a partir de su experiencia tienen la posibilidad de aportar ideas y discutir las ya establecidas. Involucrando la mayor parte de la organización se consigue contextualizar el uso de esta metodología en los procesos de gestión de conocimiento establecidos a nivel corporativo.

### b. Adquisición y preparación de datos

Esta etapa permitirá obtener los datos necesarios de acuerdo con la definición del problema e identificar las fuentes en las cuales se puede acceder a ellos. En este punto se realiza un análisis de datos propio del proyecto, logrando determinar la calidad e importancia de los mismos a través de las diversas técnicas y métodos de minería de datos. Posteriormente, se definirán las actividades necesarias para la construcción del conjunto de datos final, que será utilizado en el desarrollo del modelo.

### c. Diseño y modelado

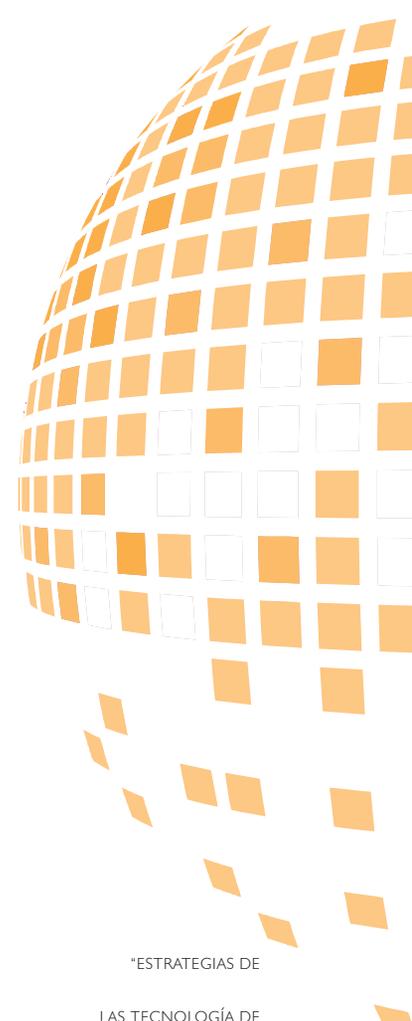
En esta fase se revisarán, seleccionarán y evaluarán diferentes técnicas de modelamiento, apropiadas para la minería, las cuales serán aplicadas al conjunto de datos preestablecido en la etapa anterior, logrando un ajuste mejorado del modelo. En esta etapa se describe un plan detallado para el entrenamiento, prueba y evaluación de los modelos, considerando las técnicas y modelado adecuados para considerar la confidencialidad de la información.

### d. Interpretación y transformación

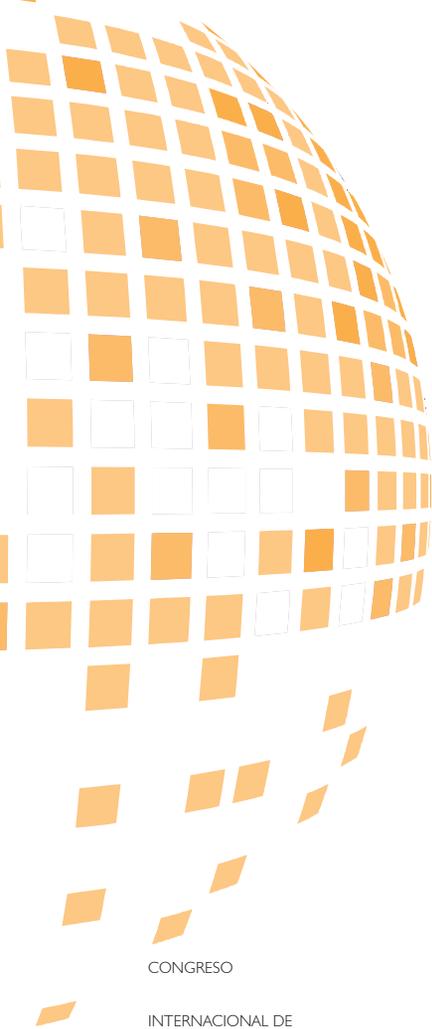
En esta etapa es necesario seleccionar las técnicas de minería apropiadas para realizar la verificación del modelo y realizar las adecuaciones necesarias para que se ajuste al problema que se quiere solucionar. También, es importante destacar los datos y registrar los resultados de análisis que son relevantes para complementar la base de conocimiento (Las bases de conocimiento surgieron como respuesta a las necesidades planteadas en inteligencia artificial y permiten almacenar no sólo datos, sino un conjunto de reglas que sirven para obtener información que no se encuentra almacenada de forma explícita) y facilitar el diseño de las pruebas de evaluación del modelo implantado.

### e. Despliegue, difusión y evaluación

En esta fase, se realiza la implantación del modelo en la organización y los resultados serán evaluados por el usuario final y la organización en general, de acuerdo con los objetivos y las pruebas planteados. Se debe recordar que existen objetivos encaminados hacia la minería y la gestión,



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

**COMTEL 2009**

los cuales, aunque han sido combinados para desarrollarse en las diferentes fases, deben haber sido cumplidos de forma independiente.

## 5. Trabajos Futuros

Los temas relacionados con la gestión del conocimiento, el descubrimiento del conocimiento y la minería de datos son día tras día de más actualidad. Por esta misma razón, surgen con mayor frecuencia proyectos de investigación en estas líneas.

La metodología GC-CRISP es una propuesta inicial, basada en lo que una parte de la teoría permite concluir para establecer relaciones estrechas entre la minería de datos y la gestión del conocimiento. Sin embargo, con el auge de estos temas aplicados a las organizaciones y a problemas de índole social, sería interesante definir casos de estudio particulares cuyo desarrollo se base en la aplicación de la metodología. Entonces, se permitiría determinar ventajas y desventajas con base en los resultados obtenidos y posiblemente se generarían ajustes basados en el uso o generación de nuevas técnicas ya sea en el contexto de la minería de datos, de la gestión del conocimiento o de ambos.

## 6. Conclusiones

De acuerdo con los enfoques analizados, es coherente decir que grandes volúmenes de datos que han sido recolectados sin que puedan ser extraídos, validados, analizados y usados como base para la toma de decisiones no tiene sentido para las organizaciones. Por esto, no es descabellado decir que la minería de datos apoya la gestión del conocimiento y enriquecen los procesos organizacionales.

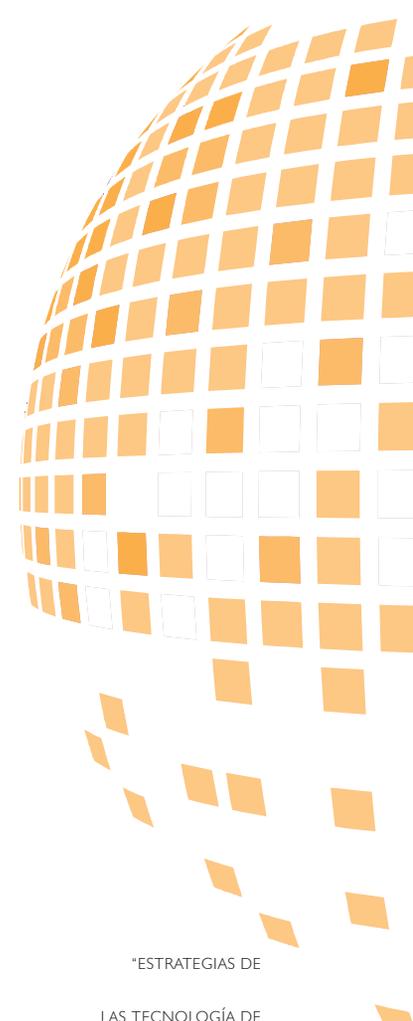
Actualmente, existen diversos conjuntos de técnicas que permiten a las organizaciones buscar el conocimiento a partir de las grandes cantidades de datos que a diario almacenan, uno de las más utilizados es la minería de datos. Así mismo, existen muchas formas de realizar el proceso de minería, pero las más conocidas son: la metodología CRISP-DM y la metodología SEMMA, las cuales han sido desarrolladas por proveedores de herramientas.

La metodología CRISP-DM establece una secuencia lógica de pasos necesarios para la construcción óptima de modelos para el minado de los datos, por lo que se ha convertido en un estándar mundial como referencia para el desarrollo de este tipo de proyectos.

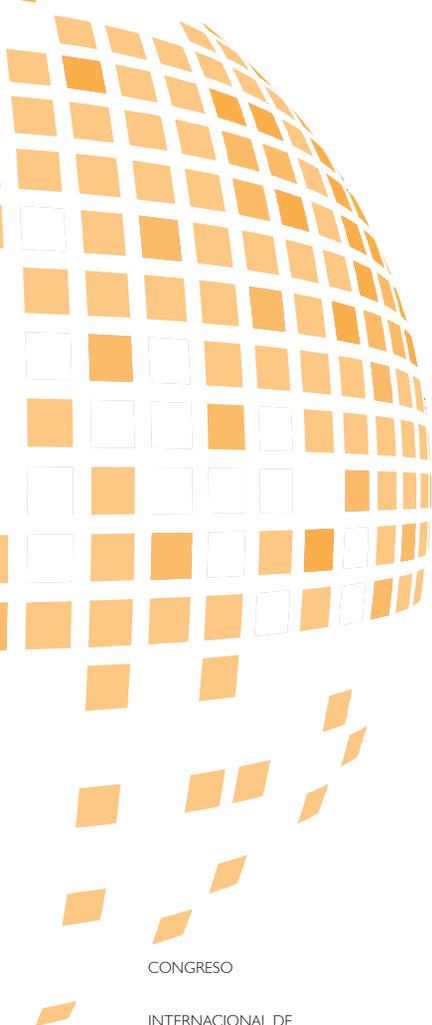
Con la metodología GC-CRISP se pretende proponer una estrategia que permita recopilar las características de la minería de datos y la gestión de conocimiento, vistas como disciplinas complementarias en pro del crecimiento de las organizaciones. De acuerdo con el contexto en el cual se aplique, puede ser más o menos flexible, la idea es que de acuerdo con lo planteado se realiza la retroalimentación para la empresa tanto en conocimiento tácito como explícito.

## 7. Referencias

- [1] D. T. Larose, "Discovering knowledge in data: An Introduction to data mining", New Jersey: John Wiley & Sons, 2005, ch. 1.
- [2] J. Roddick and B. Lees Brian, "Paradigms for spatial and spatio-temporal data mining. In *Geographic Data Mining and Knowledge Discovery*", London: Taylor & Francis, 2001.
- [3] Y.J. Marcano and R. Talavera, "Minería de datos como soporte a la toma de decisiones empresariales". vol, 52, 2007, pp.105-111. ISSN 1012-1587.
- [4] J.A. Harding, M. Shahbaz, Srinivas, and A. Kusiak (2006). "Data Mining in Manufacturing: A Review", *Transactions of the ASME on Journal of Manufacturing Science and Engineering*, ASME, Vol, 128, pp. 969-976. Available: <http://www.icaen.uiowa.edu/~ankusiak/Journal-papers/Harding.pdf>
- [5] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. (2000). *CRISP-DM 1.0. Step-by-step data mining guide*. [Online]. Available: <http://www.crisp-dm.org/CRISPWP-0800.pdf>
- [6] M. Bramer, "Principles of Data Mining", Verlag London: Springer, 2007, ch 1.
- [7] L. A. Kurgan, and P. Musilek, "A survey of Knowledge Discovery and Data Mining process models". *The Knowledge Engineering Review*. vol 21, 2006, pp.1-24. ISSN: 0269-8889
- [8] J. Hernández, M. J. Ramírez, C. Ferri, "Introducción a la minería de datos", Pearson Prentice Hall, 2007, ch 2.
- [9] J. Bulkley, S. Gayle, B. Hicks, and R. Stephens (1999). *Adding the where to the who*. [Online]. Available:  
  
<http://support.sas.com/rnd/datavisualization/papers/AddingWhereToWho.pdf>
- [10] SAS. (2008) *SAS Enterprise Miner 5.3 Specifications*. [Online]. Available: <http://www.sas.com/technologies/analytics/datamining/miner/factsheet.pdf>
- [11] K. J. CIOS, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan. "Data mining. A knowledge Discovery Approach", Springer, 2007, ch. 1.
- [12] M. L. Vásquez, C. L. Gómez. A. M. López, and J. Robledo. (1999). "Knowledge Management Systems Assessment: A Conceptual Framework and A Methodological Proposal", *International Association for Management of Technology*. Available: <http://www.iamot.org/paperarchive/153C.PDF>



"ESTRATEGIAS DE  
LAS TECNOLOGÍAS DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

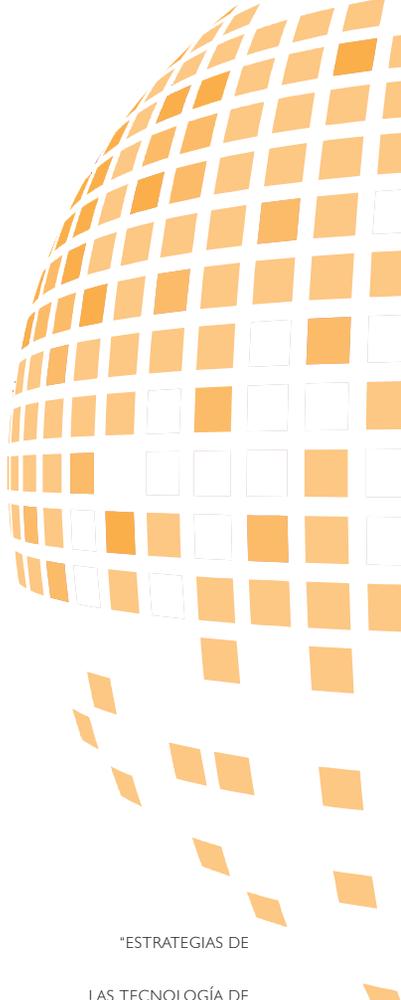
INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

- [13] Librería ZAINTEK. Diputación Foral de Bizkaia. Departamento de Promoción Económica. (2001).Gestión de conocimiento. [Online]. Available:  
[http://www.zaintek.net/ebizkaia/servlet/EbizkaiaServlet?P=mDOC&dir=1&nombre=B12\\_13.pdf&L=1](http://www.zaintek.net/ebizkaia/servlet/EbizkaiaServlet?P=mDOC&dir=1&nombre=B12_13.pdf&L=1)
- [14] European Committee for Standardization (2004). European Guide to good Practice in Knowledge Management, ch 4. [Online]. Available:  
<ftp://cenftp1.cenorm.be/PUBLIC/CWAs/e-Europe/KM/CWA14924-04-2004-Mar.pdf>
- [15] J. Firestone (2000). "Knowledge Management Process Methodology: An Overview". Knowledge and innovation: journal of the KMCI. [Online]. Available: <http://whitepapers.zdnet.co.uk/0,1000000651,260095628p,00.htm>
- [16] J. Swan, and H. Scarbrough, "The Paradox of Knowledge Management". INFORMATIK. Knowledge Management. vol. 1, 2002, pp.10-13.
- [17] S. Sinha, M. Lindvall, and I. Rus (2004). "Software Systems Support for Knowledge Management". Project Performance Corporation. [Online]. Available: <http://www.ppc.com/Documents/SW-Systems-Support-for-KM-Full.pdf>
- [18] Resnick, Marc. Knowledge management in the information age. Proceedings of the Internacional Academy of E-business. 2nd Annual Conference, Florida, 2002.
- [19] G.Lawton, "Knowledge Management: Ready for Prime Time?," IEEE Computer, vol. 34, 2001, pp. 12-14.
- [20] J. Luftman, "La competencia en la era de la información. La alineación estratégica en la práctica". Oxford: University Press. 2001 .
- [21] B. Carrillo (2009). "Importancia de las tecnologías de la información y la comunicación en el proceso educativo". [Online]. Available: [http://www.csicsif.es/andalucia/modules/mod\\_ense/revista/pdf/Numero\\_14/BEATRIZ\\_CARRILLO\\_1.pdf](http://www.csicsif.es/andalucia/modules/mod_ense/revista/pdf/Numero_14/BEATRIZ_CARRILLO_1.pdf)
- [22] K. Bartlmae, and M. Riemenschneider (2000). "Case Based Reasoning for Knowledge Management in KDD-Projects. Concepts, Organizational Setting, Categorization into KM and Application in the case of Knowledge Discovery in Databases". 2000. [Online]. Available: [http://reference.kfupm.edu.sa/content/c/a/case\\_based\\_reasoning\\_for\\_knowledge\\_manag\\_106807.pdf](http://reference.kfupm.edu.sa/content/c/a/case_based_reasoning_for_knowledge_manag_106807.pdf)

- 
- [23] C. Brewster, F. Ciravegna, and Y. Wilks (2001), "Knowledge Acquisition for Knowledge Management: Position Paper". [Online]. Available: <http://www.dcs.shef.ac.uk/~fabio/paperi/ontolearning.pdf>
- [24] P. Cimiano, F. Ciravegna, J. Domingue, S. Handschuh, A. Lavelli, S. Staab, and M. Stevenson (2003), "Requirements for Information Extraction for Knowledge Management". [Online]. Available: [http://www.dcs.shef.ac.uk/~fabio/paperi/Philipp\\_Cimiano-et-al.pdf](http://www.dcs.shef.ac.uk/~fabio/paperi/Philipp_Cimiano-et-al.pdf)
- [25] A. Veloso, B. P. Wagner, and M. B. Carvalho (2001), Knowledge Management in Association Rule Mining. IEEE ICDM. [Online]. Available: <http://homepages.dcc.ufmg.br/~adrianov/papers/DMKM01/Veloso-ieee.pdf>
- [26] European Committee for Standardization (2004). "European Guide to good Practice in Knowledge Management", ch 1. [Online]. Available: <ftp://cenftp1.cenorm.be/PUBLIC/CWAs/e-Europe/KM/CWA14924-01-2004-Mar.pdf>
- [27] K. Mertins, P. Heisig, and J. Vorbeck. Knowledge management: concepts and best practices. Spring:2003, ch 1.

"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"