

SUBSECRETARÍA DE EDUCACIÓN SUPERIOR
DIRECCIÓN GENERAL DE EDUCACIÓN SUPERIOR
TECNOLÓGICA
INSTITUTO TECNOLÓGICO DE ORIZABA



“60 Años de Excelencia en Educación Tecnológica”

SECRETARÍA DE
EDUCACIÓN PÚBLICA

SEP

Instituto Tecnológico de Orizaba

MAESTRÍA EN INGENIERÍA ADMINISTRATIVA

TEMA:
MINERÍA DE DATOS Y TEXTOS

MATERIA:
FUNDAMENTOS DE INGENIERÍA ADMINISTRATIVA

CATEDRÁTICO:
DR. FERNANDO AGUIRRE Y HERNÁNDEZ

PRESENTA:
LUIS GERARDO RODRÍGUEZ FRANCO

ORIZABA, VER.

4/ ABRIL /2016

Avenida Oriente 9 Núm. 852, Colonia Emiliano Zapata, C.P. 94320
Orizaba, Veracruz, México
Teléfonos: (272) 7 24 40 96, (272) 7 24 40 16, (272) 7 24 45 79
Fax: (272) 7 25 17 28

e-mail: orizaba@itorizaba.edu.mx
www.itorizaba.edu.mx

CONTENIDO

OBJETIVO	1
INTRODUCCION	1
¿QUÉ ES LA INFOMULTICULTURALIDAD?	¡Error! Marcador no definido.
INFORMACION	¡Error! Marcador no definido.
COMUNICACIÓN	¡Error! Marcador no definido.
PROCESO BÁSICO DE LA COMUNICACIÓN.....	¡Error! Marcador no definido.
CARACTERÍSTICAS DE LA INFORMACIÓN.....	¡Error! Marcador no definido.
CULTURA	¡Error! Marcador no definido.
CULTURA ORGANIZACIONAL	¡Error! Marcador no definido.
MULTICULTURAIDAD.....	¡Error! Marcador no definido.
CONCLUSIONES	¡Error! Marcador no definido.
REFERENCIAS	¡Error! Marcador no definido.
AGRADECIMIENTOS.....	1

MINERÍA DE DATOS Y TEXTOS

OBJETIVO

Hacer mención de lo que es la metodología de alcance de procesos y dar una breve descripción de cada uno de los pasos y cuál es el propósito del uso de esta técnica.

INTRODUCCION

Hoy en día las organizaciones empresariales cuentan con mucha información la cual deben de manejar de la manera más eficiente posible, entre la información que se debe manejar se encuentran ventas, clientes, cobros, en caso de ser un hospital pacientes, tratamientos, y así sucesivamente dependiendo del giro al cual esté destinada la organización, es por ello que para tener una mejor apreciación y control de la información de las organizaciones se utilizan equipos de cómputo y de almacenamiento como lo son los computadores personales, memorias USB, Cds, memorias flash, etc, de igual manera se hace la utilización de estos dispositivos ya que el impacto de inversión para adquirir uno de estos equipos se ha ido abaratando con el paso de los años, sin embargo igual se realiza ya que estos sistemas de información llegan a ser mucho más confiables que el propio ser humano.

Toda esta información que se guarda dentro de los sistemas de información reside en las bases de datos que se utilizan en las labores de las organizaciones, las cuales son llamadas bases de datos operacionales, las cuales reciben su nombre debido a que mediante ellas las organizaciones realizan distintas actividades tales como lo son los envíos de mercancías a los clientes, registro de estudiantes, tratamientos de pacientes, procedimientos de cobranza y otros.

Una vez que se realizan estas operaciones se procede a realizar una depuración y se resume la información obtenida mediante las fuentes primarias las cuales vendrían siendo todas las tareas antes enlistadas, a recolección, depuración y resumen de información se transfiere a lo que se le llama como bodegas de datos, las cuales por decir una metáfora podrían ser fotografías periódicas las cuales son utilizadas para tener una noción del estado en el que se ha encontrado la empresa y como tener una aprendiendo del pasado.

Es de esta manera que los empresarios pueden tener indicadores ideales para controlar el rumbo de la empresa, ya que puede tener la oportunidad de indagar y explorar en muchas situaciones en las cuales se puede considerar que son de interés y de preocupación para el logro de los objetivos de la empresa.

Fue así como hace poco tiempo apareció lo que es la minería de datos, la cual viene a ayudar a los altos directivos de las organizaciones a tomar las mejores decisiones para la empresa en la cual labora. La minería de datos funciona a través de una serie de “mineros” los cuales son una serie de algoritmos previamente creados los cuales se ven a la tarea de realizar una tarea exhaustiva dentro de la información que almacena la organización en sus sistemas de información, el material que buscan estos algoritmos no es más que una serie de tendencias, anomalías, desviaciones o situaciones que podrían ser de interés, las cuales pudiesen ser desconocidas por las empresas. Estos algoritmos o mineros, ayudan a los directivos a poder dirigir de una manera más fácil a la organización, llevándola así por el camino correcto.

Los mineros utilizan “además de las bases de datos, la inteligencia artificial (procedimientos para hallar grupos en situaciones similares, clasificar eventos nuevos en categorías conocidas, etcétera) y la estadística. Pero a diferencia de esta última, que toma una muestra de los datos y la estudia, la minería de datos

estudia todos los datos. Mientras más datos se analicen, más precisa es, y su poder de detección y predicción aumenta.” (Martínez Luna, 2011)

Todo lo anterior fue narrado con respecto a lo que comprende la minería de datos, sin embargo existe otro tipo de minería que de la misma manera podría ayudar a las empresas a conseguir esos objetivos tan deseados que ha establecido en un principio.

Para nosotros como seres humanos, el conocimiento es una de las bases de nuestra existencia, la cual delimita hacia donde nos dirigimos y de igual manera nuestras ambiciones. La mayor parte del conocimiento que ha generado la raza humana se encuentra de manera escrita la cual se puede nombrar como lenguaje natural los cuales son los periódicos, revistas, libros, informes técnicos, etcétera. Sin embargo, no todas las personas tienen la misma capacidad para poder manejar los contenidos bibliográficos, es así como podemos decir que las tareas más convencionales a las cuales nos vemos necesitados todos los seres humanos a lo largo de nuestra vida es a la de interactuar con escritos para poder tener algún beneficio. Las competencias que debería tener un buen lector y buscador de información serían:

- Buscar la información necesaria
- Comparar fuentes de información diferentes, y obtener conclusiones
- Manejar los textos, por ejemplo, traducirlos, editarlos, etc. (Montes y Gómez, 2011)

Al observar nuestras carencias para el manejo de información, la lingüística computacional llega a ser una herramienta muy fuerte para ayudarnos al procesamiento de textos, ya que mediante ésta técnica, el análisis de la información se puede realizar de manera automática, resolviendo así los problemas que tienen la mayoría de las personas.

Así como la minería de datos busca una serie de patrones dentro de un conjunto de datos, la minería de texto realiza la misma actividad pero tomando como datos los textos que se le pueda alimentar a un sistema computacional, además de que también se ve a la tarea de poder detectar desviaciones y asociaciones entre cada uno de los textos que se puedan analizar.

LA MINERÍA DE DATOS

Mediante la revolución de la era digital, los procesos para poder manejar la información se han vuelto más eficientes que en la anterioridad, es así como decimos que el proceso de la información dentro de los sistemas digitales consta básicamente de cinco pasos los cuales son:

- Capturar
- Procesar
- Almacenar
- Distribuir
- Transmitir

Mediante el uso de la informática, las grandes organizaciones del globo han ido recopilando grandes cantidades de datos históricos que se han ido obteniendo con la experiencia, sin embargo la información sigue creciendo en los sistemas de información computacional haciendo estas cantidades cada vez más grandes.

Sin embargo el manejo de estas grandes cantidades de información es un tanto complicada y es por ello que la minería de datos ha nacido, “surge como un intento de buscarle sentido a la explosión de información que actualmente puede ser almacenada” (Mitra & Acharya, 2003)

Así mediante la utilización de la tecnología es posible almacenar distintos tipos de datos ya sea imágenes, videos, textos y datos numéricos en una interfaz

relativamente sencilla permite facilitar un buen manejo multimedia de la información.

Se puede decir que mediante este tipo de mezcla de información, los procesos estadísticos convencionales para analizar la información obtenida resultan insuficientes ya que las técnicas estadísticas se enfocan a la utilización de muestras, al contrario de la minería de datos que utiliza a todo el universo de datos para tener una mejor apreciación y solución.

De esta manera llegamos a la definición de lo que es la minería de datos la cual es “el proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores aparentemente caóticos que tienen una explicación que pueden descubrirse mediante diversas técnicas de esta herramienta.” (Ángeles Larrieta & Santillán Gómez, 2001)

La minería de datos se utiliza dentro de las empresas para poder aprovechar el valor de la información contenida dentro de las bases de datos para detectar como se mencionó anteriormente patrones preestablecidos para que los altos directivos de las organizaciones puedan tener mejor conocimiento del negocio que manejan y así realizar procesos de toma de decisiones más eficientes.

VENTAJAS DEL USO DE LA MINERÍA DE DATOS CON RESPECTO A OTRAS TÉCNICAS DE MANEJO DE INFORMACIÓN

La minería de datos nace de las necesidades de manejar información contenida dentro de las bases de datos de las organizaciones, este procedimiento tiene una serie de ventajas sobre otros procesos que son utilizados para el manejo de información como lo son:

- La minería de datos proporciona a los altos mandos empresariales un conjunto de relaciones y conocimiento que en muchas ocasiones no se sabía que existía dentro de la organización.
- La minería de datos ayuda a las empresas a elegir las rutas por donde llevarán el curso de las empresas, así como a conseguir ventajas competitivas contra sus rivales de mercado, ya que mediante el uso de la minería de datos se conocerá información que solo la empresa conoce de manera exclusiva.
- Nosotros como seres humanos poseemos la capacidad para detectar patrones y anomalías de una manera por decirlo así superficial, es por ello que mediante el uso de la minería de datos se podrá percibir de una mejor manera patrones que a simple vista son difíciles de localizar por nuestra simple apreciación.

ESTRUCTURA DE LA MINERÍA DE DATOS

Ahora bien, hablando acerca de la estructura de la minería de datos, esta consta básicamente en el uso de un algoritmo o algún programa computacional para realizar las actividades de búsqueda dentro de las grandes cantidades de información contenidas en la base de datos.

El uso de estos programas y algoritmos es con el fin de poder detectar tendencias y patrones que se encuentran de alguna manera ocultos en los datos históricos de las organizaciones.

Estos programas son los que anteriormente llamamos mineros, estos mineros, programas o algoritmos, son creados por los usuarios en los cuales se utilizan varias técnicas de exploración de datos, las técnicas que se pueden utilizar son:

- Cluster
- Asociaciones
- Clasificaciones
- Visualizaciones

- Redes neuronales
- Algoritmos genéricos
- Detección de desviaciones

Todos estos métodos antes mencionados requieren de una base de datos muy grande para que así puedan tener una mayor eficiencia.

Estos programas tienen la función de recopilar la información previamente obtenida y como consiguiente realizan las actividades de selección y búsqueda en los datos históricos, después de hacer lo anterior si se llegara a encontrar algo interesante se le muestra al usuario.

Los “mineros” tienen una ventaja con respecto a otros métodos de búsqueda de información, la cual es que no necesitan de algún software especializado para realizar búsquedas. Estas actividades de búsqueda se realizan en los servidores de las empresas y toda la red de Pc’s que son utilizadas para la captura de datos e información.

CICLO DE LA MINERÍA DE DATOS

La minería de datos trabaja sobre un ciclo que contiene cuatro pasos, ya que los resultados obtenidos después de que el ciclo termina, pueden volver a alimentar al ciclo y así sucesivamente.

1. Primeramente los usuarios que llevarán a cabo el proceso de la minería de datos, deberán identificar los problemas que tiene la organización, empresa o negocio de igual manera deben localiza los datos que pueden dar una especie de valor agregado a la empresa y también se deben localizar las áreas de la empresa donde la información es extremadamente cambiante.

2. Una vez realizado lo anterior, el usuario se verá a la tarea de detectar el mejor algoritmo que utilizará para analizar los datos históricos obtenidos para que así los programas mineros puedan trabajar de una manera eficiente de acuerdo a los criterios de búsqueda previamente establecidos.
3. Se debe incorporar la información obtenida mediante el proceso de la minería de datos a la toma de decisiones, proporcionando los hallazgos obtenidos al comité que se ve involucrados en tomar decisiones, de la misma manera se debe de dar conocimiento de los problemas detectados a las áreas involucradas para que se pueda aplicar una correcta solución.
4. Finalmente se debe de realizar una medición de los resultados obtenidos proporcionados a la persona o comité encargados de la toma de decisiones de acuerdo a los problemas hallados de acuerdo a los criterios de búsqueda previamente establecidos.

USOS Y APLICACIONES DE LA MINERÍA DE DATOS

Algunas de las tareas más importantes que se pueden desempeñar mediante la utilización de la minería de datos son las siguientes:

- **Comercio y banca:** segmentación de clientes, previsión de ventas, análisis de riesgo.
- **Medicina y farmacia:** diagnóstico de enfermedades y la efectividad de los tratamientos
- **Seguridad y detección de fraude:** reconocimiento facial, identificaciones biométricas, acceso a redes, etc.
- **Recuperación de información no numérica:** minería de texto, minería web, búsqueda e identificación de imagen, video, voz y texto de bases de datos multimedia.
- **Astronomía:** identificación de nuevas estrellas y galaxias.

- **Geología, minería, agricultura y pesca:** identificación de áreas de uso para distintos cultivos o de pesca o de exploración en bases de datos de imágenes de satélites.
- **Ciencias ambientales:** identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales (plantas depuradoras de aguas residuales) para mejorar su observación, gestión y/o control.
- **Ciencias sociales:** estudios de los flujos de la opinión pública.
- **Planificación de ciudades:** identificar barrios con conflicto en función de valores sociodemográficos. (Riquelme, Ruíz, & Gilbert, 2006)

MINERÍA DE TEXTOS

La minería de textos es la parte más nueva que se conoce acerca del área de investigación enfocada al procesamiento de textos. La definición que se le puede otorgar a la minería de textos es muy similar al de la minería de datos ya que ambos buscan lo mismo pero apegado a diferentes tipos de información.

La minería de textos es “el proceso de descubrimiento de patrones interesantes y nuevos conocimientos en una colección de textos, es decir, la minería de texto es el proceso encargado del descubrimiento de conocimientos que no existan explícitamente en ningún texto de la colección, pero que surgen de relacionar el contenido de varios de ellos (Hearst, 1999)

El proceso de la minería de textos consta básicamente de dos etapas las cuales son:

- **Etapas de procesamiento:** En la primera etapa, los textos que se puedan manipular se transforman en una serie de representaciones estructuradas de tal manera que se promueva la facilidad de un análisis posterior.

- **Etapa de descubrimiento:** En esta etapa se procede a realizar un análisis de las representaciones intermedias, esta tarea se realiza con el objetivo de así poder descubrir y encontrar patrones interesantes dentro de los textos de interés, así como también se busca obtener nuevos conocimientos.

De acuerdo a los procedimientos utilizados dentro de la etapa del procesamiento de los textos, es el tipo de representación de contenido que se obtendrá. Las estrategias que pueden ser utilizadas para el procesamiento de textos en la minería de datos son las siguientes.

Etapa de pre-procesamiento	Tipo de representación	Tipo de descubrimientos
1. Categorización	4. Vector de temas	7. Nivel temático
2. Full-text	5. Secuencia de palabras	8. Patrones de lenguaje
3. Extracción de información	6. Tabla de datos	9. Relaciones entre entidades

Fig. 1.1 Estado del arte de la minería de texto (Montes y Gómez, 2011)

Como se pudo apreciar en la figura 1.1 los tres tipos de métodos que existen para el análisis de textos se ven un tanto limitados en la presentación de los resultados, lo cual dificulta de gran manera el poder descubrir y tener conocimiento de algunas cosas más complejas como lo pueden ser:

1. Consensos
2. Tendencias
3. Desviaciones

Sin embargo para poder conseguir una mejor apreciación de lo antes mencionado se recomienda que se puede hacer uso de gráficos conceptuales, con los cuales se podrá tener una mejor representación de los textos analizados.

Aun así, el análisis de información de acuerdo con los gráficos conceptuales trae consigo dos tipos de problemas los cuales se ven relacionados con el análisis sintáctico y con el análisis semántico de los textos. Algunos ejemplos de textos transformados en gráficos conceptuales son:

- Partes de artículos científicos
- Partes de expedientes médicos
- Partes de casos legales

Sin embargo no existen métodos que permitan interpretar de una correcta manera los gráficos conceptuales, a lo cual la minería de textos puede ser una parte fundamental para el tratamiento de éste tipo de información y darle el mejor sentido posible de acuerdo a los parámetros que se estén utilizando para el proceso de minería de textos.

CONCLUSIONES

Como se pudo observar, la minería de datos es una herramienta muy importante para poder interpretar los rumbos de una empresa, tomando en cuenta datos históricos obtenidos a través del tiempo, este tipo de minería podrá descubrir tendencias que existan acerca de algún problema relacionado con la organización o podrá otorgarle algún tipo de ventaja con el conocimiento de cierta información exclusiva descubierta a través de la utilización del ciclo de la minería de datos, por su parte, la minería de textos aporta casi lo mismo que la minería de datos, pero apegado al descubrimiento de nuevo conocimiento partiendo de un gran conjunto de textos.

REFERENCIAS

1. Ángeles Larrieta, M. I., & Santillán Gómez, A. M. (2001). *Minería de datos: concepto, características, estructura y aplicaciones*. UNAM.
2. Hearst. (1999). *Untangling Text Data Mining Proc. of ACL '99: The 37th Annual Meeting of the Association for Computational Linguistics*. Maryland: University of Maryland.
3. Martínez Luna, G. L. (Octubre de 2011). Minería de datos: Cómo hallar una aguja en un pajar. (UANL, Ed.) *Ingenierías, XIV(53)*, 63. Recuperado el 23 de marzo de 2016
4. Mitra, S., & Acharya, T. (2003). *Data mining: multimedia, soft computing and bioinformatics*. John Wiley & Sons.
5. Montes y Gómez, M. (2011). *Minería de texto: Un nuevo reto computacional*. México, D.F.: Instituto Politécnico Nacional.
6. Riquelme, J. C., Ruíz, R., & Gilbert, K. (2006). Minería de datos: conceptos y tendencias. *Inteligencia Artificial, 10(29)*.

AGRADECIMIENTOS

Quiero agradecer de una manera muy especial a la materia de Fundamentos de Ingeniería Administrativa de la Maestría en Ingeniería Administrativa que estudio en el Instituto Tecnológico de Orizaba, pero principalmente a mi profesor el Dr. Fernando Aguirre y Hernandez por incentivar el deseo de investigar y leer acerca de distintos temas tan interesantes.