

Grandes planes para grandes datos!

Las organizaciones inteligentes siempre han confiado en los datos para **tomar decisiones estratégicas de negocios**. Sin embargo, “los grandes datos ” tiene sus retos que deben abordarse antes de que puedan tener un impacto real.

El término “grande datos” se ha convertido en un algo expansivo que lo abarca todo. Aparentemente, tiene tantas definiciones como aplicaciones. Entonces, ¿qué quiere decir exactamente “grandes datos”, y que significa para las empresas y la normalización?

Jim Melton es un experto autorizado y escritor en el área sobre el lenguaje de base de datos. Preside el ISO/IEC JTC 1/SC 32 sobre gestión de datos e intercambio y cuenta con más de 25 años de experiencia. El Dr. Klaus-Peter Eckert es un colaborador científico del Instituto Fraunhofer de Sistemas de Comunicación Abierta. Durante más de 30 años, ha estado involucrado en proyectos relacionados con la interoperabilidad y plataformas de aplicaciones y servicios distribuidos. ¿Cómo definimos los datos? Aunque me he encontrado con varias descripciones , aquí transcribo la definición de una frase del libro “ **The Ethics of Big Data**” que me gustaría usar para entrar en el tema: **“Este aspecto de la tecnología es único, ya que su propia naturaleza (su esencia) es crear, conectar, correlacionar, agregar, almacenar, procesar e informar sobre cantidades masivas de información.”** Para hacerlo más simple, los grandes datos representan grandes cantidades de información compleja para tomar decisiones de negocios mejor informadas. Pero, como explican Eckert y Melton, los grandes datos tiene sus propios desafíos arraigados que deben abordarse antes de que puedan tener un impacto real.

Los ciegos y los elefantes

convergencia de los diferentes enfoques y la interoperabilidad entre las herramientas disponibles todavía es deficiente. “Tenemos varias herramientas alrededor de los grandes datos y el análisis de datos que han sido desarrolladas por las diferentes comunidades en los últimos años. Tenemos diferentes tipos de infraestructuras de TI, especialmente en la computación en la nube, que se han desarrollado de forma independiente de los grandes datos. Así, todos estos bloques de construcción técnicos están disponibles, pero que no encajan. Falta interoperabilidad. Existe una falta de arquitectura ampliamente acordada que reúna los varios componentes de soluciones de los grandes datos. Y aquí es exactamente donde los esfuerzos de la normalización entran en juego”.

Tanto Melton como Eckert creen que la complejidad de los grandes datos y las interacciones entre muchos dominios requieren mayor análisis. Esta es precisamente la misión del Grupo de Estudio ISO/IEC JTC 1. Mapeo de las complejidades Se espera que el Grupo de Estudio de reciente creación pueda dar una imagen mucho más clara de todo el alcance de los problemas, incluidas las cuestiones éticas graves, tales como si ciertos usos de los grandes datos violan los derechos civiles, sociales, políticos y jurídicos fundamentales. El resultado final será una o más recomendaciones al JTC 1 referentes a la asignación de proyectos a sus subcomités, momento en el cual el desarrollo de normas formales comenzará en serio.

El desarrollo de normas en este **laberinto de complejidades de los grandes datos** creará una plataforma para la interoperabilidad.

‘Regalando’ datos

Cuando Edward Snowden sacó a la luz documentos clasificados sobre los programas de vigilancia llevados a cabo por la NSA, la mayoría de la gente crítico que los gobiernos recolectasen datos e información personal sin consentimiento. Sin embargo, no necesitan mucho esfuerzo para conseguirlos. La NSA no escuchaba las conversaciones ni leía los documentos de textos, simplemente recogían los metadatos, datos que permiten describir datos como el lugar y la fecha en la que se hizo una foto.

Hoy en día, compartimos los datos de nuestra vida privada en redes sociales, con aplicaciones que muestran el recorrido que realizamos mientras hacemos deporte o simplemente permitiendo que las aplicaciones de nuestro smartphones guarden nuestra localización. Además, instalamos aplicaciones y aceptamos ‘Términos y condiciones de uso» sin informarnos ni cuestionarnos si realmente necesitan esos datos.



Permisos que piden las 10 apps de linterna más descargadas en Android

Para comprobar la importancia de los datos y si realmente eran útiles y desvelaban información sobre nuestra vida privada, la Universidad de Stanford realizó un experimento y pidió a 546 voluntarios que se instalaran una app en su teléfono que recopilaría la misma información que, en teoría, la NSA había estado recopilando, como la fecha de la llamada, la duración o el destinatario, sin escuchar ninguna llamada. Los resultados fueron sorprendentes. Se pudo extraer información sobre su religión dependiendo a los centros religiosos que llamaban o sobre las enfermedades gracias a los expertos que consultaron. Además, también relacionaron varias llamadas y pudieron descubrir que una chica estaba embarazada porque realizó una llamada de bastante duración con su hermana a una hora demasiado inusual de la mañana y un par de días después, llamó a un centro de Planificación Familiar, al que volvió a llamar durante diversas ocasiones en las siguientes semanas.

Otro claro ejemplo del peligro de los datos que ofrecemos, sin ser conscientes, en las fotos que publicamos en Internet es 'I know where your cat lives' (<http://iknowwhereyourcatlives.com/>). Esta página web, creada por Owen Mundy, utiliza los metadatos de más de un millón de fotografías con la etiqueta cat de páginas como Instagram o Flickr y las mapea utilizando las coordenadas geográficas. El objetivo de este experimento era precisamente mostrar el poder de los datos que colgamos de manera inconsciente cada vez que publicamos una foto o realizamos un comentario y como 'regalamos' a las empresas nuestra privacidad, ya no solo online sino también física.

Estos ejemplos pueden ser los más claros e impactantes por el conocimiento que pueden llegar a tener sobre nosotros las empresas, pero no solo en Internet recogen nuestros datos. Hoy en día, todas las empresas tratan (o deberían de tratar) obtener datos sobre sus clientes, como por ejemplo las tarjetas cliente que ofrecen descuentos, para, a través de herramientas de Big Data, descubrir sus comportamientos y optimizar sus ofertas y productos.

El futuro incierto de la tecnología

A menudo, los humanos nos solemos preguntar cómo será el futuro que nos espera; cómo será ese futuro que surgirá de las nuevas tecnologías. Si miramos atrás en el tiempo, nos podremos dar cuenta de cómo los avances tecnológicos han influido en el desarrollo de la sociedad y cómo estos avances han provocado también cambios en la sociedad. La tecnología trae consigo muchas comodidades para mejorar la calidad de nuestras vidas pero, a su vez, trae también cambios en la forma en la que nos relacionamos los seres humanos dentro de una sociedad. Por ejemplo, antes de tener teléfonos móviles la forma en la que contactaban unas personas con otras era diferente. Por eso, podemos afirmar que los avances tecnológicos afectan de una manera importante a la sociedad.

Además, otro de los aspectos que a mí personalmente me preocupa mucho es la viabilidad que pueda tener el futuro tecnológico que nos espera. Es decir, si realmente es sostenible el desarrollo tecnológico a gran escala.

Puede que nunca os hayáis preguntado si realmente es viable o no la revolución tecnológica que nos espera. Ahora mismo se habla de Big Data, IoT, Smart Cities... y otra serie de tendencias tecnológicas futuras que se están empezando a implantar. Pero, ¿qué tienen en común todas estas tendencias?

La respuesta es sencilla, todas estas tecnologías necesitarían una cantidad inmensa de componentes electrónicos para su puesta en escena. Por ejemplo, para poder tener una Smart City, la ciudad en cuestión tendría que estar compuesta por un ecosistema tecnológico repleto de componentes electrónicos. Por supuesto, cada uno de estos componentes electrónicos contendrían condensadores, resistencias y otro tipo de materiales. El problema que existe en la actualidad en torno a esto es que para la fabricación de dichos materiales el mineral más utilizado es la tantalita.

Como bien hemos dichos, la principal aplicación de la tantalita es crear condensadores en equipos electrónicos. La ventaja principal de este elemento en los condensadores es que tiene una alta eficiencia volumétrica, que permite reducir el tamaño, tiene una alta fiabilidad y estabilidad en un amplio rango de temperatura (-55 °C a 125 °C). Estas características hacen que los condensadores de otros materiales como la cerámica no puedan igualarlo. Sin salir de la electrónica, también se puede aplicar la tantalita para fabricar resistencias de alta potencia.

En este punto os preguntaréis, ¿y cuál es el problema? Si existe un material adecuado para la fabricación de condensadores y resistencias, ¿por qué no utilizarlo?

La pregunta que yo os recomendaría que os hicieseis es ¿de dónde se obtiene actualmente la tantalita?

La respuesta a esa pregunta es del coltán. El coltán es un mineral compuesto por columbita y tantalita y se explota con la finalidad de poderle extraer la tantalita que haya en él para la fabricación de condensadores y otros componentes electrónicos mencionados anteriormente.



El coltán se extrae de las minas de coltán. Estas minas están en la superficie de la tierra y tienen un difícil acceso. Este es el motivo por el cual se emplean niños para la extracción de dicho mineral, básicamente, porque son los únicos que pueden entrar por dichos agujeros para extraer el material. Como podréis ir viendo, es algo realmente grave e inhumano utilizar niños para explotar minas. Además, se corre el peligro de que por el simple hecho de estar en la superficie de la tierra estas minas se puedan derrumbar con facilidad y los niños se puedan quedar atrapados dentro extrayendo el mineral.



Por si fuera poco, tenemos que tener en cuenta que el 80% de las reservas de coltán que hay en el mundo se encuentran en la República Democrática del Congo. Como este mineral está considerado como un recurso no renovable altamente estratégico, existe una guerra en el Congo desde 1998 por el control de los yacimientos. Esta guerra, directamente relacionada con la explotación de este mineral, arroja un saldo de más de 5,5 millones de víctimas, lo que supone el mayor número de muertes desde la Segunda Guerra Mundial.

Vista la dependencia que tienen los componentes electrónicos del coltán, como poco podemos pensar que el futuro tecnológico que nos espera es incierto. Es decir, no puede darse un desarrollo tecnológico global con el objetivo de mejorar la calidad de vida de los seres humanos cuando, en realidad, lo que este desarrollo está provocando y alimentando es una guerra que se ha cobrado más de 5.5 millones de víctimas. Por eso me parece incierto el futuro tecnológico, porque no se cuál será la forma de sustituir este mineral (tantalita) por otro que no provoque miseria y devastación en el mundo.

Porque, ¿os gustaría vivir un desarrollo tecnológico de la manera en la que se está llevando a cabo en el presente? A mi **NO**.

Big data o Big...Brother

Hace unos días leía en la revista **XLsemanal** un [artículo](#) relacionado con un tema que nos viene como anillo al dedo. El titular era:

¿Se dejaría vigilar todo el tiempo en el trabajo?

Pensémoslo 5 segundos. [.] Seguramente nuestra respuesta, así a bote pronto sería **NO**. En el trabajo o en cualquier lugar, lo de sentirnos «



espíados» no termina de resultarnos cómodos. Pero... ¿tampoco nos sentimos así en las **redes sociales**, **asistentes personales**, **pulseras de actividad**, diferentes **aplicaciones móviles** y otras fuentes de captación de nuestra información? Puede parecer que no pero parándonos a pensar un momento nos daremos cuenta de que estamos **bastante más vigilados** de lo que seguramente nos gustaría...

Sin embargo, también es cierto que debido a todos los **patrones** que se identifican con los análisis de **Big Data**, se pueden llevar a cabo muchísimas **prácticas beneficiosas** para muchas organizaciones, desde hospitales con sus pacientes hasta **recomendaciones personales** de cualquier cosa que nos imaginemos. El caso que mostraba el artículo



de la revista anteriormente nombrada era precisamente un hospital estadounidense en el que los profesionales están **monitorizados** en **toda su jornada laboral**. Todo con unos **objetivos** claros: minimizar costes. Cada persona lleva un **receptor** en el que se puede saber en todo momento **dónde está**, **cuánto tiempo pasa en el baño**, **en la sala de café**, etc. Por supuesto, el dispositivo no es para nada obligatorio, aunque todos lo llevan.

Otra parte que me ha parecido curiosa es que **Big Data** está (o estará aún más) bastante relacionada con **recursos humanos** de la empresa. A día de hoy, hay aplicaciones que son utilizadas para contratar a los empleados que mediante una **entrevista** hecha por ordenador de unos 15 minutos, la máquina **revela**

informes con información de carácter personal: dotes de mando, preocupaciones, aficionado al riesgo, y otros datos muy interesantes para los jefes de personal. El **análisis** es **llamativo** por el hecho de que no se fija en las respuestas dadas, sino en el **volumen de la voz**, la **velocidad del habla**, el **número de negaciones** o **lo largas que son sus pausas**. Los partidarios del Big Data aseguran que con este tipo de mecanismos y otros parecidos, los despidos y contratos serán **más objetivos**.



Lo que está claro es que nos movemos en una era en la que **Big Data y similares están en auge**. Nuestras ciudades y «artilugios» más comunes se están convirtiendo en inteligentes ([Smart cities](#), [IoT](#), ...), nuestra información es recopilada para una mejor experiencia en la red, para predecir acontecimientos, recomendarnos productos, etc. Podemos sentirnos como si realmente viviésemos en un **Gran Hermano global**, pero... si nos hace más mal que bien podríamos desconectar nuestros Smartphones, no tener Facebook, no utilizar tarjetas de crédito, ... Todo es opcional.

Para finalizar el **post** y el **año** os dejo con un artículo en el que se muestra [cómo cambiará nuestra vida en los próximos años con unos ejemplos](#).



Presente y Futuro de los Sistemas de Información Gerencial

Todos somos conscientes de la importancia de la información y la correcta gestión de ésta, sobre todo cuando se trata de desarrollar la actividad económica a nivel global; y existen diversos sistemas de información que permiten manejarla de la manera adecuada en favor de la propia empresa.

Pero al cobrar una importancia considerable, el sistema de información se trata de una tecnología presente en la gran mayoría de las empresas, principalmente aquellas con una actividad económica importante, por lo que debemos ir un paso más allá para mantener una ventaja competitiva y mantener la tecnología a la última, en la medida de lo posible. No debemos olvidar que la actualización tecnológica, es costosa, y a menudo laboriosa, y en ocasiones contraproducente, ya que al haberse habituado el personal a operar con cierto sistema, el cambio repentino de este puede suponer más inconveniente que ventaja. Pero al margen de esto, los sistemas deben adquirir diversas capacidades que los hagan mejores que la competencia, y que facilite la consecución de las estrategias de la empresa.

El principal reto en este ámbito, es la **automatización en la homogeneización y análisis** de todas las fuentes de información de la empresa. A menudo el sistema no da soporte a todos los requisitos de información, o se hace uso de diversas herramientas para obtener toda la información, con lo cual no se puede gestionar toda ella con relativa facilidad.

Por esto es que últimamente las empresas han tendido optimizar e integrar en sus sistemas todas estas fuentes, dirigiéndose a una tendencia de interconexión con otras empresas como podría ser el caso del ya antiguo IED o EDI en inglés, para intercambio electrónico de datos,

llegando a desarrollarse protocolos (OFTP), frameworks, plataformas (Orchestra, pimcore, ...) y estándares (ASC X12, EDIFACT, ...) que cumplan dicha función, llevando a hacer uso de las nuevas tecnologías y servicios de internet (WS-BPEL), para la integración, con conceptos en auge como pueden ser el cloud computing y el big data, y estrategias de orquestación comercial.

Numerosos gestores de negocios reconocen la gran influencia del "big data" en actividades pertenecientes a acciones de inteligencia competitiva, y es una necesidad imperiosa el tener un análisis automatizado antes de 2020.

Unido a todo lo anterior, cuya finalidad es facilitar la tarea de recopilar y almacenar toda la información que pudiera ser relevante para el negocio así como el intercambio de información entre las diferentes entidades involucradas, nos interesa sacar todo el partido posible, y no quedarnos en un mero análisis de ese gran volumen de información que se encuentra a nuestra disposición.

He ahí donde entra en juego el otro aspecto a tener en cuenta hoy en día, y en los años venideros para una empresa exitosa. Manejamos conceptos como pueden ser data mining, y artificial intelligence para usando complejas funciones matemáticas realizar la difícil labor de **correlación de datos**, y con la ayuda de los sistemas de inteligencia artificial, contrastar los datos y poder sacar patrones difícilmente identificables por las personas, teniendo en cuenta los datos de los que se disponen (aprendizaje supervisado), así como anticiparse a lo que está por venir.

Podemos ver la gran diversidad de aplicaciones que se le puede dar a la inteligencia artificial en el área de la empresa y los sistemas de información con ejemplos como:

- Descubrimiento de conocimiento en las bases de datos
- Estandarización, gestión y mejora de la calidad de producto
- Incremento del margen de beneficio
- Diseño y fabricación asistido por ordenador
- Anticipación del precio y la demanda
- Automatización de sistemas de control
- etc...

En definitiva, la tecnología está optimizando y facilitando la vida de los empresarios, así como en todos los ámbitos de nuestra sociedad, permitiendo delegar tareas triviales a las máquinas, en favor de una especialización por parte de las personas.

[Empresas que ya combatieron el fraude](#)

En este post se cuentan las experiencias de algunas empresas para combatir el fraude.

PayPal

Esta empresa necesita evitar los casos de fraude a través de una aplicación que pueda hacerlo en tiempo real. Este caso de Big data se debe a que ellos manejan información a través de PC, teléfonos, como medio de pago estándar en miles de sitios webs y muchas plataformas más que causan que l



os casos de fraude se acumulen cada vez más. Por este motivo es que necesitaban alguna herramienta para poder predecir dónde se podría llegar a cometer fraude antes que este comience a realizarse.

Con el objetivo de lograr paliar el fraude, realizan un análisis continuo y en tiempo real en grandes conjuntos de datos y aplican patentes de Google con la indexación de la web para transferirlo al problema de calcular el riesgo de fraude para los consumidores o comerciantes individuales

CSM-ACE

En la CSM-ACE de 2012, Neil Meikle (PwC) presentó el siguiente caso de estudio.

Una productora de televisión y compañía de emisión descubrió un fraude de facturación. El cliente sospechó de otras instancias de facturación falsa durante dos años. Durante este tiempo se realizaron 200.000 transacciones y 9.500 ventas. Estas transacciones exhibían un gran rango de valores y se desconocían qué transacciones eran fraudulentas. La solución a este problema se logró haciendo clustering de los proveedores.

Para ello se agrupó a los proveedores basándose en sus características. De esta forma se pudo localizar los diferentes valores "anómalos". Tras este análisis se encontraron 42 vendedores anómalos de los cuales dos fueron confirmados como fraudulentos