

# NOVÁTICA



CEPIS **UPGRADE**

Revista de la Asociación  
de Técnicos de Informática

Nº 211, mayo-junio 2011, año XXXVII

The cover features a central image of a person in a dark suit and black shoes, walking through a circular opening. The person is carrying a brown, worn leather briefcase in their right hand and holding a glowing yellow and green light strip in their left. The background is a complex, abstract pattern of concentric, overlapping circles in various colors (blue, green, red, yellow) that create a sense of depth and motion, resembling a tunnel or a data visualization. The overall lighting is dramatic, with strong highlights and deep shadows.

**Business Intelligence**

**Novática**, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de **ATI** (Asociación de Técnicos de Informática), organización que edita también la revista **REICIS** (Revista Española de Innovación, Calidad e Ingeniería del Software). **Novática** co-edita asimismo **UPGRADE**, revista digital de **CEPIS** (Council of European Professional Informatics Societies), en lengua inglesa, y es miembro fundador de **UPENET** (**UPGRADE** European Network).

<<http://www.ati.es/novatica/>>  
 <<http://www.ati.es/reicis/>>  
 <<http://www.cepis.org/upgrade>>

**ATI** es miembro fundador de **CEPIS** (Council of European Professional Informatics Societies) y es representante de España en **IFIP** (International Federation for Information Processing); tiene un acuerdo de colaboración con **ACM** (Association for Computing Machinery), así como acuerdos de vinculación o colaboración con **AdaSpain**, **AIZ**, **ASTIC**, **RITSI** e **Hispaniux**, junto a la que participa en **ProInnova**.

**Consejo Editorial**

Ignacio Aguiló Sousa, Guillem Aínsa González, María José Escalona Cuaremas, Rafael Fernández Calvo (presidente del Consejo), Jaime Fernández Martínez, Luis Fernández Sanz, Didac Lopez Viñas, Celestino Martín Alonso, José Onofre Montesa Andrés, Francesc Noguera Puig, Ignacio Pérez Martínez, Andrés Pérez Payeras, Viktu Pons i Colomer, Juan Carlos Vigo López

**Coordinación Editorial**

Llorenç Pagés Casas <pages@ati.es>

**Composición y autoedición**

Jorge Lloer Gil de Ramales

**Traducciones**

Grupo de Lengua e Informática de ATI <<http://www.ati.es/gt/lengua-informatica/>>

**Administración**

Tomás Brunete, María José Fernández, Enric Camarero, Felicidad López

**Secciones Técnicas - Coordinadores**

**Acceso y recuperación de la información**

José María Gómez Hidalgo (Optenet), <jmgomez@yahoo.es>

Manuel J. María López (Universidad de Huelva), <manuel.mana@diehsia.uhu.es>

**Administración Pública electrónica**

Francisco López Crespo (MAE), <flc@ati.es>

**Arquitecturas**

Enrique F. Torres Moreno (Universidad de Zaragoza), <enrique.torres@unizar.es>

Jordi Tubella Moragas (DAC-UPC), <jordit@ac.upc.es>

**Auditoría SITIC**

Marina Touriño Troitino, <marinatourino@marinatourino.com>

Manuel Palao García-Suelto (ATI), <manuel@palao.com>

**Derecho y tecnologías**

Isabel Hernando Collazos (Fac. Derecho de Donostia UPV), <isabel.hernando@ehu.es>

Elena Davara Fernández de Marcos (Davara & Davara), <edavara@davara.com>

**Enseñanza Universitaria de la Informática**

Cristóbal Pareja Flores (DSIP-UJM), <cpajef@dsip.ucm.es>

J. Angel Velázquez Hurtado (ULSI, URJC), <angel.velazquez@urjc.es>

**Entorno digital personal**

Andrés Marín López (Univ. Carlos III), <amarin@it.uc3m.es>

Diego Gachet Páez (Universidad Europea de Madrid), <gachet@uem.es>

**Estándares Web**

Encarna Quesada Ruiz (Virati), <encarna.quesada@virati.com>

José Carlos del Arco Prieto (TCP Sistemas e Ingeniería), <jcarco@gmail.com>

**Gestión del Conocimiento**

Juan Baiget Solé (Cap Gemini Ernst & Young), <juan.baiget@ati.es>

**Informática y Filosofía**

José Angel Olivares Varela (Escuela Superior de Informática, UCLM), <joseangel.olivares@uclm.es>

Karim Gherab Martin (Harvard University), <kgherab@gmail.com>

**Informática Gráfica**

Miguel Chover Selles (Universitat Jaume I de Castellón), <mchover@lsi.uji.es>

Roberto Vivó Hernando (Eurographics, sección española), <rvivo@dsic.upv.es>

**Ingeniería del Software**

Javier Dolado Cosin (ULSI-UPV), <dolado@lsi.uhu.es>

Daniel Rodríguez García (Universidad de Alcalá), <daniel.rodriguez@uah.es>

**Inteligencia Artificial**

Vicente Boti Navarro, Vicente Julián Inglada (DSIC-UPV), <(vbotti,vinglada)@dsic.upv.es>

**Interacción Persona-Computador**

Pedro M. Latorre Andrés (Universidad de Zaragoza, AIPO), <platorre@unizar.es>

Francisco L. Gutiérrez Vela (Universidad de Granada, AIPO), <fgutierrez@ugr.es>

**Lengua e Informática**

M. del Carmen Ugarte García (ATI), <cugarte@ati.es>

**Lenguajes Informáticos**

Oscar Belmonte Fernández (Univ. Jaime I de Castellón), <bellem@lsi.uji.es>

Inmaculada Coma Taty (Univ. de Valencia), <inmaculada.coma@uv.es>

**Lingüística computacional**

Xavier Gómez Guinovart (Univ. de Vigo), <xgg@uvigo.es>

Manuel Palomar (Univ. de Alicante), <mpalomar@dsi.ua.es>

**Mundo estudiantil y jóvenes profesionales**

Federico G. Mon Trotti (RITSI), <gmon.trotti@gmail.com>

Mikel Salazar Peña (Área de Jóvenes Profesionales, Junta de ATI Madrid), <mikelboni\_uni@yahoo.es>

**Profesión Informática**

Rafael Fernández Calvo (ATI), <rfdc@ati.es>

Miguel Sarrías Gilán (ATI), <msarrias@ati.es>

**Redes y servicios telemáticos**

José Luis Marzo Lázaro (Univ. de Girona), <joseluis.marzo@udg.es>

Juan Carlos López López (UCLM), <juancarloles@uclm.es>

**Robótica**

José Cortés Arenas (Sopra Group), <jccortesa@gmail.com>

Juan González Gómez (Universidad Carlos III), <juan@learobotics.com>

**Seguridad**

Javier Arellito Bertolin (Univ. de Deusto), <jarellito@deusto.es>

Javier López Muñoz (ETS Informática-UMA), <jlm@cc.uma.es>

**Sistemas de Tiempo Real**

Alejandro Alonso Muñoz, Juan Antonio de la Puente Alfaro (DIT-UPM), <faalonso@puente@dit.upm.es>

**Software Libre**

Jesus M. González Barahona (Universidad Politécnica de Madrid), <ismael.herraz@upm.es>

Israel Herráz Tabernerro (UAX), <isra@herraz.org>

**Tecnología de Objetos**

Jesus García Molina (DIS-UJM), <jmolina@um.es>

Gustavo Rossi (LIFA-UNLP Argentina), <gustavo@sol.info.unlp.edu.ar>

**Tecnologías para la Educación**

Juan Manuel Dodero Beardo (UC3M), <dodero@inf.uc3m.es>

César Pablo Córcoles Briongo (UOC), <ccorcoles@uoc.edu>

**Tecnologías y Empresa**

Didac López Vilas (Universitat de Girona), <dldic.lopez@ati.es>

Francisco Javier Gantús Sánchez (Infra Sistemas), <fgantus@infra.com>

**Tendencias tecnológicas**

Alonso Álvarez García (TID), <aad@tid.es>

Gabriel Martí Fuentes (Interbits), <gabi@atinet.es>

**TIC y Turismo**

Andrés Aguayo Maldonado, Antonio Guevara Plaza (Univ. de Málaga), <(aguayo, guevara)@cc.uma.es>

Las opiniones expresadas por los autores son responsabilidad exclusiva de los mismos. **Novática** permite la reproducción, sin ánimo de lucro, de todos los artículos, a menos que lo impida la modalidad de © o copyright elegida por el autor, debiéndose en todo caso citar su procedencia y enviar a **Novática** un ejemplar de la publicación.

**Coordinación Editorial, Redacción Central y Redacción ATI Madrid**  
 Padilla 66, 3º dcha., 28006 Madrid  
 Tlf: 91 4029391; fax: 91 3093685 <novatica@ati.es>  
**Composición, Edición y Redacción ATI Valencia**  
 Av. del Reino de Valencia 23, 46005 Valencia  
 Tlf: fax: 963330392 <secreval@ati.es>  
**Administración y Redacción ATI Cataluña**  
 Via Laietana 46, ppal. 1º, 08003 Barcelona  
 Tlf: 934125236; fax: 934127713 <secregen@ati.es>  
**Redacción ATI Aragón**  
 Lagasca 9, 3-B, 50006 Zaragoza.  
 Tlf: fax: 976235181 <secreara@ati.es>  
**Redacción ATI Andalucía** <secreand@ati.es>  
**Redacción ATI Galicia** <secregal@ati.es>  
**Suscripción y Ventas** <<http://www.ati.es/novatica/interes.html>>, ATI Cataluña, ATI Madrid  
**Publicidad**  
 Padilla 66, 3º dcha., 28006 Madrid  
 Tlf: 91 4029391; fax: 91 3093685 <novatica@ati.es>  
**Imprenta:** Derra S.A., Juan de Austria 66, 08005 Barcelona.  
**Depósito legal:** B 15.154-1975 - ISSN: 0211-2124; CODEN NOVAEC  
**Portada:** Resolución en marcha - Concha Arias Pérez / © ATI  
**Diseño:** Fernando Agresta / © ATI 2003

**editorial**

**El papel de las TIC en los movimientos sociales** > 02

**en resumen**

**Inteligencia de negocios en clave de presente** > 02

*Llorenç Pagés Casas*

**Noticias de IFIP**

**Reunión del TC-1 (Foundations of Computer Science)** > 03

*Michael Hinchey, Karin Breitman, Joaquim Gabarró*

**Reunión anual del TC-10 (Computer Systems Technology)** > 04

*Juan Carlos López López*

**Actividades de ATI**

**V Edición del Premio Novática** > 05

**monografía**

**Business Intelligence**

(En colaboración con **UPGRADE**)

*Editor invitado: Jorge Fernández González*

**Presentación. Business Intelligence: analizando datos para extraer**

**nueva información y tomar mejores decisiones** > 06

*Jorge Fernández González*

**Business Information Visualization: Representación de la información empresarial** > 08

*Josep Lluís Cano Giner*

**BI Usability: evolución y tendencia** > 16

*R. Dario Bernabeu, Mariano A. Garcia Mattio*

**Factores críticos de éxito de un proyecto de Business Intelligence** > 20

*Jorge Fernández González, Enric Mayol Sarroca*

**Modelos de construcción de Data Warehouses** > 26

*José María Arce Argos*

**Data Governance: ¿qué?, ¿cómo?, ¿por qué?** > 30

*Óscar Alonso Lombart*

**Business Intelligence y pensamiento sistémico** > 35

*Carlos Luis Gómez*

**Caso de estudio: Estrategia BI en una ONG** > 39

*Diego Arenas Contreras*

**secciones técnicas**

**Arquitecturas**

**Extensiones al núcleo de Linux para reducir los efectos del envejecimiento del software** > 43

*Ariel Sabiguero, Andrés Aguirre, Fabricio González, Daniel Pedraja, Agustín Van Rompaey*

**Derecho y tecnologías**

**La protección de datos personales en el desarrollo de software** > 50

*Edmundo Sáez Peña*

**Enseñanza Universitaria de la Informática**

**Reorganización de las prácticas de compiladores para mejorar el aprendizaje de los estudiantes** > 56

*Jaime Urquiza Fuentes, Francisco J. Almeida Martínez, Antonio Pérez Carrasco*

**Estándares Web**

**Especificación y prueba de requisitos de recuperabilidad en transacciones WS-BusinessActivity** > 61

*Rubén Casado Tejedor, Javier Tuya González, Muhammad Younas*

**Referencias autorizadas** > 70

**sociedad de la información**

**Informática práctica**

**Criptografía mediante algoritmos genéticos de una comunicación cifrada en la Guerra Civil** > 71

*Tomás F. Tornadijo Rodríguez*

**Programar es crear**

**El problema del decodificador**

(Competencia UTN-FRC 2010, problema C, enunciado) > 75

*Julio Javier Castillo, Diego Javier Serrano*

**Triángulo de Pascal y la Potencia Binomial**

(Competencia UTN-FRC 2010, problema E, solución) > 76

*Julio Javier Castillo, Diego Javier Serrano, Marina Elizabeth Cardenas*

**asuntos interiores**

**Coordinación editorial / Programación de Novática / Socios Institucionales** > 77

**Tema del próximo número: "Innovación y emprendimiento en Informática"**

## El papel de las TIC en los movimientos sociales

Encontramos constantemente, entre las noticias de actualidad, referencias directas o indirectas a actividades relacionadas con el mundo de la informática. Los grandes movimientos sociales, a los que asistimos como espectadores o participantes, no hubieran podido producirse sin el protagonismo indiscutible de las redes sociales, de las aplicaciones de distribución instantánea de contenidos, o de todo ese conjunto de aplicaciones que surge diariamente alrededor del mundo de la movilidad.

Los sistemas de información, que desde sus orígenes no han parado de evolucionar y extenderse en todos los niveles del tejido empresarial y económico, han irrumpido ahora también en el tejido social como elemento ya casi indispensable en la vida de muchas personas, en un proceso de evolución continua que genera diariamente nuevas maneras de entender y utilizar las tecnologías de la información, haciéndolas, en ocasiones, auténticas protagonistas en muchos de los acontecimientos sociales más relevantes.

Un reto y una oportunidad en un escenario caracterizado por un crecimiento exponencial en la diversidad de actores y herramientas al que hemos de adaptarnos, haciéndose necesario un punto de encuentro como ATI, donde damos protagonismo a quienes lideran los cambios y fomentamos un ecosistema sostenible en el que convivan armónicamente profesionales de todas las procedencias.

En línea con esta realidad se han creado recientemente en ATI varios grupos de interés (focalizados en mercado profesional senior, gobierno TI o internacionalización) que servirán para que profesionales de todas las edades amplíen su visión en las últimas tendencias.

Y creemos en este modelo de crecimiento profesional porque, inmersos en una carrera tecnológica donde lo de ayer ya no vale y el hoy es una versión beta del futuro, lo innovador es la norma y como tal no podemos esperar a que madure para adoptarlo.

Volviendo a los titulares de la actualidad, los medios de comunicación nos hablan de

movimientos en la sociedad civil que reclaman una transformación ética. Con un modelo aún por concretar, asambleario en origen, ha dejado huella en la política de muchos países, con especial transcendencia en Egipto y Túnez.

ATI, como asociación, también tiene objetivos sociales y desea ser útil a la sociedad. Por esto, apostamos tradicionalmente por mantener canales de comunicación con representantes políticos, sindicales, asociaciones empresariales, universidades y otras entidades docentes u otras asociaciones informáticas.

Veremos cómo influyen en España los movimientos nacidos alrededor del 15-M, pero más que esperar consecuencias directas en nuestro sector, constatamos que el modelo asociacionista se encuentra de nuevo en auge.

**La Junta Directiva General de ATI**

## en resumen Inteligencia de negocios en clave de presente

### Llorenç Pagés Casas

Coordinación Editorial de **Novática**

Es sabido que la raíz etimológica del nombre de nuestra revista nos obliga a apuntar hacia la presentación de "**nov**"edades relacionadas con nuestras tecnologías informáticas y con las prácticas necesarias para sacar provecho de ellas.

Esto, sin duda, nos condiciona en bastantes ocasiones a hablar en clave de futuro, ya que lo que es "**nov**"edad suele tardar un tiempo (aunque en estos tiempos "vertiginosos" pueda ser cada vez menos) en ser llevado a la práctica.

Sin embargo, en la monografía de este número quizás la principal novedad en cuanto a planteamiento es que vamos a hablar *en ra-bioso presente*.

Efectivamente, para hablarnos de "*Business Intelligence*" (un área de trabajo que en castellano podría denominarse "inteligencia de negocios" o quizás "inteligencia empresarial") y con la inestimable ayuda de nuestro editor invitado **Jorge Fernández González** (director de Consultoría en *Business Intelligence* de Abast Systems), hemos optado principalmente por seleccionar un grupo de destacados profesionales para que nos cuenten sus

experiencias, vivencias y lecciones aprendidas sacadas directamente del ámbito empresarial.

Y es que los sistemas y herramientas de los que hablamos en esta monografía son directamente herederos de los "antiguos" sistemas de ayuda a la toma de decisiones o sistemas decisionales que se empezaron a concebir hace ya varias décadas a partir de la aparición de las hojas de cálculo y de las herramientas gráficas de presentación.

A todos nos habrán contado alguna vez que el modo de construcción de esos sistemas decisionales (por contraposición con los transaccionales) suele ser bastante más artesanal de lo habitual. Y así podríamos decir que cada decisor y cada decisión "son un mundo", hasta llegar a la conclusión de que seguramente el mejor método para ponerse al día en este tema es acercarse a los expertos de la empresa e irles preguntando uno a uno, "¿Qué tal te va a tí?".

Personalmente, pienso que nos hemos de felicitar por haber conseguido brillantes y bien elaboradas "respuestas" por parte de todos

los autores. Quede constancia por ello de nuestro agradecimiento.

Este número se completa además, en el resto de bloques, con una rica selección de artículos que tratan desde interesantes reportes de investigación procedentes de congresos en los que ATI es entidad colaboradora, hasta otros contenidos de carácter más práctico que en este caso se concretan en un artículo de **Edmundo Sáez Peña** conteniendo interesantes recomendaciones sobre el desarrollo de sistemas que cumplan con los requisitos de protección de datos personales, y un instructivo ejercicio de descifrado de mensajes con connotaciones históricas a cargo de **Tomás F. Tornadizo Rodríguez**.

Espero que disfrutéis de esta amplitud de temas... como preámbulo por supuesto de mayores disfrutes estivales.

Un fuerte abrazo,



# Reunión del TC-1 (Foundations of Computer Science)

Michael Hinchey<sup>1</sup>, Karin Breitman<sup>2</sup>, Joaquim Gabarró<sup>3</sup>

<sup>1</sup>Lero, the Irish Software Engineering Research Center, presidente del TC-1; <sup>2</sup>Pontificia Universidade do Rio de Janeiro, secretaria del TC-1; <sup>3</sup>ALBCOM, Universitat Politècnica de Catalunya, representante de ATI en el TC-1

El TC-1 se reunió en la Universidad del Sarre, Saarbrücken (Alemania) el 3 de abril de 2011. Ante todo, Michael Hinchey dio la bienvenida a los asistentes e invitados. Explicó que Francia se ha reincorporado de nuevo a IFIP.

Se trataron los siguientes puntos:

1) Se informó sobre la organización de los próximos WCC (*World Computer Congress*). Dicha conferencia ha sido deficitaria durante los últimos cinco años. Entre las causas se cita el precio de la inscripción, la existencia de múltiples sesiones paralelas en distintos tópicos y la falta de interés por los eventos comunes. Como consecuencia, IFIP esta replanteándose su formato. Se piensa en organizarlo contando con la participación de profesionales de la informática y políticos (Comisión Europea) de primer nivel. El tema giraría alrededor de una sociedad de la información segura, fiable e innovadora: “*Las tecnologías de la información juegan un papel clave en la sociedad global del siglo 21. La tecnología informática es cada vez más un motor de la economía, facilita la innovación en muchas industrias y organizaciones gubernamentales, entre otras cosas, apoya a la democracia y al desarrollo personal...*”. Maurice Nivat preguntó en qué medida la sociedad de la información constituye realmente un apoyo (y no lo contrario) a la democracia. Tras debatir estos temas se sugirió que los grupos TC-9, TC-1.7 y TC-11 podrían estudiar este punto.

2) Luego pasó a discutirse el futuro de la conferencia TCS (*Theoretical Computer Science*) organizada por TC-1 y hasta ahora ligada a WCC. Michael Hinchey plantea la posibilidad de organizar TCS en Dagstuhl. Jos Baeten sostiene que aunque hay varias conferencias en el área de TCS no hay ninguna que cubra toda la teoría. Con los años TCS ha visto una gran asistencia. Cree que hay espacio para organizar TCS con algún otro evento, sugiere organizarla en 2012 coincidiendo (pero como evento independiente) con el WCC. Todos coinciden en que TCS debe proseguir.

3) Jose Fiadeiro planteó la cuestión de la evaluación de las conferencias y la clasificación de TCS. Actualmente TCS está en la

categoría C, sobre todo debido al hecho de que se ve como parte de WCC. Para asegurar la amplitud de temas, Andrzej Tarlecki comentó que no cree que una conferencia que cubra temas muy amplios se pueda organizar siguiendo el modelo tradicional. También cree que la co-localización con otros eventos es secundaria. Joaquim Gabarró comenta que sería interesante un debate sobre el futuro de las Ciencias de la Computación. ¿Qué nos gustaría que fueran en el futuro? Quizá esto aumentaría el interés de los participantes evitando ponentes que solo están interesados en presentar su trabajo y que luego desaparecen.

4) En el siguiente punto se presentaron los informes de los distintos grupos de trabajo. Cathy Meadows del WG 1.7 informó que Taska es ahora un evento principal en ETAPS. Tienen 3 años para demostrar la viabilidad.

5) Las finanzas fueron el siguiente punto. Mike comenta que varios eventos no entendieron que tenían que pagar a IFIP (10 euros/día, 5 euros/día en casos difíciles). De estas cuotas el 75% va al TC y 25% para financiar IFIP. Mike también comenta que este modelo está siendo reconsiderado ya que plantea numerosos problemas. Recordó que cuando el evento no paga tiene que hacerlo el TC. Otra fuente de ingresos es a través de las publicaciones. El nuevo modelo, para las publicaciones de la serie IFIP, tiene un rendimiento de 3 euros por página publicada, donde el 25% se destina a la cooperación técnica y el 75% restante a IFIP. Se mencionó que este modelo puede incentivar la aceptación de trabajos largos de calidad dudosa a fin conseguir más dinero.

6) Karin Breitman presentó la versión preliminar del nuevo sitio web de TC-1. Los miembros tendrán acceso a fin de poder mandar comentarios y sugerencias. Se comentó que la manera de ordenar a los miembros, según naciones, siguiendo el criterio IFIP quizá no sea el ideal.

7) El siguiente punto trató de los criterios necesarios para ser miembro de un TC. Mike informó que IFIP comprobará si los miembros de categoría A asisten regularmente.

También se sugirió que los miembros de categoría C se deberían reelegir cada dos años.

8) Las próximas reuniones del TC-1 tendrán lugar en Estonia (marzo de 2012) coincidiendo con ETAPS y en Países Bajos (septiembre de 2012) coincidiendo con TCS.

9) Finalmente, como resultado de las discusiones planteadas, se decidió crear tres paneles o grupos de trabajo sobre los temas: Educación, Evaluación y Financiación. El grupo de Educación se centrará en la educación informática en secundaria y debería trabajar conjuntamente con el TC-3. El grupo de Evaluación se inspira en la comunidad de matemáticos que ha organizado grupos para estudiar los distintos índices y su importancia en la evaluación de investigadores individuales y de los departamentos. Jos Baeten mencionó que en <http://www.informatics-europe.org/> se puede encontrar un buen informe. El grupo de Financiación se centrará en el hecho que los presupuestos de los departamentos son cada vez menores. Como consecuencia la investigación depende cada vez más en la industria, poniendo en peligro la investigación pura. Otra cuestión a tratar es el incremento constante de carga de trabajo académico, y el impacto negativo que esto tiene sobre las actividades de trabajo voluntario, como por ejemplo, revisión de trabajos y organización de eventos.

# Reunión anual del TC-10 (Computer Systems Technology)

Juan Carlos López López

Catedrático de Universidad, Universidad de Castilla-La Mancha; representante de ATI en el TC 10

<juancarlos.lopez@uclm.es>

El pasado día 15 de marzo de 2011 se celebró en Grenoble (Francia) la reunión anual del Comité Técnico 10 (TC 10, *Computer Systems Technology*) de IFIP. Esta reunión tuvo lugar coincidiendo con el congreso DATE (*Design, Automation and Test in Europe*). Asistieron a la reunión el *Chairman* de este comité y representante de Suiza, Prof. Bernhard Eschermann, los representantes de Brasil, Alemania, Dinamarca y República Checa, y los coordinadores de los WG 10.4 y 10.5.

Tras la aprobación de la agenda de la reunión y del acta de la sesión anterior, se pasa a informar de distintas acciones consecuencia de la reunión anterior. En particular, el Prof. Franz Rammig informa sobre la decisión de que el congreso BICC (*Biologically Inspired Cooperative Computing*), organizado por el TC 10 y ligado hasta el momento al WCC (*World Computer Conference*), se celebre, sin embargo, en 2012 junto al DIPES (*Distributed and Parallel Embedded Systems*). También anuncia la decisión de organizar finalmente el WCC 2012 con un planteamiento y esquema completamente diferentes a los seguidos hasta el momento.

Asimismo, el *Chairman*, Prof. Eschermann, informa sobre la nueva política de patrocinios de congresos que llevará a cabo IFIP. Los patrocinios que impliquen participar en beneficios o pérdidas, la mayoría de los cuales llevan consigo la firma de acuerdos con otras organizaciones como IEEE y ACM, se aprobarán a partir de ahora desde la secretaría de IFIP, en un intento de llevar más control y darle más valor a la marca "IFIP".

Respecto al estado de cuentas del TC 10, se aprueba tanto el resumen financiero de 2010 como los presupuestos de 2011.

A continuación se pasa a informar sobre las actividades de los WGs (*Working Groups*). Se discute en concreto sobre la necesaria reactivación de los mismos. La nueva responsable del WG 10.5, Prof. Dominique Borrione, informa de que una de las acciones principales para llevar a cabo esa reactivación será la renovación de los miembros del WG. El Prof. Ricardo Reis informa sobre las actividades que se han llevado a cabo en Brasil, lugar en

el que se han localizado últimamente conferencias de casi todos los WGs.

Con el propósito de avanzar en los nuevos temas de interés propuestos en reuniones anteriores, se asignan inicialmente *Cloud Computing* al WG 10.4 y *Green Computing* al WG 10.5.

Seguidamente, se abre un debate sobre la forma de trabajar en el TC 10. Se pone sobre la mesa la necesidad de reuniones técnicas que permitan reunir a representantes de los distintos WGs con la finalidad de aportar nuevas ideas y proyectos enmarcados en sus áreas de investigación. La organización del WCC 2012 puede ser una buena oportunidad para contribuir con estas nuevas líneas de trabajo, aunque la lista de temas técnicos planteados hasta ahora por los organizadores de este congreso resulta muy limitada en las áreas de interés del TC 10.

Habiendo finalizado el periodo preceptivo de mandato de Vicepresidente y Secretario, se plantea la elección de los mismos. El Prof. Ricardo Reis es reelegido Vicepresidente mientras que, dada la ausencia del Prof. Paolo Prinetto, actual Secretario, se decide posponer la elección y realizarla más adelante electrónicamente.

Se propone que la próxima reunión se celebre en el marco del WCC 2012 (el lugar sugerido inicialmente para este congreso es Amsterdam), siempre y cuando sea factible la organización de una sesión técnica con los objetivos planteados anteriormente. En caso contrario, y con el fin de conseguir una mayor participación, se planteará una audio/vídeo conferencia. Con esta decisión se pone final a la reunión.

## Ramon López de Mántaras obtiene el Robert S. Engelmores Award 2011

Ramon López de Mántaras, director del Instituto de Investigación en Inteligencia Artificial del CSIC y representante de ATI en el TC 12 de IFIP, ha sido galardonado con uno de los premios más prestigiosos del mundo en el ámbito de la Inteligencia Artificial, el Robert S. Engelmores Award 2011.

Es la primera persona de fuera de los EEUU que recibe este premio, otorgado por la Asociación Americana de Inteligencia Artificial y la revista AI Magazine.

López de Mántaras ha sido premiado por el conjunto de su trayectoria, y especialmente por sus investigaciones pioneras para crear máquinas capaces de aprender a partir de sus propias "experiencias".

El premio Robert S. Engelmores, que se concede desde 2003, se caracteriza por su política de distinguir a los grandes pioneros de la investigación en Inteligencia Artificial. En los últimos años has resultado ganadores, entre otros, Edward Feigenbaum y Bruce Buchanan (inventores de los sistemas expertos) y Jim Hendler (inventor de la Web Semántica junto con Tim Berners-Lee).

Ramon, en nombre de ATI recibe nuestra más cordial enhorabuena por este merecido premio.

# V Edición del Premio *Novática*

La V Edición del Premio *Novática*, destinado al mejor artículo publicado en 2010 de nuestra revista ha tenido como ganador el artículo "*AVBOT: Detección y corrección de vandalismos en Wikipedia*", del que es autor **Emilio José Rodríguez Posada**.

En el artículo se describe AVBOT, un robot creado para proteger Wikipedia en español de ciertas modificaciones no deseadas, conocidas como "vandalismos". Aunque en un principio AVBOT fue desarrollado para Wikipedia, es posible utilizarlo en cualquier sitio web que emplee el motor MediaWiki. Desde que está en funcionamiento ha demostrado una alta efectividad.

**Emilio José Rodríguez Posada**, un joven profesional informático socio de ATI, es Ingeniero Técnico en Informática de Sistemas y colaborador de la Oficina de Software Libre y Conocimiento Abierto de la Universidad de Cádiz.

El artículo fue publicado en la Sección "*Mundo estudiantil y jóvenes profesionales*" del



**Emilio José Rodríguez Posada**

número 203 (enero-febrero de 2010), en una edición especial de dicha sección dedicada a los proyectos presentados al Concurso Universitario de Software Libre 2009, del que ATI fue colaborador, y será publicado en inglés en la revista **UPGRADE** dentro del año en curso.

El artículo está disponible en: <<http://www.ati.es/novatica/2010/203/203-51-V-Premio-Novatica-ganador.pdf>>.

Los artículos finalistas, listados por orden de apellido del primer autor, han sido:

■ "*Crear vínculos basados en el contexto para mejorar las experiencias de los turistas*", de **Carlos Lamsfus, Christoph Grün, Aurkene Alzua-Sorzabal y Hannes Werthner**.

■ "*El test de Turing: historia y significado*", de **Giuseppe O. Longo**.

■ "*El contenido de la Profesión Informática: una visión personal*", de **Fernando Piera Gómez**.

■ "*Tendencias en Procesamiento del Lenguaje Natural y Minería de Textos*", de **Javier Pueyo y José Antonio Quiles Follana**.

Jorge Fernández González  
Director de Consultoría en Business  
Intelligence de Abast Solutions

<jfernand@abast.es>

# Presentación. *Business Intelligence*: analizando datos para extraer nueva información y tomar mejores decisiones

La simplicidad del concepto bajo el que se esconde *Business Intelligence* es tanta que a veces nos resulta difícil de percibir. Este concepto no es otro que el de "incertidumbre". No saber las cosas, no tener la certeza de que es lo que ha pasado y que es lo que pasará a nuestras empresas y organizaciones. Tal y como lo define la Real Academia de la Lengua Española.

certeza. (De cierto).

1. f. Conocimiento seguro y claro de algo.
2. f. Firme adhesión de la mente a algo conocible, sin temor de errar.

Y no es más que eso lo que buscamos bajo la amalgama de conceptos vinculados a *Business Intelligence*, lidiar con la incertidumbre y con el miedo a no poder controlar los procesos de nuestra organización. Y es ahora, en el contexto actual, con la crisis económica y con la incertidumbre de cuando finalizará, es en este momento cuando los proyectos de *Business Intelligence* están más en auge que nunca, y la razón no es otra que el miedo, el miedo a errar, el miedo a no tener toda la información para tomar una decisión, que ahora sí, en estos momentos, puede determinar la supervivencia o la extinción de nuestra empresa o de nuestro modelo de negocio.

Pero antes de dar un paseo por todo lo que hay que tener en cuenta en todo proyecto de *Business Intelligence*, me gustaría incidir en el segundo término de nuestro bonito concepto. El *Business Intelligence*, es *Business* pero no es seguro que es inteligente.

inteligencia.

(Del lat. *intelligentia*).

1. f. Capacidad de entender o comprender.
2. f. Capacidad de resolver problemas.
3. f. Conocimiento, comprensión, acto de entender.

Las tres primeras acepciones de la definición de inteligencia, vemos claro que no las podemos aplicar a un sistema de información, el que es inteligente es el analista de negocio, nunca el software o la plataforma, por eso algunos están adoptando una nueva terminología sustituyendo *Business Intelligence* por "*Business Analytics*". Término que no me acaba de convencer tampoco, ya que quizás yo soy un nostálgico de aquella primera definición no tan comercial en la que a los sistemas que nos ocupan se les llamaba "*sistemas*

## Editor invitado

**Jorge Fernández González** desarrolla su actividad profesional en dos ámbitos de actuación, el principal de ellos es como profesional de Sistemas de Información. En el que acumula más de 14 años de experiencia y actualmente ejerce el cargo de Director de Consultoría en *Business Intelligence* de AbastSolutions. El segundo de sus ámbitos de actuación es el de docente universitario. Actualmente es Profesor Asociado al departamento de ESSI de la UPC y responsable de la asignatura "Sistemas de Información para las Organizaciones" de la Facultad de Informática de Barcelona. También ha sido colaborador docente en la Universitat Oberta de Catalunya (UOC), director académico y profesor en masters y postgrados de la Fundación Politécnica y realiza sesiones como invitado en escuelas de negocio como ESADE, La Salle y EAE. En el ámbito de las publicaciones, ha colaborado con la revista *Data TI* en múltiples ocasiones, fue redactor de la revista *Gestión del Rendimiento*, es miembro del grupo de expertos en español del portal *BeyeNetwork*, portal que reúne a los especialistas en BI a nivel mundial, y colabora como experto en todas las ediciones del *BARC Business Intelligence Guide*.

de ayuda a la toma de decisiones", que es lo que realmente son y que por mucho marketing que le pongan los fabricantes, es lo que siempre serán. Pero como no queremos volver a los 80, al final hemos llamado a este número especial *Business Intelligence*. Espero sepáis perdonar los anglicismos que nos hemos visto obligados a utilizar.

¿Pero cómo te ayudan estos sistemas a tomar decisiones? Pues es fácil solo hay dos vías, recopilando datos sobre los que trabajar y dotando a los usuarios de negocio de herramientas para interrogarlos de forma sencilla y útil.

El recorrido al que os invitamos en este especial de *Business Intelligence*, es precisamente el contrario, del cerebro del analista al dato del que se nutre.

Cuando estamos creando un reporte, ponemos sus métricas y sus gráficos explicativos, hemos estado trabajando en él durante mucho tiempo y vemos que hemos aportado gran cantidad de información a una problemática del negocio que antes era cuanto menos difusa, nos encontramos con el hecho de que todo ese trabajo está realizado desde un punto de vista subjetivo, el nuestro. Es más, a partir de ese "fantástico" informe que vamos a presentar y que tiene un claro camino analítico que nos lleva a conclusiones concretas, nos podemos encontrar con que éstas no sean tenidas nunca en cuenta a la hora de tomar decisiones en nuestra organización. ¿Cómo es eso posible?

Yo fui consciente de este hecho hará unos años, realizando una consultoría, evaluando los posibles escenarios finales y los resulta-

dos que se podrían derivar de que un determinado indicador/métrica, apareciera en rojo en el cuadro de mando. Mis argumentos eran que si sucedía ese hecho, era un claro indicador de que estaba pasando un hecho X y que teníamos obligatoriamente que tomar la decisión Y. A lo que un amigo mío me comentó que la decisión Y nunca se tomaría fuese cual fuese el informe, porque iba en contra de la creencia de la empresa.

¿Qué pasa si de pronto te dicen que la Tierra es plana de nuevo?, ¿qué pasa si la información que recibes entra en frontal conflicto con tu paradigma decisional?, ¿qué pasa si esa información hace que tengas que tomar una decisión que está en total disonancia con lo que tú realmente piensas?, ¿La tomarás igualmente o cambiarás el peso que esta información tiene en tu toma de decisiones?. Eso es lo que se llama disonancia cognitiva y que se resume como:

*"Es el conflicto mental que abunda en la experiencia cuando se presentan evidencias de que una creencia propia o asunción personal es incorrecta. La teoría de la disonancia cognoscitiva afirma que hay una tendencia en la gente a reaccionar para reducir tal disonancia. Una persona puede evitar la nueva información o convertirse manipulador de argumentos para mantener su creencia o juicios".*

Es en este contexto y no en otro, para qué engañarnos, en el que desarrollamos nuestro trabajo, aunque siempre tendemos a pensar que los decisores van a sacar el máximo provecho de nuestro trabajo y de los sistemas que desarrollamos. Al menos éste es nuestro propósito y es en este sentido positivista en el que presentamos los artículos de esta monografía.

El cerebro del analista tiene diferentes formas de entender o desechar la información que se le presenta, es por esto que es muy importante como se la presentamos. Por esta razón abrimos este especial con un artículo del profesor de ESADE **Josep Lluís Cano Giner**, titulado "*Business Information Visualization*: Representación de la información empresarial".

Pero una vez que se nos ha presentado esa información tenemos que tener herramientas que nos permitan usarla de forma eficiente, sobre esto nos hablarán desde Argentina los autores **R. Darío Bernabeu** y **Mariano A. García Mattío** en su artículo "*BI Usability*: evolución y tendencias".

Aunque para llevar a cabo un proyecto de BI sin contratiempo, lo mejor que podemos hacer es mirar los factores críticos de éxito del BI, para intentar repetirlos en nuestros proyectos, eso es lo que abordo junto con el profesor **Enric Mayol Sarroca** de la UPC en nuestro artículo homónimo.

Sin embargo, todos los grandes proyectos de BI tienen que tener una sólida base, siendo el diseño del *Data Warehouse* una parte fundamental. Para comprender mejor las raíces del BI contamos con el excelente artículo de **Jose María Arce Argos** reputado especialista con muchos años de experiencia y profesor asociado de ESIC, titulado "*Modelos de construcción de Data Warehouses*".

Pero no podemos construir información sin tener una política de datos, sin saber los orígenes sobre los que construimos, es por esto que **Oscar Alonso Llombart**, analista de Penteo y con una larga trayectoria en

proyectos de BI, nos muestra todo lo que tenemos que tener en cuenta a la hora de gobernar los "ladrillos" de la información, en su artículo "*Data Governance, ¿qué?, ¿cómo?, ¿por qué?*".

Pero no todos seguimos los mismos caminos, y **Carlos Luis Gómez** de Ibertia nos muestra uno nuevo en su artículo "*Business Intelligence y pensamiento sistémico*", un artículo para leer pensando.

Y por último, desde Chile, **Diego Arenas Contreras** comparte su experiencia real, en un caso de estudio de "*Estrategia BI en una ONG*", explicando como BI puede ayudar a organizaciones sin ánimo de lucro.

Esperamos que este especial sea de vuestro agrado y que os despierte las ganas de seguir aprendiendo sobre este extraño concepto llamado *Business Intelligence*.

### Nota del editor

Como decíamos en la introducción de este número ("En resumen", página 2), no son muchas las ocasiones en las que, como ha ocurrido en esta monografía, podemos conseguir que un elenco de expertos trabajando en la industria escriban para contarnos, casi diríamos que "en vivo y en directo", sus vivencias actuales sobre el tema de que se trate. La reflexión que viene a continuación está motivada por la cuestión de la cantidad de anglicismos contenida en los distintos artículos y que el propio editor invitado comenta en esta presentación.

No sin antes haberlo pensado bien, e incluso haberlo comentado en los foros de ATI, hemos creído oportuno no proceder a ningún tipo de corrección lingüística de los términos contenidos en los artículos respetando al 100% los modos de expresarse de los autores, puesto que entendemos que reflejan con precisión los modos de expresión y las terminologías usadas hoy en día en la práctica empresarial. Sin embargo, esperamos que el lector encuentre en las "Notas" añadidas a cada artículo debida respuesta a las dudas terminológicas que le puedan surgir a partir del uso de esas palabras de la jerga técnica que no le resulten conocidas o habituales.

Por último comentar que no va a ser éste siempre el mismo criterio que usemos en **Novática** en materia de usos lingüísticos. Para monografías y temas de carácter más teórico y "futurista" es lógico que pidamos a los autores que envíen sus artículos considerando el uso de sinónimos en castellano de los términos que aparezcan procedentes de otros idiomas. De ahí, la nota que aparece en la página 77 ofreciendo explícitamente la ayuda que el Grupo de Lengua e Informática de ATI puede aportar a los autores en este sentido.

### Referencias útiles sobre "*Business Intelligence*"

Las referencias que se citan a continuación, junto con las proporcionadas en cada uno de los artículos, tienen como objetivo ayudar a los lectores a profundizar en los temas tratados en esta monografía permitiendo contrastar ideas y obtener información actualizada.

#### Libros

- **J. M. Curto.** *Introducción al Business Intelligence*. Barcelona: Editorial, UOC, 2010.
- **J.L. Cano.** *Business Intelligence: competir con información*. Madrid: Fundación cultural Banesto, 2007.
- **W.H. Inmon.** *Building the Data Warehouse*. Hoboken: John Wiley & Sons, 4th Edition, 2005.
- **B. Kimball.** *Data Warehouse Toolkit Classics*. Hoboken: John Wiley & Sons, 2009.

#### Portales

- **TodoBI.** <<http://todobi.blogspot.com/>>.
- **DataPRIX.** <<http://www.dataprix.com/>>.

- **BI-SPAIN.com.** <<http://www.bi-spain.com>>.
- **BeyeNetwork Spain.** <<http://www.beyenetwork.es>>.

#### Blogs

- **Josep Curto Díaz.** *Information Management*, <<http://informationmanagement.wordpress.com>>.
- **Pau Urquizu.** *BI Fácil*, <<http://www.businessintelligence.info>>.
- **Rémi Grossat.** *Inteligencia de negocio*, <<http://www.intelineg.com>>.
- **Salvador Ramos.** *SQL Server Sí!*, <<http://www.sqlserversi.com>>.
- **Jose María Arce Argos.** *Business Intelligence Blog (BIB)*, <<http://josemariaarce.blogspot.com/>>.
- **Aníbal Goicochea.** *Tecnologías de la información y estrategia*. <<http://anibalgoicochea.com/>>.
- **Jorge Fernández González.** *Sistemas*

*decisionales, algo mas que BI.* <<http://sistemasdecisionales.blogspot.com>>.

#### Instituciones

- **The Data Warehouse Institute (TDWI).** <<http://www.tdwi.org>>.
- **The MDM Institute.** <<http://www.tcdii.com/>>.
- **The Data Governance Institute.** <<http://www.datagovernance.com>>.
- **Inmon Institute.** <<http://inmoninstitute.com/>>.
- **Kimball University.** <<http://www.kimballgroup.com/html/kimballu.html>>.

Josep Lluís Cano Giner

Profesor del Departamento de Dirección de Sistemas de Información, ESADE, Universidad Ramon Llull

<josepluis.cano@esade.edu>

## Business Information Visualization: Representación de la información empresarial

### 1. Introducción

Las aplicaciones utilizadas por las organizaciones cada vez generan mayores cantidades de información. Esta información se maneja en tiempo real o casi real. Generar conocimiento y tomar decisiones son las principales razones por las que las organizaciones almacenan la información, además de soporte a las operaciones y para cumplir las obligaciones legales. Ambas razones dependen del criterio de las personas, que deben extraer de la visualización de la información presentada aquellos aspectos clave que permitan reconocer patrones o tendencias ocultas. La visualización se convierte así en la interficie entre los ordenadores y las mentes de las personas. Las capacidades cognitivas de los humanos tienen limitaciones, por visualización se entiende el proceso de transformación de datos, información y conocimiento en una representación para ser usada de manera afín a las capacidades cognitivas de los humanos.

### 2. Ejemplos de la historia de la visualización de información

Diversos autores han investigado sobre la historia de la visualización de la información. Entre ellos destaca Tufte [1]. Desde tres autores distintos, se han seleccionado tres ejemplos correspondientes a tres representaciones con el fin de mostrar tanto las bondades como los inconvenientes del uso de la visualización de la información.

En 1786, el ingeniero escocés William Playfair se dio cuenta de que las transacciones económicas se podían representar fácilmente de manera gráfica. Además, según su opinión, la representación en series temporales y en diagramas de barras simplificaba su comprensión y retención. El autor publicó el "Commercial and Political Atlas" donde se representaba el comercio exterior de Inglaterra; libro que, además y por primera vez, incluía un nuevo tipo de gráfico: el diagrama de pastel. En la **figura 1** se muestra un gráfico de las exportaciones e importaciones entre Inglaterra y Dinamarca junto con Noruega [2]. En éste claramente se señala el momento en que el signo del saldo de la balanza comercial entre los países junto con el crecimiento del saldo a favor de Inglaterra cambia.

Uno de los ejemplos más famosos de presentación de información pertenece a Charles Minard, un ingeniero civil francés que usó la

**Resumen:** Los directivos y directivas cada vez disponen de más información y menos tiempo para acceder a ella, ya que deben tomar decisiones rápidamente. Su correcta representación se puede convertir en una pieza clave a la hora de facilitar la toma de decisiones. En el artículo, se parte de una revisión de la historia y de la importancia de la visualización de la información. Asimismo, se muestra un ejemplo de cómo mejorar dicha visualización y se concluye con las nuevas necesidades surgidas en torno a este aspecto, requeridas tanto por las organizaciones como por sus directivos.

**Palabras clave:** Business Intelligence, representación gráfica, Sistemas de Información Directivos, visualización de información.

#### Autor

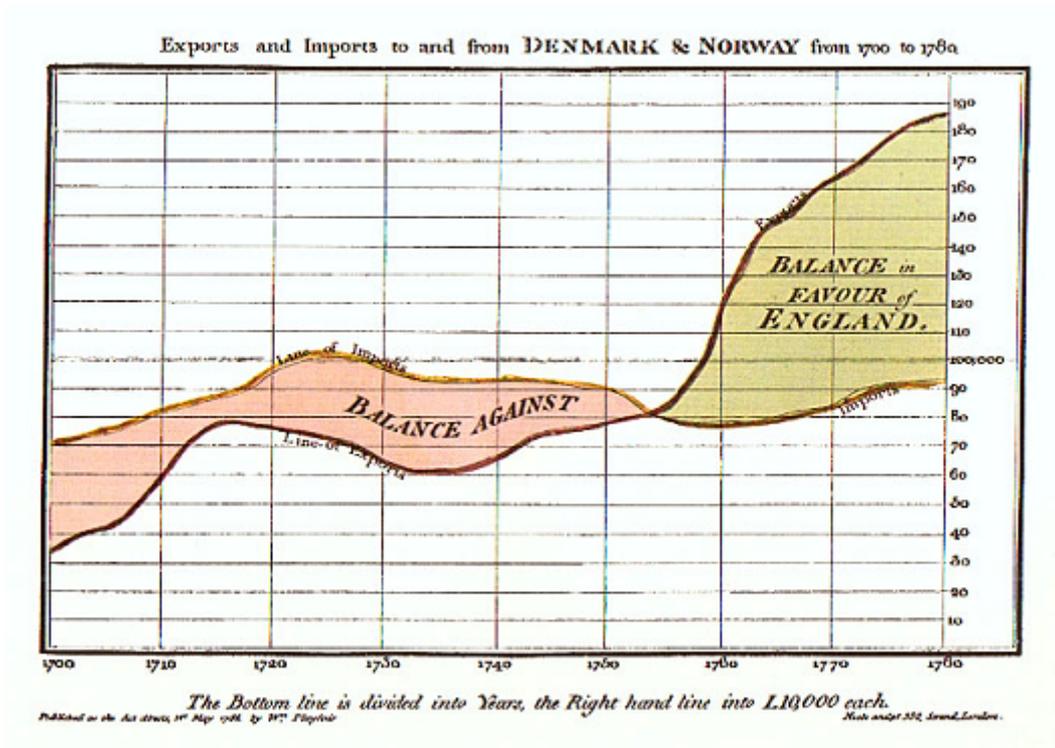
**Josep Lluís Cano Giner** es profesor del Departamento de Dirección de Sistemas de Información de ESADE desde el año 1989. Licenciado en Ciencias Empresariales y máster en Dirección de Empresas (ESADE). Licenciado en Administración y Dirección de Empresas (Universidad Politécnica de Catalunya). Obtuvo el Diploma de Estudios Avanzados (DEA) por la UPC. En la actualidad, investiga en el campo de los factores que inciden en el aumento del uso de los Sistemas de Información Directivos por parte de los directivos. Es consultor de empresas con una dilatada experiencia en la planificación estratégica de sistemas de información y en selección de soluciones, especialmente en *Business Intelligence*. Entre sus publicaciones académicas se encuentran artículos y diversos *Business Case*, entre ellos: "Government Intelligence en las universidades públicas catalanas", "Aventis", "Aguirre&Newman", "Bodegas Torres", "Raventos i Blanc at a crossroad". También es el autor del libro "*Business Intelligence: Competir con Información*".

visualización para mostrar la historia de la trágica marcha a Moscú de Napoleón<sup>1</sup> en 1812. En la gráfica de la **figura 2**, mostró en una barra de color (en sombreado claro para el lector de esta revista), cuyo grosor indicaba el tamaño del ejército (originalmente de 422.000 efectivos), cómo este se iba reduciendo a medida que se acercaba a la capital rusa. A la vez, otra barra, de color negro, indicaba las tropas que regresaron de Moscú (a la llegada eran tan solo 10.000 efectivos). En la base del gráfico se anotaron las temperaturas al aire libre, que fueron el principal problema de los soldados. En el medio del gráfico, se puede observar cómo la barra de color negro se ensancha, debido a la incorporación de unas tropas rezagadas que habían intentado avanzar por el flanco izquierdo, así como la dramática disminución del grosor al cruzar un río con aguas heladas. Al final del retorno, se puede comparar el grueso de las dos barras: la de color, los que partieron; y la negra, los que regresaron. Un simple gráfico nos enseña, de una forma muy potente, la historia. Robert Spence [3] se pregunta si se podría escuchar la obertura n.º 1812 de Tchaikovsky<sup>2</sup> combinada con la visualización del gráfico.

En 1954, Darrell Huff publicó "*How to Lie with Statistics*" [4] en el que exponía cómo se podía manipular la representación gráfica de

las estadísticas para evidenciar diferentes intereses, algunas veces contrapuestos. Evidentemente su gran aportación fue enseñarnos a hacerlo correctamente. En la **figura 3** ofrecemos un ejemplo de la representación de un gráfico de líneas muy útil para mostrar tendencias o predicciones. En el eje de las abscisas (el horizontal) se indican los meses del año; mientras que en el de las ordenadas (el vertical) se señala el volumen, por ejemplo, de ventas en billones de dólares. En el gráfico de la izquierda la representación de la información es la correcta, el eje de las ordenadas comienza en 0 y las distancias entre los valores de los dos ejes son equivalentes; por su parte, en el gráfico de la derecha, el eje de las ordenadas comienza en 20, con lo que la expresión del personaje que aparece en el gráfico se vuelve de profunda admiración por los resultados obtenidos.

Al repasar tres de los ejemplos históricos de la visualización de la información se ha subrayado, en el primer caso, la importancia de la representación de la información para comprender qué sucede; en el segundo, que una buena representación de la información nos permite comprender mejor una realidad; y en el tercero, que si la representación de la información está manipulada, intencionadamente o no, nos puede llevar a una interpretación



**Figura 1.** Exportaciones e Importaciones de y para Dinamarca y Noruega desde 1700 a 1780. Fuente: W. Playfair.

errónea de los hechos. Si la representación es la adecuada podremos tomar decisiones pero, ¿qué sucedería si alguien hubiera manipulado la representación con otros fines?

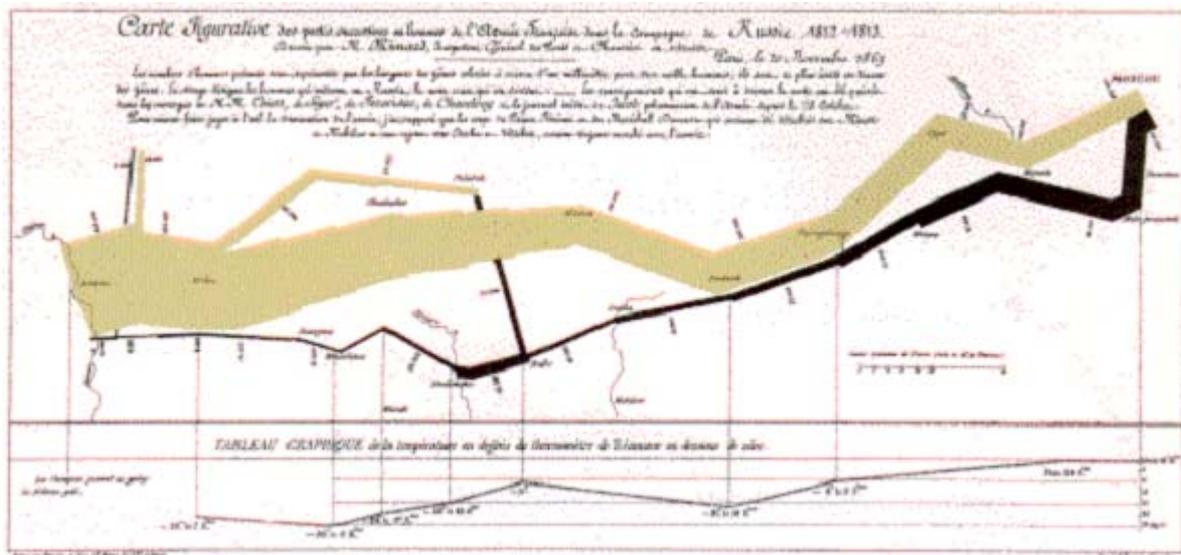
**3. Visualizando información**

La visualización de la información se utiliza en distintos campos, desde la medicina, pasando por la ingeniería, la estadística, los negocios, o incluso en el deporte. De este último, hemos elegido la **figura 4**, publicada en un reconocido periódico nacional<sup>3</sup>, con el

fin de demostrar la dificultad que la presentación de la información en un gráfico alberga o puede albergar. En el gráfico publicado se muestran los récords de los 100 metros a lo largo de la historia: los atletas, su nacionalidad, la fecha en la que se consiguieron y las marcas obtenidas. En el gráfico aparecen también unas barras grises junto a la representación de un corredor. El lector se puede sorprender al ver que la barra más larga está junto al menor tiempo. A mí me ocurrió lo mismo.

Después de analizar el gráfico durante un tiempo me di cuenta de que lo que el autor intentaba representar era cómo hubiera sido la llegada si en la carrera hubieran participado los 10 corredores que ostentan el récord del mundo de los 100 metros (Carl Lewis y Leroy Burrell lograron dos récords, por lo tanto deberían correr en dos calles).

Trabajando sobre el gráfico anterior, se podrían proponer algunas mejoras, como por ejemplo añadir la línea de llegada en el gráfico



**Figura 2.** Mapa de las fuerzas de Napoleón en la campaña de Rusia. Fuente: Charles Minard, 1861.

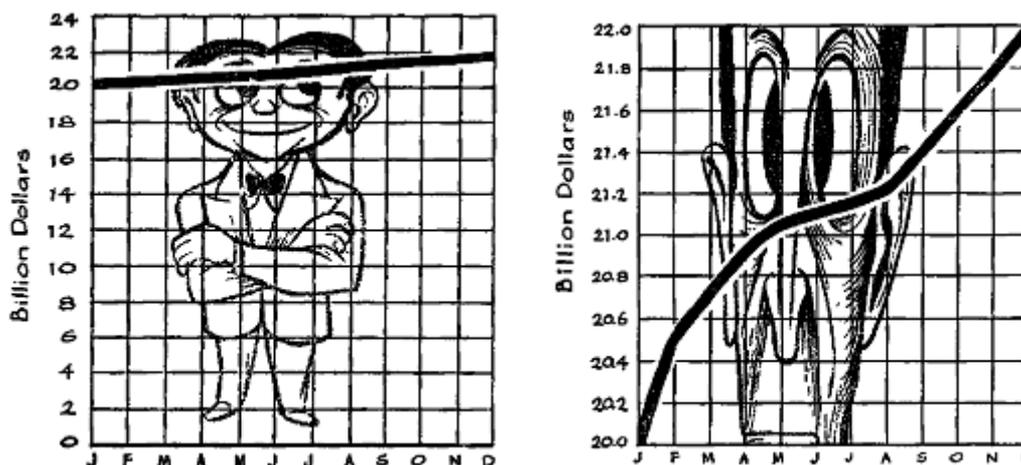


Figura 3. Gráfico en el que se muestra como cambia la figura al cambiar la escala del eje.  
Fuente: Darrell Huff.

y modificar la posición de los tiempos de las marcas para facilitar su lectura y comprensión. En la figura 5 se propone una posible solución.

En el supuesto de que la representación propuesta facilite la interpretación de la información, podríamos cuestionarnos ahora si la distancia representada entre los corredores corresponde a la real. A partir de la información proporcionada podemos calcular que la distancia entre el primer y el décimo récord sería de 1,81 metros, es decir, en el tiempo de

9,77 segundos en que Asafa Powell recorre 100 metros, Jim Hines hubiera recorrido 98,19 metros. Así pues, ¿la distancia entre el corredor y la línea de meta corresponde a 1,81 metros? Para responder a esta pregunta recurrimos a representar las distancias valiéndonos de los gráficos<sup>4</sup> de una hoja de cálculo. En la figura 6 se muestra el resultado obtenido.

A simple vista parece que las distancias entre los corredores han decrecido en relación con el gráfico original. ¿Qué ha sucedido? En el

gráfico original no se presentan los valores del eje de abscisas, por lo que podríamos preguntarnos en qué valor comienza. En la figura 7 se muestran dos gráficos. El de la izquierda comienza en 97 metros, y el de la derecha en 95 metros.

Si cambiamos el límite inferior de las abscisas, al visualizar las gráficas, la persona que las está analizando puede interpretarlas de forma distinta, lo que le podría llevar a tomar una decisión diferente.

■ EVOLUCIÓN DEL RÉCORD DE 100 METROS

Atleta	Fecha	Marca
Asafa Powell (JAM)	14-6-2005	9,77s
Tim Montgomery (EE UU)	14-9-2002	9,78s
Maurice Greene (EE UU)	16-6-1999	9,79s
Donovan Bailey (CAN)	27-7-1996	9,84s
Leroy Burrell (EE UU)	6-7-1994	9,85s
Carl Lewis (EE UU)	25-8-1991	9,86s
Leroy Burrell (EE UU)	14-6-1991	9,90s
Carl Lewis (EE UU)	24-9-1988	9,92s
Calvin Smith (EE UU)	3-7-1983	9,93s
Jim Hines (EE UU)	14-10-1968	9,95s

Figura 4. Gráfico de la evolución de los récords de los 100 metros.  
Fuente: El País, 15/06/2005

■ EVOLUCIÓN DEL RÉCORD DE 100 METROS

Atleta	Marca	Fecha	Meta
Asafa Powell (JAM)	9,77s	14-6-2005	1
Tim Montgomery (EE UU)	9,78s	14-9-2002	2
Maurice Greene (EE UU)	9,79s	16-6-1999	3
Donovan Bailey (CAN)	9,84s	27-7-1996	4
Leroy Burrell (EE UU)	9,85s	6-7-1994	5
Carl Lewis (EE UU)	9,86s	25-8-1991	6
Leroy Burrell (EE UU)	9,90s	14-6-1991	7
Carl Lewis (EE UU)	9,92s	24-9-1988	8
Calvin Smith (EE UU)	9,93s	3-7-1983	9
Jim Hines (EE UU)	9,95s	14-10-1968	10

Figura 5. Propuesta de gráfico de la evolución de los récords de los 100 metros. Fuente: Elaboración propia

También cabe preguntarse si la distancia recorrida es, efectivamente, la mejor variable para representar las diferencias entre los distintos récords del mundo de los 100 metros lisos. O, si bien, podemos recurrir a la velocidad, es decir, a los metros por segundo para evidenciar la diferencia entre los distintos corredores. En la figura 8 se muestra la gráfica de las velocidades que se obtuvieron en los 10 récords. En este caso solo se presenta la gráfica en la que el eje de abscisas comienza

con el valor 0, ya que si comenzara por otros valores nos sucedería lo mismo que en el caso anterior. La diferencia en metros por segundo es de tan solo 0,19 entre el primero y el último.

Vayamos un poco más allá en el análisis. Si retomamos la perspectiva histórica nos podríamos preguntar ¿cuál ha sido la evolución de los 10 mejores récords del mundo de los 100 metros lisos a lo largo del tiempo? Podemos recurrir para ello a un gráfico que

nos indique los valores a lo largo del tiempo. El que nos propone por defecto la hoja de cálculo es el mostrado en la figura 9.

De nuevo, una lectura rápida nos podría llevar a una conclusión errónea: ha mejorado mucho ya que la pendiente es pronunciada. Pero si representamos en el eje de las ordenadas el valor 0, el gráfico cambia significativamente (ver figura 10).

Nos podría parecer que en el gráfico anterior no se ha representado ninguna marca. Sin embargo, si se observa con atención puede percibirse que hay una línea entre los valores de 9 y 10 segundos, con una pendiente negativa pequeña a lo largo de aproximadamente unos 37 años. Es decir, se han tardado casi 37 años en reducir la marca a 18 centésimas de segundo, lo que equivaldría a unas 5 milésimas de segundo por año. Parece que la segunda gráfica representa mejor esta realidad que la primera. En el límite, dos representaciones diferentes de los mismos datos nos pueden llevar a dos conclusiones que incluso se pueden contradecir entre ellas.

Hemos conseguido, con la representación gráfica, que aquellas personas que estaban interesadas en el hecho a analizar hayan comprendido realmente lo que ha sucedido en los últimos 10 récords del mundo de 100 metros lisos hasta el año 2005, las pocas diferencias que hay entre ellos y la dificultad por mejorarlos. Tal vez, hubiera sido mejor echar mano de otro tipo de representación gráfica: la tabla

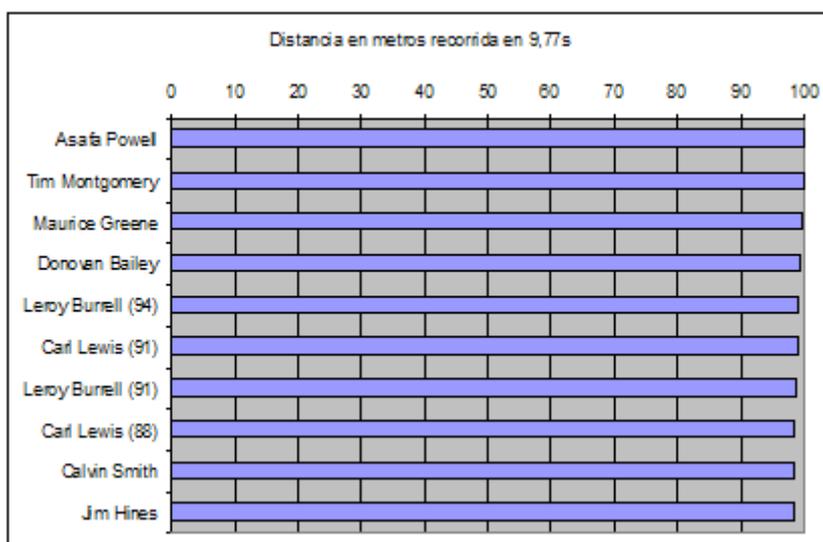


Figura 6. Gráfico de la evolución de los récords de los 100 metros, comenzando en 0 metros. Fuente: Elaboración propia

Fecha	Record	Atleta	Nacionalidad	Marca (segundos)	Mejora de tiempo %	Velocidad (metros/segundo)	Distancia en metros recorrida en 9,77s
14/06/2005	1	Asafa Powell	JAM	9.77	0.00%	10.24	100.00
14/09/2002	2	Tim Montgomery	EEUU	9.78	0.10%	10.22	99.90
16/06/1999	3	Maurice Greene	EEUU	9.79	0.20%	10.21	99.80
27/07/1996	4	Donovan Bailey	CAN	9.84	0.72%	10.16	99.29
06/07/1994	5	Leroy Burrell (94)	EEUU	9.85	0.82%	10.15	99.19
25/08/1991	6	Carl Lewis (91)	EEUU	9.86	0.92%	10.14	99.09
14/06/1991	7	Leroy Burrell (91)	EEUU	9.90	1.33%	10.10	98.69
24/09/1988	8	Carl Lewis (88)	EEUU	9.92	1.54%	10.08	98.49
03/07/1983	9	Calvin Smith	EEUU	9.93	1.64%	10.07	98.39
14/10/1968	10	Jim Hines	EEUU	9.95	1.84%	10.05	98.19

Tabla 1. Evolución de los récords de los 100 metros. Fuente: Elaboración propia

	Objetivos de los gráficos	Gráfico original de los récords de los 100 metros lisos
1	Sugiere que solo debemos mostrar la información. Algunas veces los que diseñan los gráficos tienden a mostrar agregaciones de información en lugar de la información en sí misma.	Los países de los corredores están junto a los nombres. Si se les hace aparecer en una columna separada facilita la lectura y se muestra mejor la preponderancia de los corredores de EEUU.
2	Sugiere que debemos asegurarnos de que el usuario piense en lo esencial del gráfico, y no en el gráfico en sí mismo.	Para interpretar el gráfico era necesario reconocer que se intentaba representar la llegada a la meta de una carrera hipotética entre los corredores de los 10 últimos récords. Se debería indicar la posición del récord.
3	Evitar todas aquellas decoraciones innecesarias.	La representación de los corredores no es necesaria.
4	Comprimir tanta información como sea posible en el menor espacio posible.	Se hubieran podido incluir los metros recorridos o la velocidad de cada récord.
5	Los gráficos deben ser diseñados para animar al usuario a hacer comparaciones entre las distintas informaciones.	Las marcas, al no estar alineadas, dificultan su comparación.
6	Los gráficos deben proveer vistas de la información a distintos niveles de detalle.	Al ser un único gráfico y no interactivo, no aplica.

Tabla 2. Principios de Tufte y análisis del gráfico de los récords. Fuente: Elaboración propia

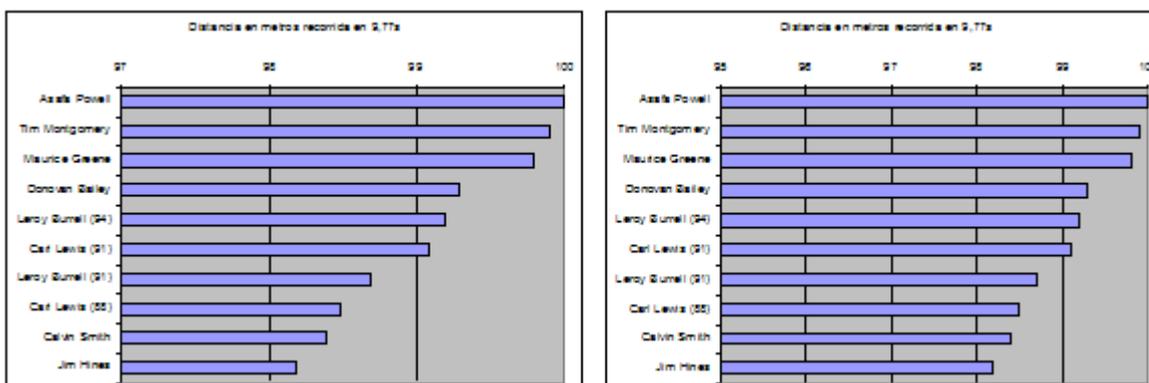


Figura 7. Gráfico de la evolución de los récords de los 100 metros, comenzando en 97 y de 95 metros. Fuente: Elaboración propia

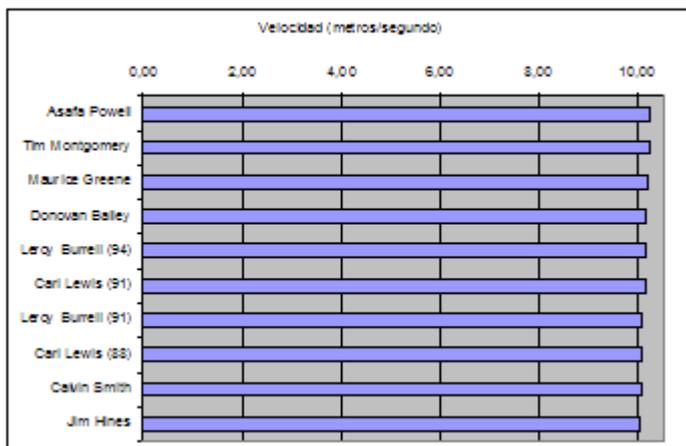


Figura 8. Gráfico de la velocidad de los récords de los 100 metros. Fuente: Elaboración propia

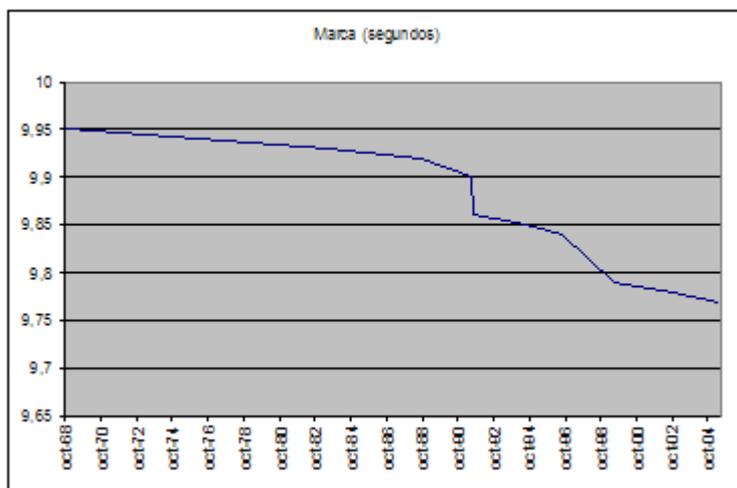


Figura 9. Gráfico de la evolución de los récords de los 100 metros, comenzando en 9,65 segundos. Fuente: Elaboración propia

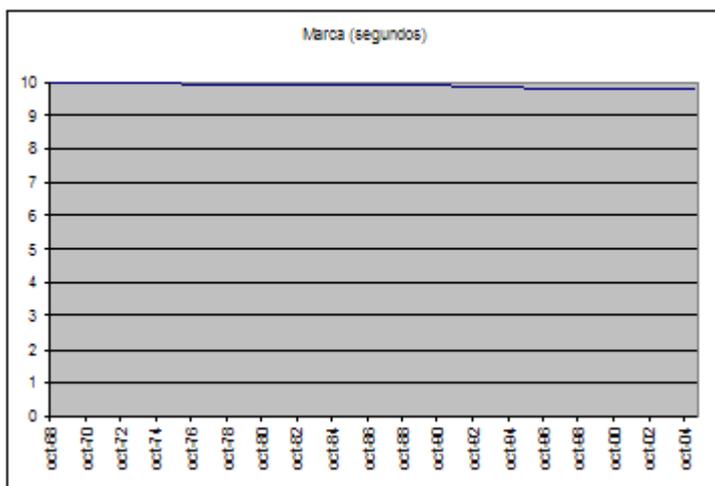


Figura 10. Gráfico de la evolución de los récords de los 100 metros, comenzando en 0 segundos. Fuente: Elaboración propia

con los valores y los cálculos utilizados para representarlos gráficamente. En la **tabla 1** se han añadido tres nuevas columnas: la de la mejora del tiempo en porcentaje respecto al mejor récord, la velocidad en metros por segundo, y la de la distancia que habrían recorrido el resto de corredores en el tiempo que Asafa Powell habría llegado a la meta.

Si la conclusión es que la mejor forma de comprender esta realidad es la tabla con los valores, entonces resulta la mejor representación. Esta decisión probablemente se relacione con una elección personal, a la vez que se encuentra influida por el conocimiento que la persona disponga sobre los resultados de las pruebas de 100 metros lisos. Es decir, no depende tan solo del que representa la información, sino también del que la visualiza. Nos podríamos preguntar si siempre tiene sentido establecer un formato de informe invariable a lo largo del tiempo, como deciden la mayoría de organizaciones, basándose en que al no cambiar el formato facilita su interpretación o deberíamos modificarlo para conseguir una mejora en la visualización si quisiéramos representar mejor algunos cambios que se han producido en los datos.

#### 4. Information Visualization

Según Card et al. [5], la visualización de la información se define como: "El uso de representaciones visuales basadas en ordenadores e interactivas de información abstracta para amplificar la cognición".

Los autores la diferencian de la visualización científica, normalmente basada en información física.

Los mismos autores han llevado a cabo una revisión de la literatura<sup>6</sup>, a la vez que justifican cómo la visualización amplifica la cognición, o dicho de otra manera, que el concepto de cognición (del latín: *cognoscere*, "conocer") hace referencia a la facultad de los seres de procesar información a partir de la percepción, el conocimiento adquirido y las características subjetivas que permiten valorarla.

Debemos señalar que en la definición propuesta se añade el término "interactivas", ya que en la actualidad, normalmente, las representaciones siempre se basan en el uso de ordenadores que permiten la interacción entre el usuario y la aplicación del ordenador.

Para aquellos lectores que quieran profundizar en la definición de "visualización", en las distintas tecnologías que la soportan o en la relación entre las teorías cognitivas y las tareas de la resolución de problemas, así como en las representaciones visuales, pueden hacerlo en el artículo de Tegarden [6].

En dicho artículo, Tegarden resume los seis

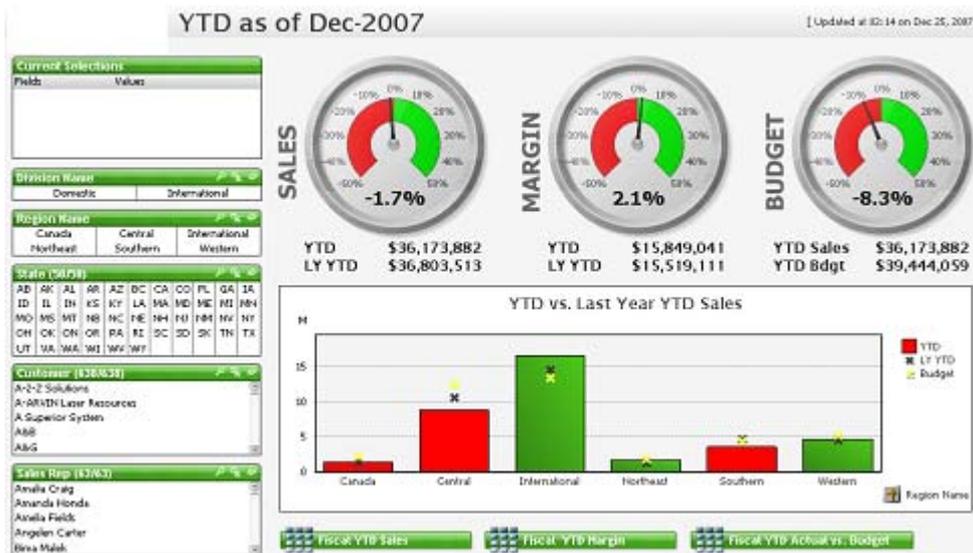


Figura 11. Cuadro de mando. Fuente: Ejemplo de QlikView.

objetivos que todo gráfico debería cumplir según Tufte [1][7]. En la **tabla 2**, dichos objetivos se relacionan con el gráfico original de los récords de los 100 metros lisos que hemos utilizado como ejemplo:

**5. Business Information Visualization**

Los directivos necesitan información para la toma de decisiones, y necesitan que se presente de forma que facilite su interpretación. Normalmente, las organizaciones desarrollan proyectos de *Business Intelligence*. Uno de los aspectos clave de estos proyectos es la correcta representación de la información. En el presente artículo, no nos vamos a referir a los conceptos básicos de la representación de la información. Para ello, se adjuntan referencias bibliográficas de distintas obras que tratan amplia y profundamente este tema. No obstante, sí que vamos a incidir en las nuevas tendencias y necesidades sobre la visualización. Entre los autores cabe destacar a Stephen Few [8]. Según él, la visualización de infor-

mación del futuro conlleva nuevas necesidades, entre ellas:

- **Dashboards o Scorecards:** los directivos necesitan acceder a información que en poco tiempo les permita analizar cuál es la situación, de manera que una vez detectado el problema, mediante unas pocas pulsaciones con el ratón, puedan descender al nivel de detalle suficiente para poder comprender qué sucede y tomar las medidas correctoras. En los *scorecards* se representan perspectivas de áreas estratégicas, objetivos, medidas y indicadores semafóricos, mientras que en los *dashboards* la información presentada puede variar mucho y se suelen incluir representaciones gráficas. En los *dashboards* o cuadros de mando la complejidad de la visualización de la información aumenta, ya que pueden presentarse informaciones o gráficos interrelacionados. Normalmente se debe, además, presentar una gran cantidad de información en un espacio muy limitado (ver **figura 11**).

- **Geo espacial:** cuando la información se basa en el territorio, su visualización sobre el terreno se hace cada vez más necesaria. Desde los conocidos *Geographic Information Systems (GIS)*, hasta Google Earth o distintos *web services* encontramos medios que nos permiten relacionar, por ejemplo, las ventas o los gastos con el territorio.

- **Scatterplots** o gráficos animados: en algunos casos necesitamos comparar dos magnitudes, por ejemplo, inversiones en publicidad y ventas a lo largo del tiempo. Para ello es necesario valerse de un nuevo tipo de representación que incluye animaciones en el transcurso temporal. Uno de los mejores ejemplos de representación animada de información es el caso de [www.GapMinder.org](http://www.GapMinder.org), en el que, por ejemplo, se puede ver la relación entre los ingresos de los países y su tasa de mortalidad infantil a lo largo del tiempo.

- **Treemaps:** un ejemplo de representación de este tipo de gráficos es el volumen de operaciones de la bolsa de Nueva York, agregado por industria y de forma que permite comparar los precios y sus cambios desde el día anterior (puede verse en [www.SmartMoney.com](http://www.SmartMoney.com)).

- **Sparklines:** son un tipo de representación gráfica caracterizada por su pequeño tamaño y la densidad de información que contiene. Normalmente se utilizan varias a la vez representando distinta información que en algunos casos puede ser complementaria. El término *sparkline* fue propuesto por Edward Tufte, que las describe<sup>7</sup> como "pequeños gráficos, simples, de elevada resolución y del tamaño de una palabra". En la **figura 12** se muestra un ejemplo de un *sparkline*.

- **Representación de relaciones:** en algunos casos, necesitamos representar relaciones entre entidades, como en el caso de las *Web sites*. Cada una de ellas actúa como nodo en una

		Close	Max	Min
AT&T		40,28	41,34	33,30
Boeing		98,15	100,59	84,79
Citigroup		53,98	55,20	48,27
Exxon Mobil		85,94	85,94	69,56
General Electric		38,12	38,12	34,09
General Motors		34,66	36,20	28,85
Intel		24,24	24,24	18,76
Microsoft		30,49	31,11	26,63

Figura 12. Tabla de evolución de las cotizaciones. Fuente: <<http://www.edwardtufte.com>>.

red, y tiene vínculos con otras. Un ejemplo de uso es Vizster de redes entre personas.

## 6. Conclusión

Las necesidades de visualización de información han ido cambiando a lo largo de la historia. El tiempo del que disponen los directivos y directivas para tomar decisiones cada vez es menor. Además, han aparecido nuevas necesidades con lo que los investigadores han propuesto y propondrán nuevas soluciones para cubrirlas. Una correcta representación de la información debería facilitar su interpretación y disminuir el tiempo que los directivos invierten en ella, y este se presenta como el principal objetivo de la visualización de información.

Si la visualización de la información no es la correcta puede llevar a los directivos y directivas a tomar decisiones equivocadas. A lo largo del artículo se ha presentado distintos ejemplos en los que la representación de información de una forma "manipulada" podría llevar a interpretaciones erróneas, con lo que el riesgo que los directivos y las directivas se equivoquen aumentan. En todo proyecto de *Business Intelligence* deberíamos asegurar que la representación gráfica es la más adecuada, por ello necesitamos especialistas que nos lo aseguren minimizando la visualización de información. Sin una correcta representación, no conseguiríamos aportar el valor que se espera, y difícilmente retendremos el interés de los directivos y las directivas en el uso de esta solución.

## Referencias

- [1] **E.R. Tufte.** *The Visual Display of Quantitative Information*. Cheshire: Graphics Press, 1983.
- [2] **W. Playfair, H. Wainer, I. Spence.** *The Commercial and Political Atlas and Statistical Breviary*. New York: Cambridge University Press, 2005.
- [3] **R. Spence.** *Information Visualization*. Essex: ACM Press, 2001.
- [4] **D. Huff.** *How to Lie with Statistics*. Penguin; New Ed edition, 1991.
- [5] **S.K. Card, J.D. Mackinlay, B. Shneiderman.** *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann Publishers, 1999.
- [6] **D. P. Tegarden.** "Business Information Visualization". *Communications of AIS*, vol.1, art. 4, enero, 1999.
- [7] **E.R. Tufte.** *Envisioning Information*. Cheshire: Graphics Press, 1990.
- [8] **S. Few.** *Data Visualization, Past, Present, and Future*. Perceptual Edge, 2007.

## Bibliografía

- J. Best.** *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*. Berkeley and Los Angeles: University of California Press, 2001.
- J. Best.** *More Damned Lies and Statistics: How Numbers Confuse Public Issues*. Berkeley and Los Angeles: University of California Press, 2004.
- W.S. Cleveland.** *The Elements of Graphing Data*. Summit: Hobart Press, 1994.
- S. Few.** *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland: Analytics Press, 2004.
- S. Few.** *Information Dashboard Design: The Effective Visual Communication of Data*. Sebastopol, CA: O'Reilly Media, Inc., 2006.
- S. Few.** *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Oakland: Analytics Press, 2009.
- J.G. Koomey.** *Turning Numbers into Knowledge: Mastering the Art of Problem Solving*. Oakland: Analytics Press, 2001.
- D. Niederman, D. Boyum.** *What the Numbers Say: A Field Guide to Mastering Our Numerical World*. New York: Broadway Books, 2003.
- N.B. Robbins.** *Creating More Effective Graphs*. Hoboken: John Wiley and Sons, Inc., 2005.
- T. Segaran, J. Hammerbacher.** *Beautiful Data*. O'Reilly Media, 2009.
- E.R. Tufte.** *Visual Explanations*. Cheshire: Graphics Press, 1997.
- E.R. Tufte.** *Beautiful Evidence*. Cheshire: Graphics Press, 2005.
- C. Ware.** *Information Visualization: Perception for Design*, second edition. San Francisco: Morgan Kaufmann Publishers, 2004.
- D.W. Wong.** *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures*. New York: W. Norton & Company, 2010.

## Notas

- <sup>1</sup> <<http://www.math.yorku.ca/SCS/Gallery/images/minard.gif>>.
- <sup>2</sup> El lector puede hacer el ejercicio accediendo a <<http://www.youtube.com/watch?v=k-vQKZFF-9s&feature=related>> (último acceso 5.1.2011).

Esta abertura, Op. 49, fue compuesta para conmemorar la victoriosa resistencia rusa en 1812 frente al avance de la Grande Armée de Napoleón Bonaparte. La obra fue estrenada en Moscú el 20 de agosto de 1882. La obra es reconocida por su final triunfal, que incluye una salva de disparos de cañón y repique de campanas.

<sup>3</sup> Artículo "Huracán Powell" aparecido en *El País* el miércoles 15 de junio de 2005. Al pie del gráfico se indica el literal "elaboración propia".

<sup>4</sup> En todos los gráficos se ha omitido la presentación de los valores para facilitar la comprensión del efecto que se quiere mostrar con el gráfico.

<sup>5</sup> Sobre la pregunta, debemos señalar que no nos estamos preguntando ¿cuál ha sido la evolución de los récords del mundo de los 100 metros lisos a lo largo de la historia? Para responder a esta pregunta deberíamos disponer de todos los récords del mundo, y no es el objetivo de este artículo analizarlos.

<sup>6</sup> Tabla 1.3 de la página 16 del libro citado en la referencia [5].

<sup>7</sup> El lector puede acceder a más ejemplos en: <[http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg\\_id=00010R&topic\\_id=1](http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00010R&topic_id=1)>.

R. Dario Bernabeu, Mariano A. García Mattío  
Cofundadores de Grupo eGlu Business Intelligence (eGluBI), Córdoba (Argentina)

<{DarioSistemas,magm3333}@gmail.com>

## 1. Introducción

Lo que hoy se conoce con el nombre de *Business Intelligence* (BI), ha tenido un origen y evolución que cabe resaltar a fin de introducirnos en el concepto que será tratado en este artículo: "*BI Usability*".

Uno de los objetivos principales del BI es que los usuarios, encuentren la información que necesitan para tomar decisiones en tiempo y "forma". La "forma" incluye, entre otras cosas, el formato en que la información se presenta y el nivel de interacción pretendido para obtener el resultado deseado. Son los puntos anteriores los que conforman el término "*BI Usability*".

*Usabilidad* puede definirse como la facilidad de uso de un software, en la que también intervienen factores como la familiaridad del diseño, la comodidad, si es atractivo al usuario o no, el nivel de interacción que permite, el tiempo de respuesta, etc.

Se han seleccionado varias definiciones de usabilidad para complementar el concepto<sup>1</sup>.

■ La ISO/IEC 9126 define la usabilidad como "*la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso*". Asimismo, la ISO establece cuatro principios básicos en los que se basa la usabilidad: facilidad de aprendizaje, facilidad de uso, flexibilidad y robustez.

■ Jakob Nielsen, padre de la usabilidad, la define como "*un atributo de calidad que evalúa cuán fácil para su uso son las interfaces de usuario*".

■ Janice (Ginny) Redish, consultora independiente, define lo que debería permitir una interfaz a los usuarios: "*encontrar lo que necesitan, entender lo que encuentren y actuar apropiadamente, dentro del tiempo y esfuerzo que ellos consideren adecuado para esa tarea*".

*Business Intelligence* puede definirse como un concepto que integra, por un lado el almacenamiento y por el otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración. Dicho conocimiento debe ser oportuno, relevante, útil y debe estar adaptado al contexto de la organización<sup>2</sup>.

En el marco de estas aproximaciones conceptuales sobre usabilidad y BI, es posible propo-

# BI Usability: evolución y tendencia

**Resumen:** Este artículo nos introduce inicialmente en los conceptos de usabilidad y Business Intelligence, para luego definir el término "*BI Usability*". Seguidamente, presenta un gráfico histórico con los hitos más significativos, que son antecedentes de lo que en la actualidad se conoce como sistemas BI. Luego, se sistematizan estos hitos, década por década, teniendo en cuenta por un lado, la evolución/innovación en los sistemas de información BI y por el otro destacando la usabilidad de esa época. Finalmente, describe y traza la forma en que *BI Usability* se fue desarrollando a través del tiempo, y señala tendencias en cuanto a usabilidad.

**Palabras clave:** BI, evolución histórica, usabilidad.

## Autores

**R. Dario Bernabeu** es Ingeniero en Sistemas por el Instituto Universitario Aeronáutico de Argentina. Ha sido cofundador de eGluBI <www.eglubi.com.ar>. Se especializa en el desarrollo e implementación de soluciones OSBI (*Open Source Business Intelligence*): gestión de proyectos, análisis de requerimientos/necesidades, despliegue y configuración de soluciones BI, confección de procesos de integración de datos, modelado de Data Warehouse, confección de cubos multidimensionales y *Business Models*, desarrollo de reportes ad hoc, reportes avanzados, análisis interactivos, *dashboards*, etc. Es docente, investigador, *geek* y entusiasta del software libre. Su publicación más destacada es "*Data Warehousing: Investigación y Sistematización de Conceptos – HEFESTO: Metodología para la Construcción de un DW*". Es coordinador de la red social Red Open BI <www.redopenbi.com> y realiza numerosos aportes en diferentes foros, wikis, blogs, etc.

**Mariano A. García Mattío** es Ingeniero en Sistemas por el Instituto Universitario Aeronáutico (IUA) de Argentina y Especialista en Sistemas y Servicios Distribuidos por la Universidad Católica de Córdoba (UCC). Profesor titular de: "Bases de Datos 1" y "Bases de datos 2" en el IUA – Facultad de Ingeniería; "Motores de Bases de Datos 2" en el IUA – Facultad de Administración; "Paradigma de Programación Orientada a Objetos" y "Sistemas Distribuidos en la Especialización en Sistemas Embebidos" del IUA. Jefe de trabajos prácticos de "Bases de Datos Aplicadas" en la UCC; Codirector del proyecto de investigación sobre NTICS (*Nuevas TICs*) en la UCC; Integrante del proyecto de investigación "Laboratorios Virtuales" del IUA. Ha sido cofundador de eGlu BI <www.eglubi.com.ar> y es coordinador de la red social Red Open BI. Se especializa en tecnologías JAVA SE/EE, administración y diseño de Bases de Datos, y OSBI (*Open Source Business Intelligence*).

ner una conceptualización de *BI Usability*. *BI Usability* se refiere al diseño de softwares dedicados al BI que cuenten con interfaz amigable, intuitiva, fácil de utilizar (y fácil de aprender a utilizarla); interfaz que permita la creación de nuevos contenidos (análisis interactivos, *reporting*, *dashboards* o *tablets* de mando), así como su correspondiente navegación, haciendo énfasis en la presentación de dichos contenidos, todo de manera visual e interactiva, para que el usuario se sienta cómodo con su herramienta y le saque el mayor provecho a sus datos.

## 2. Evolución histórica de la usabilidad en BI

A continuación, se enumerarán los hitos principales que se fueron sucediendo y que son antecedentes de la forma que hoy han tomado los sistemas BI en cuanto a la usabilidad.

En la **figura 1** puede apreciarse el detalle de este recorrido histórico.

A continuación se mostrará el impacto de la usabilidad en cada una de estas etapas.

### 2.1. Años 60

**Sistemas de información de BI:** En la década de los 60, los sistemas estaban basados en archivos y la dependencia con respecto del hardware era casi total. Estaban principalmente orientados al almacenamiento y al tratamiento de datos, pero los mismos sistemas de almacenamiento secuenciales (cintas) impedían, en gran medida, la posibilidad de manejar información<sup>3</sup>. El surgimiento del acceso directo, junto con la creación de los primeros discos rígidos, marcó un hito, a partir del cual el software y el hardware ayudaban a procesar los datos para obtener información.

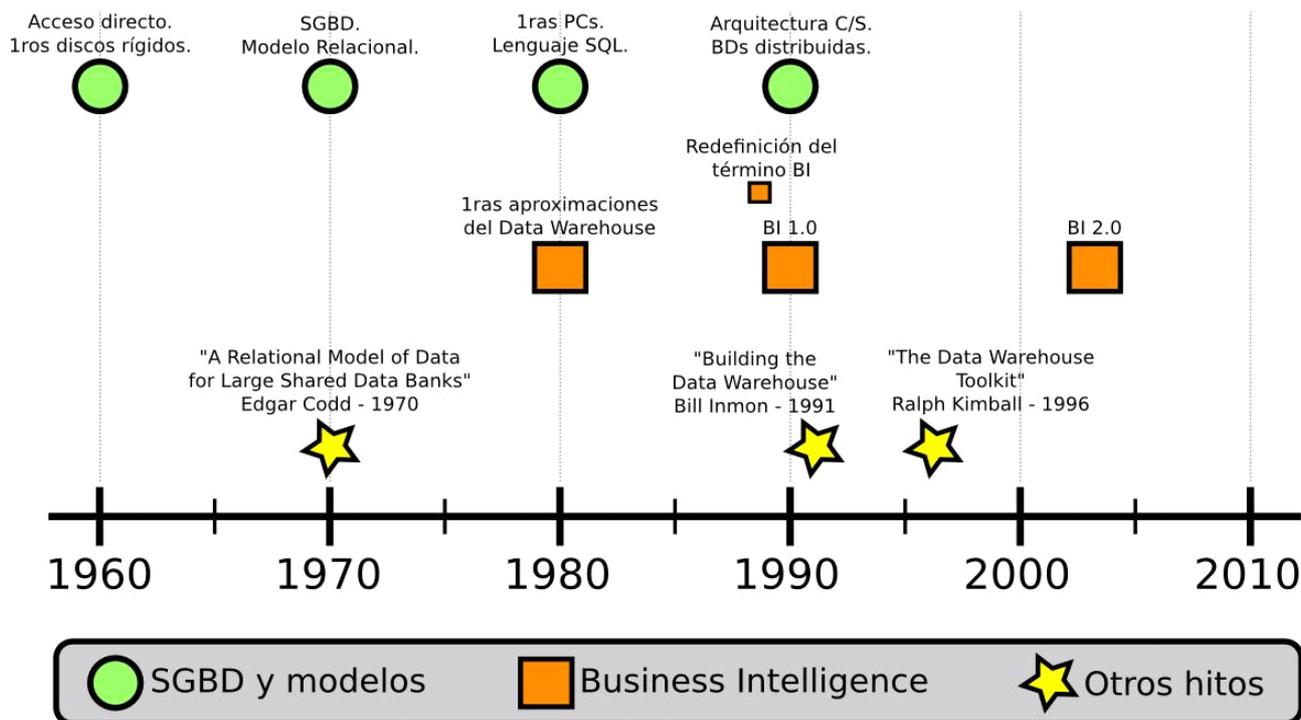


Figura 1. Principales hitos en *Business Intelligence* en comparación con otros hitos históricos.

**Usabilidad de BI en esa época:** Por esos años, la interacción de los sistemas de información con los usuarios era bastante precaria. Consistía en consolas que mostraban una serie de opciones en modo texto que luego el usuario debía seleccionar, en general se presentaban tantas pantallas como opciones disponibles, y al finalizar la selección se obtenían resúmenes de información impresos y/o listados detallados específicos. Por lo definido anteriormente, no hay lugar a dudas que en esta época no se puede hablar de BI per se.

**2.2. Años 70**

**Sistemas de información de BI:** En los 70 la tendencia fue marcada por el surgimiento de los sistemas de gestión de bases de datos (SGBD) y el modelo relacional que fue presentado en 1969 por Edgar Codd (y publicado formalmente en 1970). En esta década es posible visualizar un salto en la evolución de las bases de datos, ya que hasta entonces las mismas se basaban en su mayoría en modelos de red, jerárquicos o simplemente archivos estructurados, cuya característica preponderante era la inflexibilidad y las relaciones físicas entre las entidades.

**Usabilidad de BI en esa época:** Si bien las bases de datos tuvieron un gran impulso de la mano del modelo relacional, recién a finales de esta década se construyeron las primeras versiones de sistemas que les daban soporte. Asimismo, se produjeron mejoras sustanciales en las respuestas de requerimientos de datos e información. La interacción

con el usuario mejoró notablemente y se contaba con interfaces de texto interactivas, esto permitió mejorar la presentación de información por pantalla debido a la posibilidad de realizar *scroll*. A pesar de todo, los informes continuaban siendo estáticos y altamente orientados a la información transaccional.

**2.3. Años 80**

**Sistemas de información de BI:** En la década de los 80, con la aparición en escena de las computadoras personales (PC) se populariza la utilización de los SGBDs y en 1986 se estandariza el lenguaje SQL. Además, aparecen las primeras aproximaciones que traerán como consecuencia la posterior definición del concepto de "Data Warehouse" (DW), la que fue construida en 1992 por Bill Inmon y Ralph Kimball. En 1989 Howard Dresner redefine el término *Business Intelligence*, ya que éste fue utilizado por primera vez en 1958 por Hans P. Luhn.

**Usabilidad de BI en esa época:** Inicialmente, los proveedores de los primeros DW, solo hacían énfasis en el hardware y en la capacidad de sus SGBD, y delegaban la creación de las GUI (*Graphical User Interfaces*) en los desarrolladores/programadores de cada empresa. En aquellos años, quienes se encargaban de diseñar e implementar los DW, se topaban con muchos inconvenientes y dificultades, ya que estas personas estaban acostumbradas a trabajar con sistemas transaccionales/operacionales (OLTP), modelado relacional y, fundamentalmente, a encarar

proyectos de esa índole. Ese apego a los sistemas tradicionales hizo fracasar un alto porcentaje (algunos hablan de un 80%) de los proyectos de la época, debido a que no se comprendía que el desarrollo e implementación de un DW no puede compararse con el de un OLTP, ni mucho menos es viable intentar "adaptar" metodologías y modelos, puesto que para este nuevo concepto deben emplearse herramientas construidas específicamente.

En cuanto a la interactividad, la mejora fue muy notable, los lenguajes de programación permitían crear interfaces gráficas y de texto más amenas y orientadas al usuario. Los informes eran más personalizables y parametrizables y los primeros gráficos de información (gráfico de torta, barras, etc.) vieron la luz. Las hojas de cálculo requieren una mención especial, ya que cambiaron radicalmente la interacción entre el usuario final y la información, otorgándole la posibilidad de mantener e interactuar con sus propios datos. Pero las facilidades que brindaron las hojas de cálculo produjeron como resultado cúmulos de datos redundantes y no vinculados entre sí, debido a que no están pensadas para manejar bases de datos. Posteriormente, estos cúmulos se han ido arrastrando y requieren un gran esfuerzo para procesarlos, ordenarlos y convertirlos en un conjunto que pueda ser explotado de manera segura.

**2.4. Años 90**

**Sistemas de información de BI:** La década de los 90 encuentra a las organizaciones/empresas repletas de PCs, SGBDs perso-

nales, hojas de cálculo, etc. que conforman un conjunto de datos heterogéneos de información descentralizada y no conectada. La arquitectura conocida como cliente servidor (C/S) posibilitó el surgimiento de un nuevo paradigma en cuanto al funcionamiento y comunicación de las aplicaciones. Los SGBDs fueron una de las categorías que más aprovechó esta arquitectura, dando origen a las bases de datos distribuidas, potenciando la intercomunicación en las organizaciones/empresas y haciendo las bases de datos más consistentes y provechosas. Sin embargo, existían una serie de formatos heredados (hojas de cálculo, archivos planos, etc.) para los cuales el aporte de la arquitectura C/S no fue significativo, aunque sí lo fue la idea de estandarizar los procesos de integración de datos<sup>4</sup>.

**Usabilidad de BI en esa época:** Las diversas publicaciones de Bill Inmon y Ralph Kimball, donde explicitan cómo construir y diseñar un DW, además de definir un marco conceptual acerca del tema, ayudaron a clarificar conceptos y, sobre todo, definen un punto de referencia a partir del cual se construirían los DW y las aplicaciones relacionadas al BI. Es en este punto en el que surgen las primeras aplicaciones de software orientadas a DW, tales como: IBM OLAP Server, Cognos, Business Object, SAS, Microstrategy, Oracle, etc. Estas herramientas son las denominadas aplicaciones "BI 1.0" y sus rasgos más importantes o destacables se formulan de la siguiente manera:

- Limitadas en cuanto a analizar grandes volúmenes de datos en un tiempo aceptable ya que las estructuras de almacenamiento físicas no estaban optimizadas para tal fin. Tampoco existían herramientas para mejorar el rendimiento en DW como: *clusters* multidimensionales, tablas agregadas automantenidas, *buffers* con estructuras multidimensionales, etc.
- Acotadas en cuanto a las fuentes de datos posibles.
- Sin consenso general sobre diseño de GUIs para la administración y navegación.

Se puede resumir diciendo que en general la flexibilidad no era una virtud en estas herramientas, aunque cumplieran con las tareas básicas inherentes al DW, y más importante aún, estaban orientadas al DW.

En los primeros años del siglo XXI, hacia el año 2003 aproximadamente surge el "BI 2.0" con el desarrollo de softwares dedicados al BI, que comienzan a incorporar nuevas funcionalidades, características y tecnologías, tales como: interactividad, exploradores web, JS, Ajax, JSON, flexibilidad, interfaces gráficas de usuario (GUI) intuitivas orientadas al usuario final, *Web Services*, etc. A estos softwares dedicados se los conoce como *suites BI*.

### 3. Hechos y sucesos

A continuación se describirán los cambios, hechos y sucesos que fueron dando forma al *BI Usability*.

Se puede decir que con el paso de los años, las aplicaciones BI 1.0 vieron saldadas cuestiones importantes como almacenamiento masivo, velocidad de respuesta, modularidad, etc. Esto fue posible, en gran parte debido a los avances de hardware tales como paralelismo, multiprocesamiento, etc., y a arquitecturas e implementaciones más robustas de software como OLEDB, JDBC, *middlewares*, *frameworks*, etc. En este contexto y con el paso del tiempo, el desarrollo de las aplicaciones BI fue ganando experiencia y madurando en consecuencia.

Como ha ocurrido en muchos casos a lo largo de la historia de la informática, una vez saldadas algunas cuestiones que limitaban el crecimiento, se priorizaron otras cuestiones soslayadas o dejadas de lado hasta entonces. Una de ellas es la usabilidad. En este punto, comenzó a pensarse en la importancia de que las aplicaciones BI fuesen más atractivas, intuitivas y sencillas de utilizar para que los usuarios se sintiesen cómodos y pudiesen sacar el mayor provecho de sus datos.

Hasta el momento, la gran mayoría de las aplicaciones BI eran de tipo de escritorio, ya que los usuarios estaban acostumbrados a la velocidad de respuesta y componentes de interfaz de usuario (también llamados *widgets*) de este tipo de aplicaciones; aunque algunas presentaban una interfaz web bastante limitada. Las limitaciones de las interfaces web, estaban caracterizadas por el período en que fueron desarrolladas, "antes de la web 2.0", en el cual las páginas se cargaban de forma total por cada requerimiento, el ancho de banda era consumido por los requerimientos básicos y los *widgets* eran muy básicos y no podían competir con las versiones de escritorio. En definitiva, no eran agradables ni familiares al usuario.

### 4. La usabilidad del BI en la actualidad y tendencias futuras

En los últimos años, la aparición Ajax<sup>5</sup> y con ella la maduración y/o creación de tecnologías que permitían representar y transportar datos de manera eficiente y estándar, facilitar la creación de GUIs llamativas y potentes, y la interacción entre los datos y las GUIs (JSON, Web Services, *frameworks* JavaScript, flash, CSS, etc.) cambió el paradigma de desarrollo web, pasando de clientes livianos a clientes pesados y con grandes capacidades de procesamiento, interacción y visualización. Las aplicaciones desarrolladas con esta técnica se ejecutan del lado del cliente (cliente pesado o *fat client*), el cual solo requiere del servidor lo que necesita puntualmente (y no toda la página como antes), pudiendo hacerlo de

forma asíncrona, permitiendo de esta manera que el usuario nunca pierda la interactividad con la aplicación.

El advenimiento de Ajax y las tecnologías mencionadas en el párrafo anterior, marcan lo que hoy se conoce como BI 2.0, e inclina la balanza hacia el lado del desarrollo web<sup>6</sup>. Las aplicaciones BI 2.0 se enfocan en el diseño y presentación de las consultas, reportes, análisis OLAP, etc, a través de gráficos interactivos, objetos Flash y JavaScript, *dashboards* personalizados y altamente parametrizables, etc. Todo esto haciendo énfasis en la interfaz gráfica y en la interactividad con el usuario.

Actualmente el desarrollo y crecimiento de aplicaciones BI y tecnologías asociadas no ha mermado. Por el contrario, continuamente recibe nuevos impulsos. Algunos de los desarrollos que eventualmente podrían marcar tendencias de cambio o, por qué no, establecer nuevos hitos, son:

- Bases de Datos no SQL.
- *Cloud Computing*.
- *Sql Streaming*.
- *In-Memory OLAP*.
- Tecnologías móviles.
- Tecnologías basadas en GPS.
- Tecnologías *Touch*.
- Reconocimiento de voz.
- Prototipado ágil.

Si se habla de tendencias y desarrollo, omitir una referencia a OSBI (*Open Source Business Intelligence*) sería definitivamente un desacierto. En los últimos años, el crecimiento y aporte de la comunidad de software libre y *open source* en este sentido ha sido cuantiosa y significativa, logrando que las aplicaciones OSBI puedan competir con las aplicaciones privativas y, en algunos casos, a marcar tendencias e impulsar ideas muy innovadoras. Como es costumbre, y parte de la cultura del software libre y *open source*, existe un gran número de foros, blogs, wikis, redes sociales (por ejemplo Red Open BI<sup>7</sup>), etc., que son de gran calidad. Millones de usuarios alrededor del mundo, utilizan estos medios como forma primaria para realizar aportes y compartir conocimientos, posibilitando que los diversos proyectos crezcan continuamente y se hagan robustos, haciendo de estos softwares una opción viable y segura.

El OSBI ha brindado la posibilidad a pequeñas y medianas empresas de implementar soluciones de BI, antes denegadas por el alto costo de las herramientas, aumentando notablemente la demanda de este tipo de soluciones. La demanda no solo se remite a la cantidad, sino al aumento de requerimientos funcionales y no funcionales como: mejores interfaces, mayor interactividad, respuestas en menor tiempo, entrega de información en diversas formas y por diversos canales, etc.

## 5. Conclusiones y trabajo futuro

A modo de conclusión, puede remarcar que la evolución hacia el BI 2.0, y más particularmente el énfasis hacia la usabilidad, ha ido desarrollándose al compás de los procesos de avance y maduración de las herramientas y aplicaciones, el hardware utilizado, los requerimientos de los usuarios en cuanto a lo que se espera de un aplicación BI actual (interactividad, familiaridad, GUI intuitivas, *wizards*, etc) y el mundo OSBI.

Podemos asumir entonces que el rumbo se mantendrá hacia direcciones similares en el cambio de tecnologías. Consideramos que no es descabellado pensar un escenario no muy lejano en el cual un encargado de stock de supermercado se acerca a una góndola y le ordena a su móvil mediante un comando de voz "estado stock", mientras apunta la cámara de tal manera que abarque una serie de productos distintos; inmediatamente en la pantalla y sobre la imagen que se está capturando aparecerían agrupados por colores semitransparentes, los productos del mismo tipo, donde la intensidad del color estaría indicando el estado del stock. El encargado podría presionar sobre una de estas áreas, las que poseen un icono de alerta que le indica la necesidad de reposición inminente; es posible enriquecer el ejemplo planteando que, además de estas facilidades, aparecería una pantalla en el móvil de este encargado mostrando un gráfico histórico de stock donde se podría analizar rápidamente la tendencia y estacionalidad (entre otras cosas). Esta información permitiría a nuestro encargado tomar la

decisión de reposición de stock a través de su móvil, mediante un comando de voz "realizar pedido"; si esto sucediera aparecería en la pantalla un mensaje preconfigurado para el proveedor más conveniente (por cercanía, por precio, por eficiencia, etc.) con los datos del pedido y sin más que la orden verbal de "enviar". El trabajo de este proveedor imaginario también se realizaría desde y con su móvil.

El ejemplo anterior no es un relato de ciencia ficción ni mucho menos delirio. Estamos alineados en los rieles de este desarrollo, falta maduración pero las tecnologías ya existen y son algunas de las mencionadas anteriormente.

## Agradecimientos

A Andrea Martino por su colaboración en cuanto al estilo de redacción, ya que de otra manera el artículo hubiese sido demasiado "only-for-geeks".

## Bibliografía

- W. H. Inmon.** *Building the Data Warehouse*. Wiley, 3ª edición, 2002.
- Ralph Kimball.** <<http://www.kimballgroup.com/>>.
- C.J. Date.** *Introducción a los Sistemas de Bases de Datos*. Prentice Hall. 6ª edición, Addison-Wesley.
- G.W. Hansen, J.V. Hansen.** *Diseño y Administración de Base de Datos*. Prentice-Hall, 1997.
- R. Elmasri, S. Navathe.** *Sistemas de Bases de Datos*. Addison-Wesley Iberoamericana, 2ª edición.
- Rafael Camps Paré.** *Introducción a las bases de datos*.
- Dolors Costal Costa.** *Introducción al diseño de bases de datos*.

**Dolors Costal Costa.** *El modelo relacional y el álgebra relacional*.

**Andrew Tanenbaum, Maarten Van Steen.** *Distributed systems - Principles and Paradigms*.

## Notas

<sup>1</sup> Las definiciones siguientes fueron tomadas de la Wikipedia: <<http://es.wikipedia.org/wiki/Usabilidad>>.

<sup>2</sup> **Ing. R. Dario Bernabeu.** *DATAWAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología para la Construcción de un Data Warehouse (Versión 2.0 - 19 de julio de 2010)* < <http://www.dataprix.com/es/data-warehousing-hefesto>>.

<sup>3</sup> **Información** es un conjunto de datos organizados, ordenados y procesados que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema que lo recibe. Fuente: Wikipedia.

<sup>4</sup> **Integración de datos:** conjunto de técnicas y subprocesos que se encargan de extraer datos desde diferentes orígenes, manipularlos, integrarlos, transformarlos, volcarlos en otra fuente de datos, etc.

<sup>5</sup> **Ajax**, acrónimo de *Asynchronous JavaScript And XML* (JavaScript asíncrono y XML), es una técnica de desarrollo web para crear aplicaciones interactivas o RIA (*Rich Internet Applications*). Fuente: Wikipedia.

<sup>6</sup> El término **Web 2.0** (2004–actualidad) está comúnmente asociado con un fenómeno social, basado en la interacción que se logra a partir de diferentes aplicaciones en la web, que facilitan el compartir información, la interoperabilidad, el diseño centrado en el usuario y la colaboración en la World Wide Web. Ejemplos de la Web 2.0 son las comunidades web, servicios web, aplicaciones web, blogs, wikis, plataformas educativas, entornos para compartir recursos, redes sociales. Fuente: Wikipedia.

<sup>7</sup> **Red Open BI** es la primera red social en español dedicada al OSBI, <<http://www.redopenbi.com/>>.

**INVITA A UN AMIGO A QUE DISFRUTE, SIN COSTE ALGUNO Y DURANTE ESTE AÑO, DE LAS VENTAJAS DE SER SOCIO DE ATI**

La esencia actual de ATI sigue siendo la misma que la originó: Crear una red de profesionales que permita una mejora constante de la profesión informática, individual y colectivamente.

Solo necesitas introducir en el siguiente formulario tus datos (nombre, apellidos, número de socio y correo-e) y los datos de contacto de la persona a quien deseas invitar a ATI (nombre, apellidos y correo-e) y le remitiremos tu invitación.

No te llevará más de dos minutos y contribuirás a enriquecer vínculos asociativos, además de ayudar a fortalecer y hacer crecer esta red de profesionales.

>> Acceso al formulario: <http://bit.ly/socioinvitado>

\* La persona beneficiada gozará durante el 2011 de: descuentos en formación, ofertas especiales, invitaciones a presentaciones y eventos, consulta de la revista Novática vía intranet, participación en foros, listas de distribución, grupos de interés, acceso preferente a la bolsa de trabajo, cuenta de correo electrónico...

\*\* Esta promoción está limitada a un invitado por socio. No se podrá invitar a más de uno.

Jorge Fernández González<sup>1</sup>,  
Enric Mayol Sarroca<sup>2</sup>

<sup>1</sup>Director de Consultoría en Business Intelligence de Abast Solutions; <sup>1,2</sup> Departamento de Ingeniería de Servicios y Sistemas de Información (ESSI) de la Universidad Politécnica de Catalunya (UPC)

<jfernand@abast.es>,  
<mayol@essi.upc.edu>

## 1. Introducción

Hoy en día, estamos en el auge del movimiento de la llamada inteligencia de negocio o Business Intelligence (BI). Casi todas las organizaciones se esfuerzan por crear y mejorar sus procesos y sistemas de toma de decisión. Una gran cantidad de nuevos proyectos de BI aparecen constantemente, pero la experiencia global en los últimos años no es tan buena. Algo por lo general va mal en la ejecución de proyectos de BI, ya que la mayoría de proyectos de BI (85%) no pudo lograr sus objetivos [1].

La disciplina de Business Intelligence es todavía un área muy joven [2]. Hemos encontrado trece enfoques metodológicos diferentes para administrar un proyecto de BI [3][4][5][6][7][8][9][10][11][12][13][14][15], y la mayoría de ellos se han definido en los últimos 10 años. Pero, ¿cuáles son las razones de esta tasa tan alta de fracasos frente a la gran cantidad de alternativas metodológicas?

De hecho, podríamos decir que la gran diversidad y heterogeneidad de enfoques metodológicos para proyectos BI muestra la inmadurez que todavía existe en este ámbito. Por lo tanto, elegir una metodología de BI no es una tarea fácil. J. Thomann y D.L. Wells [16] afirman que cada proyecto de BI y cada organización debe escoger la metodología específica que mejor se adapte a la organización y características del proyecto para tener más posibilidades de éxito.

## 2. Enfoques metodológicos de Business Intelligence

En este apartado vamos a revisar muy brevemente los principales enfoques metodológicos existentes para proyectos BI, así como sus principales limitaciones.

El ciclo de vida de un proyecto de BI [17][18] implica múltiples fases, cíclicas y muchas veces ejecutándose en paralelo, de ahí la complejidad que puede llegar a tener. Se han identificado más de 900 tareas a ejecutar [17][19], cosa que hace que no sea del todo sencillo definir una metodología única. Son muchos los enfoques metodológicos que la literatura científica ha tratado. Según R. Cicchetti et al. [20] son éstos:

# Factores críticos de éxito de un proyecto de Business Intelligence

**Resumen:** El objetivo de encontrar la mejor metodología de Business Intelligence (BI) es difícil o imposible, pero creemos que sí es posible identificar las características que dicha metodología debe cumplir. En este artículo hemos analizado las características principales de los proyectos de Business Intelligence y distintos enfoques metodológicos usualmente utilizados. Este estudio revela que algunas metodologías tienen ciertas debilidades al no estar expresamente definidas para proyectos de BI, y por lo tanto no se ajustan a las características específicas del proyecto o a las necesidades de los usuarios. Para ello, hemos identificado cuáles son los principales factores críticos de éxito de estos proyectos que pueden determinar cómo tendría que ser la "Metodología de BI". Hemos realizado un análisis de las principales publicaciones en este dominio anotando los factores que los distintos autores y enfoques de BI identifican como determinantes en el éxito y/o fracaso de un proyecto de BI. Hemos propuesto una agrupación de los mismos y una uniformización del vocabulario utilizado, proponiendo finalmente un subconjunto de los factores críticos de éxito de proyectos de BI más significativos.

**Palabras clave:** Business Intelligence, Data Warehouse, factores de éxito, metodologías, proyectos.

### Autores

**Jorge Fernández González** desarrolla su actividad profesional en dos ámbitos de actuación, el principal de ellos es como profesional de Sistemas de Información. En el que acumula más de 14 años de experiencia y actualmente ejerce el cargo de Director de Consultoría en Business Intelligence de Abast Solutions. El segundo de sus ámbitos de actuación es el de docente universitario. Actualmente es Profesor Asociado al departamento de ESSI de la UPC y responsable de la asignatura "Sistemas de Información para las Organizaciones" de la Facultad de Informática de Barcelona. También ha sido colaborador docente en la Universitat Oberta de Catalunya (UOC), director académico y profesor en masters y postgrados de la Fundación Politécnica y realiza sesiones como invitado en escuelas de negocio como ESADE, La Salle y EAE. En el ámbito de las publicaciones, ha colaborado con la revista Data TI en múltiples ocasiones, fue redactor de la revista Gestión del Rendimiento, es miembro del grupo de expertos en español del portal BeyeNetwork, portal que reúne a los especialistas en BI a nivel mundial, y colabora como experto en todas las ediciones del *BARC Business Intelligence Guide*.

**Enric Mayol Sarroca** es Doctor Ingeniero en Informática desde abril del año 2000 por el programa de Software de la Universidad Politécnica de Cataluña y miembro del Departamento de Ingeniería de Servicios y de Sistemas de Información. Actualmente ocupa la plaza de profesor contratado doctor en dicho departamento y es profesor de la especialidad de Sistemas de Información del Grado en Informática de la Facultad de Informática de Barcelona, en la misma universidad. Miembro del grupo de investigación GESSI (Grupo de investigación en Ingeniería del Software para los Sistemas de Información) y del grupo SUSHITOS (Services for Ubiquitous Social and Humanistic Information Technologies and Open Source Research Group), participa en distintos proyectos de investigación. Sus intereses de investigación se centran en los Sistemas de Información, Business Intelligence, Gestión de Contenidos Digitales, e-Learning y la Ingeniería de Servicios. Es vocal de la Junta Directiva de la *Societat Catalana de Genealogia, Heràldica, Sigil·lografia, Vexil·lologia i Nobiliaria* (SCGHSVN) y el dominio de aplicación de su investigación se centra en la temática de Sistemas de Información y Genealogía.

### Plan-Driven approach o Requirement-Driven approach

Esta metodología más tradicional no parece adecuada para este tipo de proyectos [7] ya que este enfoque no satisfará las demandas futuras de los usuarios, y los usuarios difícilmente son capaces de definir y explicar como toman sus decisiones.

### Demand-Driven o User-Driven o Prototype-Driven Approach

Metodologías orientadas hacia la confección de prototipos [21] para la obtención de los requisitos que sean lo suficientemente precisos [7]. Así pues, se busca mostrar al usuario un prototipo funcional [22] para intentar captarlos lo mejor posible [23].

El punto débil de este enfoque está en asumir que todos los usuarios conocen la estrategia empresarial y se comportan de forma coherente con ella, lo cual no siempre es así. Pero

de serlo, si realmente son ellos los que van a tomar las decisiones, son ellos también los que deberían dirigir el proceso de creación del sistema de BI.

La idea se fundamenta en crear un primer prototipo basado en los objetivos empresariales y a partir de ahí los usuarios definen las necesidades de información, las preguntas que le van a hacer al sistema BI, y el mantenimiento y evolución futura [24] del mismo.

B. Afolabi y O. Thiery [5] nos hablan de la importancia de usuario para definirlo basándose en sus propias fases cognitivas (observación, abstracción elemental, razonamiento y simbolización y creatividad). Se puede adaptar nuestro sistema de BI al tipo de consultas que nos hará el usuario final (*Query Adaptation*) y al tipo de respuestas que espera recibir (*Response Adaptation*) [25].

#### **Data-Driven Approach**

Este enfoque [7] se centra en los datos: en cómo están estructurados, en quién los usa, en la forma en que los usan. Se fija en los datos con mayor tasa de acceso, aquellos que se consultan con mayor frecuencia, como se relacionan entre ellos, qué consultas suelen venir asociadas. Son los datos los que dirigen el proceso. Este enfoque se basa en la premisa de que los datos nunca mienten, mientras que de los usuarios es difícil de asegurar. El problema es que en este enfoque, a priori se deja de lado a los usuarios, los objetivos de la organización y los futuros requisitos del sistema.

#### **Value-Chain Data Approach**

Basado en la cadena de valor del *Business Intelligence* [9], es una evolución del enfoque *Data-Driven* focalizada en los datos que generaran mayor valor para el negocio, pero no resuelve las limitaciones de su predecesor.

#### **Process-Driven Approach**

Este enfoque se basa en el análisis de los procesos de negocio [14], la información que generan y la información que consumen. El proceso es la clave y se estructura la información según sea el usuario de proceso. Un aspecto que se puede perder de vista en este enfoque, demasiado centrado en el proceso, es la perspectiva global de la organización y las relaciones entre procesos, lo cual puede llevar a tener una visión incompleta o errónea de la organización.

#### **Event-Driven Approach**

Este enfoque propone dividir los procesos de negocio bajo tres puntos de vista: Datos, Función y Organización, cada una de las cuales se conecta entre sí a través de eventos. La gran ventaja de este enfoque es el análisis funcional de la organización.

V. Stefanov y B. List [6] proponen este mis-

mo modelo extendido con objetos BI y conectores de información BI, como una manera de rellenar el gap entre el negocio y los sistemas de BI. Este enfoque es muy complejo de llevar a la práctica y requiere una gran experiencia y modelos organizacionales muy maduros.

#### **Object-Process Driven Approach**

Es una de las variantes metodológicas a medio camino entre el *Event-Driven* y el *Process Driven* [13]. En este enfoque, tanto los objetos como los procesos tienen la misma importancia desde el punto de vista decisional y por tanto se deben tratar de la misma manera.

#### **Joint Approach**

Enfoque metodológico centrado en el reconocimiento de las arquitecturas funcionales cruzadas de las empresas. Los procesos no son de un solo departamento, sino que existen muchos puntos de contacto y muchas juntas [11], por lo tanto es donde se tiene que centrar el esfuerzo. La idea es que la organización es una matriz de procesos con diferentes necesidades de información, pero allí donde se juntan es donde debemos hacer el mayor esfuerzo. La dificultad del enfoque puede radicar en la dificultad en definir los procesos de gestión y control de la información en estos puntos de contacto.

#### **Goal-Driven Approach**

Este enfoque [7] se centra en el objetivo de los procesos estratégicos de la organización y se basa en el análisis de la interacción que tanto clientes como usuarios hacen para conseguir dicho objetivo. A partir de ahí establece necesidades de información e interrelaciones entre ellas que darán lugar a la estructura del sistema de *Business Intelligence*. El problema puede aparecer cuando no existe un conocimiento o alineamiento preciso entre los procesos estratégicos y los tácticos o operacionales.

#### **Triple-Driven Approach**

Vista la inmadurez de las metodologías de *Business Intelligence*, en Y. Guo et al. [12] apuestan por una combinación de las mejores ideas de cada una de las metodologías *Goal*, *Data User Driven*, creando la *Triple-Driven*, pues se considera que estos tres enfoques son perfectamente compatibles.

#### **Model Driven Approach**

Otra de las metodologías que se han usado en BI es la *Model Driven* [3][4]. Con ella, se pretende tender un puente entre el negocio y el departamento de Informática, intentando proporcionar la base para desarrollar soluciones rápidas, que evolucionen fácilmente y flexibles. Debido a su alto nivel de la reutilización de la abstracción y del código, la metodología MDA (*Model-Driven Architecture*) se ha aplicado extensamente. El

MDA permite reducir tiempo de desarrollo de software, la mejora de la calidad y del mantenimiento de la solución. Pero por el contrario, es difícil definir este modelo simplificado de la realidad y aún es difícil de implantar sobre arquitecturas SOA (*Service-Oriented Architecture*) y en organizaciones reales.

#### **Adaptive Business Approach**

La metodología *Adaptive Business Approach* [8] se basa estrictamente en aquellos aspectos realmente relevantes para el negocio y su evolución. Se centra en los problemas que el negocio tiene que resolver para adaptarse a los cambios del mercado y en los datos de que disponemos para ello. El resultado de los sistemas de *Business Intelligence* han de ser o bien la solución al problema o bien la aportación de más conocimiento sobre el problema para seguir analizando y tomando decisiones para hallar dicha solución. El centrarse en solo lo relevante para el cambio, puede dejar de lado o no considerar explícitamente otros aspectos no tan críticos del negocio, pero que determinan o influyen en aspectos más relevantes. Por lo tanto estas dependencias tienen que tenerse en cuenta explícitamente y no obviarse.

#### **Agile Approach**

Si este enfoque puede considerarse novel en el área de ingeniería del software, lo es más en el área de *Business Intelligence*. La primera referencia académica que encontramos referente al uso de las metodologías ágiles en BI es un pequeño artículo divulgativo en el año 2001 de apenas 2 páginas en el que L.T. Moss [10] comenta que sería posible realizar proyectos de *Business Intelligence* con rigor usando las metodologías ágiles. Posteriormente, J. Fernández et al. [26] analizan la correlación entre los principios básicos de las metodologías ágiles y las necesidades de una metodología de BI. Entre tanto, algunos trabajos adicionales aparecen, pero sorprende la reducida cantidad de trabajos académicos del tema respecto al enorme número de artículos de experiencias reales del tema, creciendo exponencialmente en 2011.

### **3. Como debe ser la metodología de BI**

Tanta diversidad de enfoques metodológicos denota la inmadurez que existe en este tipo de soluciones, las interrelaciones no son claras y elegir una metodología no es nada fácil. Los trabajos [16][27] nos muestran que para cada proyecto de BI y para cada organización debe seleccionarse la metodología que más posibilidades de éxito tenga.

L.T. Moss [19] nos propone las características que debería cumplir una metodología para este tipo de sistemas decisionales, a las que hemos incluido las dos características finales:

1) Ha de estar orientada al cambio y no a la

consecución de un producto final.

- 2) La gestión del proyecto debe ser de forma global y transversal a toda empresa.
- 3) Debe poder manejar múltiples subproyectos a la vez y en paralelo.
- 4) Ha de tener en cuenta todas las tareas/procesos de la empresa, sean o no críticos.
- 5) Debe basarse en la gestión de los caminos críticos del *workflow* empresarial.
- 6) Debe estar orientada a las personas y relaciones entre ellas.
- 7) Ha de estar alineada con las necesidades de negocio de la organización.

Estas características se han sintetizado a partir de la experiencia práctica y la realidad, y se presentan con un enfoque más de divulgación empresarial que académico o científico. Por ello, creemos que un enfoque complementario para determinar cuales son las metodologías más adecuadas para la gestión de un proyecto BI es basarnos en un proceso de análisis de factores críticos de éxito de todo proyecto de BI. Por eso es necesario, en primer lugar, analizar cuales son los factores de éxito de los proyectos de BI más citados y referenciados en la bibliografía académica. Actividad que queremos reflejar en los siguientes apartados.

#### 4. Factores críticos de éxito (FCE)

Un estudio realizado por la Universidad de Monash [1] concluye que el 85% de los proyectos de BI han fracasado en la consecución de sus objetivos. Así pues, ¿Qué determina el éxito y el fracaso de esos proyectos? A nuestro entender, una análisis de los factores críticos de éxito de estos proyectos ha de dar respuesta a esta pregunta.

Para introducir estos factores de éxito, los presentamos agrupados por categorías, numerados de 1) a 6), según el aspecto del proyecto BI que contemplan. Hemos dejando una última categoría para todos aquellos factores identificados en una revisión bibliográfica que hemos realizado, siendo todos ellos aspectos que los autores aseguran que influyen positivamente o negativamente al éxito o fracaso de un proyecto BI. Este último conjunto de factores, será el que intentaremos reagrupar y priorizar en el siguiente apartado.

#### 1) Relativos a las herramientas de BI

¿Están fallando las herramientas de BI respecto a lo que se espera de ellas? En el trabajo de B. Azvine et al. [28] se nos dice que las actuales herramientas de BI han estado fallando en tres puntos principales:

- En la visualización de la información de lo que ha pasado.
- En ayudar a comprender la razón o las causas de lo que ha pasado.
- En la predicción de lo que sucederá.

Si realmente depende de la validez de las herramientas de BI, entonces, no es necesario

preocuparse por la metodología a utilizar. La verdad es que esta visión [28] podría ser demasiado alarmista. Pero hay que reconocer que las herramientas de BI padecen aún de un gran déficit en los dos últimos puntos, aunque poco a poco van mejorando su fiabilidad.

#### 2) Relativos a la organización

Una de las características básicas para el éxito del BI es, sin duda, la cultura organizacional y el nivel de madurez de la organización. Pero la creación y gestión de una cultura y un nivel de madurez requiere mucho tiempo. K.R. Quinn [29] ofrece cinco reglas para el éxito y otras 5 causas del fracaso de los proyectos de BI relacionados con la organización:

- + Comprender a los usuarios.
- + Utilizar el paradigma de los clicks.
- + Distinguir claramente entre usuarios productores y usuarios consumidores de información.
- + Establecer una cultura organizativa de medición y evaluación.
- + Conseguir que el proyecto de BI sea una decisión estratégica de toda la compañía.
- Se desestiman las necesidades y deseos de los usuarios.
- Se hace demasiado énfasis en fases menos críticas.
- La información no es auto-explicable, poco énfasis en la semántica.
- No se ha establecido una estrategia de medición.
- El proyecto BI solo se ha aplicado tácticamente, no estratégicamente.

#### 3) Relativos a la gestión del conocimiento

Otro punto a tener en cuenta al evaluar el éxito de las soluciones de BI se encuentra en los déficits o limitaciones del propio Departamento de Sistemas de Información. J. Becker et al. [30] enumeran las deficiencias en la gestión del conocimiento en los departamentos de BI.

La primera limitación que identifican es la dificultad para definir estructuras adecuadas para la localización de la información. El 30% de los documentos con información relevante se almacena y manipula en ordenadores personales o portátiles aislados por lo que el acceso a dicha información está limitando y accesible solamente a un conjunto limitado de usuarios en la organización, y no se facilitan los mecanismos para compartirlas con el resto de personal interesado.

Una segunda limitación es la dificultad del personal respecto al acceso al conocimiento para ejercer sus funciones satisfactoriamente. Esta adquisición del conocimiento requiere tiempo y dinero de la empresa repercutido en las iniciativas de formación, pero estos recursos son de difícil reutilización. Si a ello se suma una alta rotación de empleados, las pérdidas pueden llegar a ser considerables.

#### 4) Relativos a aspectos intangibles

El éxito de la solución del proyecto BI ha de tener presente el componente intangible del proyecto que también debe ser evaluado. A. Counihan et al. [31] ya identificaban la dificultad de evaluar los aspectos intangibles de los sistemas de información estratégicos. Por otra parte, M. Gibson et al. [32] identificaban seis criterios para evaluar los beneficios intangibles de los sistemas de BI:

- Determinar la criticidad de las cuestiones intangibles.
- Separar los requisitos de los usuarios de los aspectos intangibles propios del proyecto.
- Convencer de la importancia de los intangibles a los gerentes de la empresa.
- Clasificar adecuadamente los intangibles para hacer más fácil su evaluación.
- Gestionar el proyecto orientado a resultados rápidos (*quick win*).
- Medir el nivel de cumplimiento de los intangibles.

#### 5) Relativos al personal y al liderazgo

Otro factor a tener en cuenta que puede determinar el éxito o el fracaso del proyecto BI es el perfil, experiencia y formación de las personas involucradas en el proyecto, la implantación y, especialmente en el liderazgo. En este sentido, A. Faulkner y A. MacGillivray [33] identifican 12 aspectos o requisitos que debería satisfacer el líder del proyecto para asegurar un posible éxito del proyecto BI:

- Saber reflexionar antes que actuar sobre los valores de la organización.
- Focalizar los objetivos del proyecto en las necesidades más urgentes de la organización, a fin de que los resultados sean visibles lo más pronto posible.
- Ser un antropólogo amateur, identificando las necesidades del negocio y quienes las están gestionando, para proveerles de herramientas fáciles de usar según sus habilidades, necesidades y preferencias personales.
- Planificar para el éxito. Es decir, comprender como la organización evalúa el éxito y dirigirse hacia él.
- Ser un niño de pequeño, preguntando siempre el porqué de todo.
- Conseguir que el proyecto sea una iniciativa de cambio e innovación para toda la organización.
- Dialogar, dialogar y dialogar.
- Integrar e involucrar a los ejecutivos y directores intermedios como co-líderes del proyecto, para que sientan el proyecto como suyo.
- Ser proactivo, anticipar la resistencia al cambio y convertirse en el paladín de la causa de BI.
- Aprender de los demás.
- Evaluar el coste y el riesgo de la alternativa a no usar herramientas BI, y usarlo como argumento.
- Tener una mente abierta y una visión global de la evolución que puede tener el BI dentro de la organización.

Por otro lado, T. Chenoweth et al. [34] afirman que la interacción entre la tecnología y el contexto social de las empresas claramente determina el éxito o el fracaso de un banco de datos común a la organización. Y al mismo tiempo, reconocen que esta misma interacción puede determinar la extensión y evolución de un sistema de BI. En este sentido, proponen siete intervenciones clave, como una lista ordenada de preguntas básicas a considerar en todo proyecto BI.

- ¿El proyecto tiene el apoyo de la dirección?
- ¿Los futuros usuarios apoyan el proyecto?
- ¿Los usuarios accederán con el sistema a un amplio abanico de datos?
- ¿Los usuarios necesitan herramientas restrictivas?
- ¿Los usuarios entienden la relación existente entre la información que les proporcionará el sistema BI y los procesos de negocio que llevan a cabo?
- ¿Los usuarios perciben al departamento de SI como soporte y ayuda en la realización de sus procesos de negocio?
- ¿Existen usuarios líderes o 'power users'?

La principal conclusión en este apartado es que el usuario es el centro de todo, y que la implicación del usuario en el proyecto BI va a determinar claramente el éxito o el fracaso del mismo

### 6) Otros factores críticos de éxito

En esta sección proporcionamos una lista exhaustiva de FCE obtenida a partir de una revisión de la literatura científica y académica. En los trabajos analizados, se proponen diversos FCE sin focalizarlos en temas específicos de *Business Intelligence* como en las agrupaciones anteriores. En este caso, son más genéricos pero igualmente relevantes y críticos.

**M.D. Solomon** [35] nos propone una guía de aspectos a considerar cuando se lleva a cabo la implementación de una solución de BI o de *Data Warehouse*<sup>1</sup>.

- El usuario participa en la definición del nivel de servicio y los requisitos de información.
- Identificar los sistemas de provisión de datos.
- Definir plan de calidad.
- Elegir un modelo de diseño adecuado.
- Escoger la herramienta ETL (*Extract, Transform and Load*) a usar.
- Realizar cargas de datos incrementales preferentemente.
- Escoger cuidadosamente la plataforma de desarrollo BI y el SGBD adecuado.
- Realizar procesos de conciliación de datos.
- Revisar y modificar periódicamente la

planificación.

- Proveer soporte al usuario.

**L.T. Moss** [19] resume los 10 errores más frecuentes en la gestión de proyectos de BI y de *Data Warehouse*.

- No usar ninguna metodología.
- No disponer del equipo adecuado.
- Los usuarios y decisores no participan en el proyecto.
- Descomposición del proyecto en etapas inadecuadas.
- Inexistencia de una planificación del proyecto.
- Inexistencia de un plan de calidad.
- Pruebas de calidad incompletas o inadecuadas.
- No prever el volumen de datos a monitorizar o depurar.
- Ignorar metadatos y semántica de datos.
- Depender en exceso (ser esclavo) de la herramienta de gestión del proyecto.

**D. Briggs et al.** [36][37] proponen los siguientes FCE para sistemas decisionales:

- Patrocinio del proyecto.
- Gestión de las expectativas del usuario.
- Uso de prototipos.
- Búsqueda del resultado rápido (*quick win*)
- Escoger un problema de la organización medible.
- Modelización y diseño del *data warehouse*.
- Selección del caso de negocio adecuado.
- Alineación con la estrategia organizativa.
- Selección cuidadosa de las herramientas.
- Usuarios finales involucrados.
- Gestión del cambio organizativo.
- Consideración de la cultura de la organización.
- Centrarse en la gestión de los datos.
- Nivel de escalabilidad y flexibilidad del proyecto y la solución.
- Transmitir el conocimiento en proyectos subcontratados.
- Uso de estándares.
- Aprovechamiento de la experiencia de los miembros del equipo.
- Soporte al usuario final.

**B.H. Wixom y H.J. Watson** [38][39] identifican los siguientes factores como los más relevantes para un proyecto BI:

- Apoyo a la gestión de la organización.
- Existencia de un líder del proyecto.
- Uso adecuado de los recursos.
- Participación del usuario final.
- Equipo con competencias adecuadas.
- Disponer de fuentes de datos adecuadas.
- Considerar la información y su análisis como parte de la cultura de la organización.
- Alineamiento con la estrategia de la organización.
- Gestión y control del BI eficaz.

**D. Sammon y P. Finnegan** [40] proponen los 10 mandamientos del proyecto BI:

- Iniciativa ligada a las necesidades del negocio.
- Existencia de patrocinio de la dirección.
- Gestión de las expectativas del usuario.
- Proyecto transversal a la organización.
- Control de calidad.
- Flexibilidad del modelo de datos.
- Gestión orientada a los datos.
- Proceso automático de extracción de datos.
- Conocimiento.
- Experiencia.

**R. Weir et al.** [41] proponen este conjunto de mejores prácticas de BI:

- Realizar cambios incrementales.
- Construcción del sistema adaptable.
- Gestionar expectativas del usuario.
- Equipo mixto entre técnicos y usuarios finales.
- Contacto directo con la organización y el negocio.
- No perseguir la perfección.

**R.S. Abdullaev y I.S. Ko** [42] analizan las lecciones aprendidas de diversas experiencias en la construcción de BI:

- La centralización de datos en un *data warehouse* y su agregación en varios *data marts*<sup>2</sup> permiten un acceso rápido y de confianza a la información solicitada.
- La definición de listados estándares para todos los usuarios favorece el intercambio de información entre departamentos de una manera más clara y consistente.
- Algunos modelos de informes predefinidos se han de implementar con el fin de proporcionar a los decisores la funcionalidad para añadir o eliminar elementos necesarios y crear informes específicos.
- Es necesario un equipo responsable de alinear las especificaciones de informes estándar con las necesidades locales y que facilite la ejecución del proyecto de BI.
- Debe existir un fuerte compromiso de la dirección para resolver cualquier conflicto y gestionar los cambios que ocurran durante el desarrollo del proyecto.
- La integración de técnicas "Six Sigma" en la infraestructura TI de la organización contribuye a un sistema BI robusto.
- La infraestructura TI ha de centrarse en una sola plataforma proporcionada por proveedores bien conocidos.

**I.S. Ko y R.S. Abdullaev** [43] identifican los aspectos especialmente críticos durante el desarrollo de proyectos de BI:

- Requisitos del mercado y del cliente en lugar de requisitos internos.
- Disponer de personal representante de cada departamento dedicados al proyecto.

- Equipo formado por personal competente en el proyecto.
- Adopción de una metodología de desarrollo de proyectos BI.
- Realizar y seguir una planificación del proyecto.
- Estandarización de los datos.
- Control de calidad de los datos.
- Existencia de metadatos.
- Usar solo las herramientas necesarias.

**W. Yeoh et al.** [44] realizan una recompilación de FCE para proyectos de BI:

- Soporte al proyecto por parte de la alta dirección.
- Disponer de los recursos adecuados.
- Apoyo comprometido por parte de la organización.
- Participación formal del usuario a lo largo de todo el proyecto.
- Soporte, formación y entrenamiento.
- Caso de negocio establecido y consensuado.
- Visión estratégica de BI integrada con las iniciativas de la compañía.
- Ámbito del proyecto claramente definido.
- Adopción de un enfoque de resultados incrementales.
- Proyecto orientado a resultados rápidos (quick wins).
- Equipo poseedor de la perfecta combinación de capacidades.
- Participación de consultoría externa en fases iniciales del proyecto.
- Experiencia en el dominio del negocio.
- Equipo multifuncional.
- Sistemas proveedores de datos estables.
- Entorno técnico estratégico, escalable y extensible.
- Prototipo usado como prueba de concepto.
- Disponer de fuentes de datos de calidad.
- Métricas y clasificaciones comunes establecidas por la organización.
- Modelo de metadatos escalable.
- Gobierno de los datos por la organización.

## 5. Clasificación y agrupación

La revisión bibliográfica realizada para identificar factores críticos de éxito, mejores prácticas y recomendaciones de cómo llevar a cabo un proyecto BI ha de permitir identificar los requisitos deseables de la 'Metodología' más adecuada a un proyecto BI concreto.

En primer lugar hemos analizado, clasificado y agrupado todos FCE, mejores prácticas y recomendaciones enumeradas en el apartado anterior en tres grupos básicos:

- **Factores Primarios:** los aspectos o factores propuestos por más de cinco autores.
- **Factores Secundarios:** los aspectos o factores propuestos por más de un y menos de

cinco autores.

- **Factores de Autor:** aquellos que sólo han sido identificados por un solo autor.

Los autores que proponen alguno de los factores primarios y secundarios, usan terminología y semánticas ligeramente distintas. A continuación, intentamos describir estos FCE con una terminología y semántica unificada respecto a todas las apariciones del factor en la bibliografía consultada.

Por lo tanto, un proyecto BI ha de:

### Factores Primarios

- Estar enfocado a la gestión de datos.
- Usar una metodología de gestión del proyecto.
- Considerar primordialmente las necesidades y características del decisor.
- Ser una herramienta de soporte a la gestión de la organización.
- Proporcionar resultados rápidos ya en las primeras fases.
- Usar herramientas seleccionadas con cuidadosamente.
- Tener los usuarios finales involucrados en el proyecto.

### Factores Secundarios

- Disponer de recursos adecuados.
- Contemplar la cultura organizativa.
- Disponer de un diseño de *data warehouse* adecuado.
- Tener un equipo de trabajo capacitado y con experiencia y conocimiento.
- Tener el soporte y apoyo de la organización y dirección.
- Estrategia y dirección del BI alineadas con estrategia del negocio.
- Estar orientado a la gestión del cambio de la organización.
- Hacer cargas de datos incrementales.
- Tener una definición clara y precisa de los resultados e informes para el usuario.

Los Factores Primarios tienen una media de 9,4 ocurrencias, mientras que para los Secundarios la media es de 3,4. El rango de número de apariciones para los Factores Primarios es de [15, 6] y para los Secundarios es [5, 2]. Los Factores Primarios corresponden al 66% de todos los FCE identificados en la revisión bibliográfica realizada, mientras que Factores Secundarios representan el 33% restante. Por lo tanto, entendemos que la selección de los FCE propuestos es representativo de los FCE de un proyecto BI y que toda metodología de gestión de proyectos BI debería soportar.

## 6. Conclusiones

Con el surgimiento de la Inteligencia del Negocio o *Business Intelligence* en los últimos años, todas las organizaciones hacen esfuerzos para crear o mejorar sus procesos de decisión y sistemas decisionales. Una gran cantidad de nuevos proyectos de *Business*

*Intelligence* aparecen constantemente, pero la experiencia no es tan buena como era de esperar, porque una enorme cantidad de proyectos de BI (85%) no alcanzan sus objetivos iniciales.

Una gran diversidad y heterogeneidad de enfoques metodológicos para la gestión de proyectos BI muestra el estado de la novedad y la inexperiencia que todavía existe en este ámbito. Elegir una metodología de BI no es una tarea fácil. La implementación y gestión de un proyecto de BI puede suponer múltiples fases y más de 900 tareas. Por lo tanto, no es tan fácil de identificar una metodología de soporte satisfactoriamente a un proyecto de tanta complejidad y en todos sus aspectos y dimensiones.

En este artículo hemos resumido los principales enfoques metodológicos existentes para la gestión e implementación de un proyecto de BI. Hemos detectado las principales limitaciones de éstas y analizado cuales tendrían de ser las características de 'la metodología de proyectos BI'. Para identificar estas características, nos hemos basado en los Factores Críticos de Éxito de proyectos de *Business Intelligence* que han sido propuestos recientemente en las publicaciones académicas y científicas más relevantes.

## Referencias

- [1] U.M. Fayyad. Tutorial report. Summer school of DM. *Monash Uni Australia*.
- [2] R. Preston. Business Intelligence Still In Its Infancy. *Information Week*. Último acceso: julio 2010, <<http://www.informationweek.com/story/showArticle.jhtml?articleID=196801521>>.
- [3] P. Chowdhary, G.A. Mihaila, H. Lei. Model Driven Data Warehousing for Business Performance Management. *ICEBE 2006*: pp 483-487.
- [4] P. Chowdhary, K. Bhaskaran et al. Model Driven Development for Business Performance Management. *IBM Systems Journal*, vol 45, nº 3, pp 587-605.
- [5] B. Afolabi, O. Thiery. *Using Users' Expectations to Adapt Business Intelligence Systems*. Último acceso: julio 2010, <<http://arxiv.org/ftp/cs/papers/0608/0608043.pdf>>.
- [6] V. Stefanov, B. List. Bridging the Gap between Data Warehouses and Business Processes: A Business Intelligence Perspective for Event-Driven Process Chains. *EDOC 2005*: pp. 3-14.
- [7] J. Rowan. *Design Techniques for a Business Intelligence Solution*. Auerbach Publications 2003.
- [8] T. Bäck. Adaptive business intelligence based on evolution strategies: some application examples of self-adaptive software. *Inf. Sci.* 148(1-4): pp. 113-121.
- [9] M.K. Brohman, M. Parent, M. Pearce, N. Wade. The Business Intelligence Value Chain: Data-Driven Decision Support in a Data Warehouse Environment: An Exploratory Study. *HICSS 2000*.
- [10] L.T. Moss. Business Intelligence Methodologies, Agile with Rigor. *Cutter IT Journal*. vol. 14, no 12, pp. 19-26.
- [11] S. March, A.R. Hevner. Integrated decision support systems, A data warehousing perspective. *Decision Support Systems*, 2005.
- [12] Y. Guo, S. Tang, Y. Tong, D. Yang. Triple-driven data modeling methodology in data warehousing: a case study. *DOLAP 2006*: pp: 59-66.
- [13] D. Dori, R. Feldman, A. Sturm. An OPM-based Method for Transformation of Operational System Model to Data Warehouse Model. *SwSTE 2005*, pp. 57-66.
- [14] C. Kaldeich, J. Oliveira e Sá. Data Warehouse Methodology: A Process Driven Approach. *CAiSE 2004*, pp. 536-549.
- [15] L. Niu, G. Zhang. A Model of Cognition-Driven Decision Process for Business Intelligence. *Web Intelligence*, 2008 pp. 876-879.
- [16] J. Thomann, D.L. Wells. Implementing Data Warehousing Methodology- Guideline for Success. *TDWI 2000*.
- [17] L.T. Moss, Sh. Atre. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision Support Applications*. Addison Wesley Longman, 2003. ISBN-10: 0201784203.
- [18] G.R. Gangadharan, S.N. Swami. Business Intelligence Systems Design and Implementation Strategies. *IT1 2004*, pp. 139-144.
- [19] L.T. Moss. Ten Mistakes to avoid for Data Warehouse Projects Managers. *TDWI'S best of Business Intelligence Vol 3*, pp. 16-22.
- [20] R. Cicchetti et al. (Eds.). A Comparison of Data Warehouse Development Methodologies Case Study of the Process Warehouse. *DEXA 2002, LNCS 2453*, pp. 203-215, 2002.
- [21] T.N. Huynh, J. Schiefer. Prototyping Data Warehouse Systems. *DaWaK 2001*, pp. 195-207.
- [22] W. Yang, P. Hou, Y. Fan, Q. Wu. The Research of an Intelligent Object-Oriented Prototype for Data Warehouse. *ICIC (1)*, 2006, pp. 1300-1305.

- [23] R. Winter, B. Strauch. A Method for Demand-driven Information Requirements Analysis in Data Warehousing Projects. *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*.
- [24] H. Engström, S. Chakravarthy, B. Lings. A User-centric View of Data Warehouse Maintenance Issues. *BNCOD 2000*, pp. 68-80.
- [25] D. Schuff, K. Corral, O. Turetken. Comparing the Effect of Alternative Data Warehouse Schemas on End User Comprehension Level. *ICIS'05* <<http://mis.temple.edu/sigdss/icis05/>>.
- [26] J. Fernández, E. Mayol, J.A. Pastor. Agile Business Intelligence Governance: Su justificación y presentación. *ITSMF 2008*, <[http://www.uc3m.es/portal/page/portal/congresos\\_jornadas/congreso\\_itsmf/Agile%20Business%20Intelligence%20Governance.pdf](http://www.uc3m.es/portal/page/portal/congresos_jornadas/congreso_itsmf/Agile%20Business%20Intelligence%20Governance.pdf)>.
- [27] J. Thomann, D.L. Wells. Evaluating Data Warehousing Methodologies- An Evaluation Process. *TDWI 1999*.
- [28] B. Azvine, Z. Cui, D.D. Nauck. Towards real-time business intelligence. *BT Technology Journal Vol 23 No 3 July 2005*, pp. 214-225.
- [29] K.R. Quinn. Establishing a culture of Measurement, a practical guide to BI. *White Paper Information Builders 2003*.
- [30] J. Becker, L. Vilkov, C. Brelage. Multidimensional Knowledge Spaces for Strategic Management - Experiences at a Leading Manufacturer of Construction and Mining Equipment. *DEXA Workshops 2004*, pp. 482-487.
- [31] A. Counihan, P. Finnegan, D. Sammon. Towards a framework for evaluating investments in data warehousing. *Inf. Syst. J.* 12(4), pp. 321-338.
- [32] M. Gibson, D. Arnott, I. Jagielska. Evaluating the Intangible Benefits of Business Intelligence: Review & Research Agenda. *IFIP TC8/WG8.3 International Conference*, 2004.
- [33] A. Faulkner, A. MacGillivray. A business lens on business intelligence - 12 tips for success. *ODTUG 2001*.
- [34] T. Chenoweth, K. Corral, H. Demirkan. Seven key interventions for data warehouse success. *Commun. ACM* 49(1), pp. 114-119.
- [35] M.D. Solomon. Ensuring a successful data warehouse initiative. *ISM Journal winter 2005*, pp. 26-36.
- [36] D. Briggs, D. Arnott. Decision Support Systems Failure: An Evolutionary Perspective (Working Paper. No. 2002/01). Melbourne, Australia: *Decision Support Systems Laboratory, Monash University*.
- [37] D. Briggs. A Critical Review of Literature on Data Warehouse Systems Success/Failure (Working Paper. No. 2004/01). Melbourne, Australia: *Decision Support Systems Laboratory, Monash University*.
- [38] B.H. Wixom, H.J. Watson. An empirical investigation of the factors affecting data warehouse success. *MIS Quarterly Vol. 25 No. 1*, pp. 17-41, marzo 2001.
- [39] B.H. Wixom, H.J. Watson. The Current State of Business Intelligence. *IEEE Computer (COMPUTER)* 40(9), pp:96-99.
- [40] D. Sammon, P. Finnegan. The Ten Commandments of Data Warehousing. *The DATA BASE for Advances in Information Systems*, Vol. 31, No. 4.
- [41] R. Weir, T. Peng, J.M. Kerridge. Best Practice for Implementing a Data Warehouse: A Review for Strategic Alignment. *DMDW 2003*.
- [42] R.S. Abdullaev, I.S. Ko. A Study on Successful Business Intelligence Systems in Practice. *JCIT 2(2)*, pp. 89-97.
- [43] I.S. Ko, R.S. Abdullaev. A Study on the Aspects

of Successful Business Intelligence System Development, *Computational Science - ICCS 2007*.

[44] W. Yeoh, J. Gao, A. Koronios. Towards a Critical Success Factor Framework for Implementing Business Intelligence Systems: A Delphi Study in Engineering Asset Management Organizations. *CONFENIS 2007*.

## Notas

<sup>1</sup> Un **Data Warehouse (DW)** es una base de datos usada para generación de informes. Los datos son cargados desde los sistemas operacionales para su consulta. Pueden pasar a través de un almacén de datos operacional para operaciones adicionales antes de que sean usados en el DW para la generación de informes (Traducción libre de la introducción al concepto que se encuentra en la Wikipedia en inglés el 24/6/2011: <[http://en.wikipedia.org/wiki/Data\\_warehouse](http://en.wikipedia.org/wiki/Data_warehouse)>). Se suele considerar que el término equivalente en castellano es "Almacén de datos": "En el contexto de la informática, un almacén de datos (del inglés data warehouse) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Se trata, sobre todo, de un expediente completo de una organización, más allá de la información transaccional y operacional, almacenado en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos (especialmente OLAP, procesamiento analítico en línea)". (Wikipedia en castellano 24/6/2011: <[http://es.wikipedia.org/wiki/Almac%C3%A9n\\_de\\_datos](http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos)>).

<sup>2</sup> Un **Data mart** es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades. <[http://es.wikipedia.org/wiki/Data\\_mart](http://es.wikipedia.org/wiki/Data_mart)>.

José María Arce Argos  
Gerente de Business Intelligence & CRM en Oesía

<chema.arce.argos@gmail.com>

## Modelos de construcción de Data Warehouses

### 1. Introducción

Es complejo intentar desvelar, a través de unas líneas todo lo que puede incluir un modelo de *Data Warehouse*<sup>1</sup> (DW) e incluso llegar a un acuerdo en su significado y tipo de estructuración física.

En cualquier caso, a estas alturas de la madurez de *Business Intelligence* (BI) en España, todos coincidimos en que el único pilar básico para el desarrollo de soluciones de negocio es sin lugar a dudas el DW. Almacenamiento pensado, diseñado y construido por y para unas necesidades agresivas de análisis, análisis completamente impredecibles.

En España llevamos más de 16 años hablando de DW y de BI, aunque lamentablemente tras las siglas DW se esconde un gran desconocido, obviando su objetivo, sus técnicas y sus posibles tipos de modelos. Desconocido al haberse puesto de moda desde 1998 otros "tecnicismos" informáticos (*Customer Intelligence*, BSC, CRM, etc.).

Las nuevas herramientas, más sofisticadas, han derivado erróneamente en profesionales más técnicos y con menos visión del negocio, obviando temas muy relacionados con los modelos de datos de DW y su representación formal en un modelo específico. En cualquier caso, empezamos por el principio.

Los modelos de un *Data Warehouse* mantienen necesariamente una relación directa con los tipos de almacenamiento que se determinen e incluso deben y pueden ser optimizados según decenas de criterios. Los cuales van desde el tipo de consulta más demandado, optimización de acceso por claves e incluso tablas agregadas, hasta la optimización según la herramienta de explotación utilizada. Por todo ello, no es sencillo establecer un único criterio o patrón de cara a la correcta construcción de un DW. El presente artículo únicamente pretende sentar unas bases mínimas para ayudar a la selección del almacenamiento más adecuado, así como algunos consejos de cara a realizar un modelo de DW que permita un éxito en su iniciativa de BI, o por lo menos, adquirir algún conocimiento para poder comprender o entender a los consultores o responsables de desarrollarlo.

Además, es importante reflejar alguna idea previa sobre todo ello. Un buen sistema de BI

**Resumen:** A través del presente artículo nos adentraremos en las técnicas, modelos y consejos más básicos para crear valor en nuestras organizaciones. Todo ello de la mano del *Data Warehouse* y de sus posibles modelos de implementación. El artículo profundiza, comparando algunos de los tipos de modelos posibles y sus consecuencias. Prestando especial atención en la necesidad de las empresas de poder medir, como única salida hacia el éxito empresarial.

**Palabras clave:** Datawarehouse, modelo en estrella, modelo E-R, modelos multidimensionales.

### Autor

**José María Arce Argos** es Gerente de *Business Intelligence & CRM* en Oesía. Ha desarrollado su carrera profesional en empresas tan importantes como Leinsa, IBM, Ernst & Young, SAS Institute, Bull España, Altran SDB e Inad. Cuenta con 22 años de experiencia en consultoría, de los cuales 16 años dedicados al *Business Intelligence*. Redactor y colaborador en la revista "Gestión del Rendimiento", colaborador en TodoBI, profesor durante 10 años en el Máster de "Sistemas de Información e Investigación de Mercados" de la *Business Marketing School* (ESIC), ponente en diversos eventos en IIR, SAS, CUORE Oracle, Univ. de Zaragoza, etc. A lo largo de su trayectoria profesional ha diseñado, coordinado, participado o dirigido iniciativas de BI en clientes como Publiespaña, British American Tobacco, Halifax, ONT, REE, Continente (Carrefour), Grupo Abengoa, ABN Anro Bank, B. Santander, Open Bank, Ing Direct, Dirección Gral. de Estadística (JCYL), TeleMadrid, Hospital San Cecilio de Granada, DFA, RED.ES, Egailan, INEM, INSS, Policía Municipal de Madrid, Canal de Isabel II, etc. Es autor del Blog BIB: <<http://josemariaarce.blogspot.com/>>.

requiere de un DW, de hecho casi todas las iniciativas de BI tienen por debajo o se apoyan mayoritariamente en un DW, aunque no se diga e incluso se oculte. También es cierto que las herramientas de explotación cada día son más poderosas e incluso permiten realizar agrupaciones, desgloses, etc., incluso contra un modelo, digamos cortésmente, poco evolucionado. Lo cual posiblemente denota un error en el planteamiento, pues siempre será mejor que ciertos conceptos figuren en modelo que generados al vuelo por una herramienta, pues podríamos tener el mismo concepto "N" veces y calculado de formas diferentes. En el caso de producirse esta situación implicaría la muerte del sistema BI, básicamente por la desconfianza sobre la calidad de los datos e incluso por posibles resultados contradictorios. En este ejemplo no está fallando el BI, está fallando el diseño del modelo de DW.

Para acabar esta introducción, comentar que en decenas de clientes hemos escuchando quejarse de importantes problemas de rendimiento en sus DW, de la necesidad de poner más maquina, cambiar de herramienta, meter nuevas estructuras más optimizadas, etc. En la mayoría de los casos los problemas no se solucionan así. El 90% de los problemas de los sistemas de BI residen en un mal modelo de datos (DW).

### 2. Almacenamientos y alternativas

Sin entrar en las características que debe cumplir un DW, pues creo que ya está todo escrito, mencionar que el DW tiene como única finalidad la capacidad de analizar la información de interés desde diversos puntos de vista, a lo largo del tiempo, entre otras muchas cosas. Ello obliga a que esté diseñado y construido de una forma específica, siendo pues necesario cumplir con el término OLAP (*OnLine Analytical Process*), en contra de los diseños tradicionales, los cuales ahora los denominamos OLTP (*OnLine Transaction Processing*).

Años antes del DW ya se intentaron montar este tipo de sistemas, a base de duplicar el mundo operacional. El tiempo demostró que dicho camino no era el más efectivo y que tenía serias limitaciones, tomando fuerza la necesidad de que el DW esté diseñado con características OLAP.

Recordemos que los sistemas OLAP cumplen con más de 50 posibilidades de "navegación" por la información, como rotar, bajar, profundizar, expandir, colapsar, etc. Todo esto y mucho más es simplemente imposible en los sistemas tradicionales (OLTP), motivo extra para olvidarse de replicar sistemas, solución ya probada (años 70) y que no dio los resultados esperados.

Existen diversas alternativas para la arquitectura de un DW. En cuanto a tipos de almacenamiento, mencionaremos las más habituales:

- Solución **MOLAP** (*Multidimensional OLAP*): Montar el almacenamiento bajo una base de datos multidimensional propietaria (MDDDB de SAS, ESSABE, etc.).
- Solución **ROLAP** (*Relacional OLAP*): Montar el almacenamiento bajo una base de datos relacional (SQLServer, Oracle, DB2, etc.).
- Soluciones **Híbridas**, denominadas **HOLAP**. Las cuales combinan ambas alternativas en una única solución, interesante y sabia alternativa según las necesidades.

Debemos considerar que detrás de unas siglas se esconden unas claras diferencias tecnológicas y unas ventajas e inconvenientes que deben conocerse a priori. Las diferencias más llamativas o significativas las podemos clasificar en cuatro puntos:

**Necesidades de procesamiento y tiempo de respuesta:** Como no existen soluciones ideales cada una ofrece diversas capacidades: Mientras los modelos ROLAP deben calcular y buscar los datos al vuelo, lo cual consume tiempo y retarda la respuesta, los modelos MOLAP tienen todo precocinado y sus tiempos son mejores. Lógicamente las necesidades *batch* para crear los MOLAP son claramente superiores a los ROLAP. También mencionar que los volúmenes de datos gestionados en ROLAP son muy, pero que muy superiores a los MOLAP, a más datos más tiempo. En algún cliente he intentado crear MOLAP con importantes volúmenes y simplemente fue imposible. Cada tecnología vale para una cosa.

**Gestión e incremento en las visiones de negocio:** La visiones de negocio o dimensiones, utilizando términos de BI, son casi infinitos en el mundo ROLAP, pero no por el contrario en el mundo MOLAP. Los modelos MOLAP empiezan a flaquear en la frontera de las 10 dimensiones, pues al crear el MOLAP se deben de calcular todos los cruces o combinaciones posibles, por lo tanto ojo al número de combinaciones usadas.

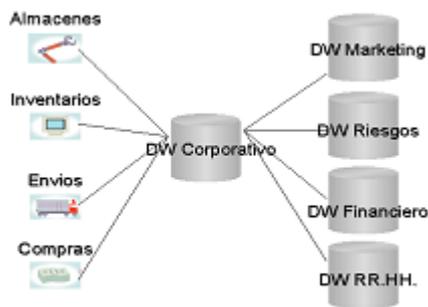


Figura 1. Arquitectura ROLAP para Data Warehouse.

**Necesidades cambiantes de agrupaciones o consolidaciones:** En este caso los modelos ROLAP tienen una capacidad más camaleónica, pues todo se calcula al vuelo, siendo más flexibles y potentes. Por contra, los modelos MOLAP requieren de una precocina en *batch*. En general, incluir cualquier concepto olvidado en un DW es algo laborioso, especialmente en función de donde se deba incluir. Incluir una nueva métrica en una tabla de hechos (*tabla fact*)<sup>2</sup> es lo más costoso, pues no vale solo con ponerla, será necesario recargar esa métrica y todo su historial asociado. Es casi mejor pecar por exceso, un olvido en una tabla de hechos se paga muy caro.

**Volúmenes y capacidades:** Posiblemente el éxito de los modelos ROLAP viene especialmente por esta capacidad. Son los únicos que pueden enfrentarse a volúmenes de Gb e incluso Teras de información, algo simplemente imposible en el mundo MOLAP. Lo cual lógicamente hace que sus respuestas no ofrezcan unos tiempos óptimos, pero en otras soluciones simplemente no se pueden hacer. La arquitectura ROLAP es la clara vencedora por capacidad y flexibilidad. Dejando el mundo MOLAP a soluciones o necesidades muy controladas, pequeñas, departamentales o, hablando en cristiano, de "andar por casa". Es normal encontrarse con evoluciones en base a *Data Marts*<sup>3</sup>, decisión muy extendida e incluso razonable años atrás. Especialmente usadas en las primeras iniciativas de BI, pues resultan rápidas de desarrollar y nos permiten que en nuestras compañías, "los mayores", entiendan las capacidades y bondades de dichos sistemas. Les gustarán mucho y obtendremos financiación para continuar. Pero debemos ser conscientes de las limitaciones que tendremos a medio o largo plazo, pues tendremos silos de información y limitadas capacidades de cruces, lo cual terminará obligando a desarrollar más y más *Data Marts*.

Posiblemente habremos salido de una pesadilla operacional para entrar en una pesadilla informacional, gracias a la proliferación de *Data Marts* en nuestra organización.

La otra, o mejor dicho, la única alternativa eficaz, flexible y que siempre nos permitirá un crecimiento y mantenimiento acorde a nuestras necesidades, si hablamos de DW, entendiéndolo como una solución integrada, consolidada y de amplia cobertura empresarial (no me atrevo a decir corporativo, pero lo pienso), es montar el DW bajo arquitectura ROLAP, como se ilustra la figura 1.

Bajo esta figura podemos ofrecer lo mejor de ambos mundos. Dado la limitación física de los MOLAP, dejemos los cimientos del BI bajo ROLAP, y diseñemos estructuras MOLAP para accesos controlados, rápidos y eficaces a los departamentos o necesidades

específicas. Como ya hemos comentado, una solución Híbrida (HOLAP) es una sabia salida que ofrece lo mejor de cada arquitectura.

Una vez tratado ligeramente el tema de los almacenamientos posibles y considerando que nos decantamos por su gestión bajo un motor relacional, debemos considerar que los modelos tradicionales no están pensados para las necesidades que debe de soportar un DW. Lo cual requiere un diseño específico, como ya hemos comentado, el cual permita su explotación OLAP. Tradicionalmente nos encontramos con dos tendencias: los modelos en estrellas y los modelos copo de nieve.

**3. Tipos de modelos en DW**

Tal vez antes de entrar en faena, me gustaría comentar que en definitiva hablamos de lo mismo, pues de un modelo en estrella llevo a un copo de nieve y viceversa, por lo tanto, no merece la pena pegarse en qué es mejor (guerrilla promovida por los fabricantes, a mediados de los 90, y sus intereses particulares), pues finalmente son más de lo mismo.

Aquellos que han sido alumnos míos durante mis 10 años siendo profesor en un Máster de prestigio sobre "Sistemas de Información e Investigación de Mercados", bien lo saben y ha quedado demostrado. Sin embargo existen importantes limitaciones e incluso costes ocultos, como los asociados a los mantenimientos, a las mejoras, a la dificultad en los nuevos procesos, etc.

Con independencia de un tipo de modelo u otro, es conveniente quedarnos con algunas ideas que nos van a permitir buscar la mejor solución y evitar algunas trampas en el diseño de un DW. Con el ánimo de ser lo más claro posible evitaré usar tecnicismos. Entre las muchas coletillas adquiridas en estos años y relacionadas con el DW os comentaré:

■ **"Divide y vencerás"**: Se puede empezar a diseñar un DW sin conocer al 100% las necesidades de toda la organización, desglosa el gran proyecto por dominios de información y "ataca" uno a uno, sin perder la visión del gran sistema. Recuerda: "*Piensa en grande, haz en pequeño*". También aplicable al "*divide y vencerás*" a la forma de diseñar las dimensiones. Considerando la forma de acceso del usuario final y la "navegación" deseada, considera que cuanto más "normalizado" se diseñe tendrás más flexibilidad y un mantenimiento más sencillo que será más comprensible por los DBA más tradicionales.

■ **"Diseñar cabezas de ciervo"**: Existen trampas en un DW como son las "*trampas de abanico*" y "*trampas de abismo*", las cuales exclusivamente ocurren bajo un modelo mal diseño. Un modelo simple, es como la cabeza de un ciervo. La cabeza es la tabla que contiene los valores numéricos a analizar, las métricas o indicadores, la *tabla Fact*. Mientras los



cuernos son sus dimensiones. Las dimensiones jamás se tocan o cruzan entre sí. El único punto en común que tienen son la cabeza del ciervo. Los cuernos solamente tienen un punto de contacto con la cabeza... con este simple y claro ejemplo, nunca tendrás problemas

■ **"Si metes basura, sacas basura"**: Expresión asociada a la importancia de los procesos ETL (*Extract, Transform and Load*) para la carga de los datos, a sus controles de calidad del dato, etc. Todos los esfuerzos, todos los diseños y cualquier otra actividad no tendrá ningún valor, sin la credibilidad y calidad de los datos.

■ **"El exceso de análisis conduce a la parálisis"**: Es más que discutible como se abordan algunos proyectos de DW, especialmente los grandes (corporativos). Como ya hemos comentado se debe fragmentar por dominios e irlo abordando o desarrollando por iteraciones, en caso contrario estamos muertos, no ofreceremos resultados nunca tendremos otro bonito fracaso.

**4. ¿Qué tipo de modelo diseño?**

Como hemos comentado, aunque ambos modelos pueden derivar el uno en el otro, en función de aplicar técnicas de normalización o desnormalización, el resultado final sí tiene su importancia. Hoy por hoy casi todas las herramientas de explotación pueden "atacar" a cualquier modelo, pero solamente algunas son capaces de aprovechar al máximo las capacidades de los modelos copo de nieve, en general, todas se mueven por el mundo de las estrellas.

Tradicionalmente los fabricantes de herramientas, basadas en modelos estrellas, han sido especialmente críticos con su competencia, pues se posicionaba con otro tipo de modelo. Ello provocó y todavía se nota, mensajes muy erróneos lanzados al mercado. Del mismo modo, actualmente nos quieren vender ideas como: hacer un DW en 10 minutos, todo ello sin técnicos, o de la vital importancia de disponer de análisis en un iPad, etc., cuando la mayoría de los sistemas funcionan por los pelos y no son adecuadamente explotados. En general, las necesidades de los usuarios finales no están alineadas con las estrategias comerciales de algunos fabricantes, mal que les pese.

En función de sus necesidades, conocimientos, herramientas y estrategias pueden optar por un tipo de modelo u otro. La **figura 2**

visualiza un modelo en estrella, en el centro la tabla de hechos donde se encuentran los indicadores numéricos a analizar, a su alrededor una tabla por dimensión o visión del negocio. Dentro de cada una se encuentran todos los conceptos asociados a dicha dimensión.

No soy enemigo de los modelos en estrella, pues mi maestro y casi padrino en BI fue Ralph Kimball. También es cierto que por aquellos años, aquella forma de diseñar fue una revolución en sí misma y tampoco había muchas más alternativas.

Puntualizar que en un proyecto de BI, no existe una sola estrella, de hecho, algunos hablan de constelaciones... En una base de datos de DW, podríamos encontrar decenas de *tablas fact* y sus dimensiones asociadas, comunes y no comunes entre diversas *tablas fact*. Pues en caso contrario, si piensan que con una estrella lo tienen todo resuelto, permitanme dos comentarios:

- No malgaste su tiempo y dinero, pues algo no he sabido explicarle o para usted un fichero texto ya sirve para tomar decisiones.
- Hágalo usted mismo, con una tabla dinámica de Excel.

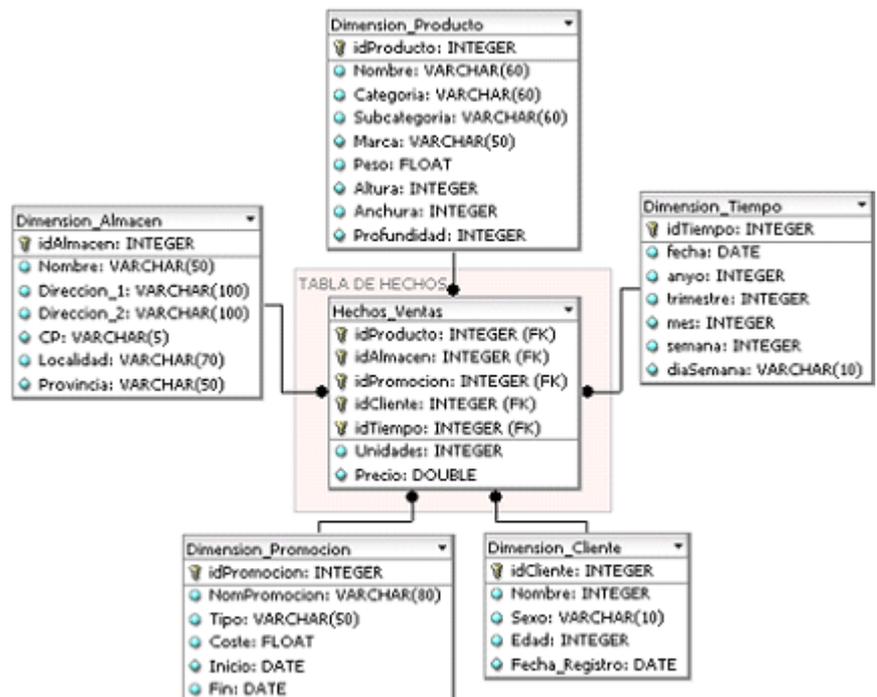
Queda claro que de cara a intentar explicar temas de modelos, nos apoyamos con la expresión mínima (una sola tabla de hechos y sus dimensiones), las cual son ilustradas en las figuras, pero la realidad no es tan simple. La otra opción es el modelo copo de nieve, el cual tiene más o menos la apariencia que muestra la **figura 3**.

Los defensores de los modelos en estrella, modelos completamente desnormalizados, han basado su defensa en algunas afirmaciones que vamos a repasar, aunque el tiempo ha demostrado que "no es oro todo lo que reluce":

■ Fácilmente entendible, consultas sencillas (por usar pocas tablas). Lo comparto, pues casi cualquiera puede entender a simple vista este modelo, digamos que no asusta al verlo. Supongo que su expansión se encuentra relacionada con su relativa sencillez al diseñarlo, pues deja todo el "marrón" a la herramienta, sobre la cual se debe definir la lógica de navegación. Como ya he indicado lo considero un error potencialmente alto. Las estrellas simples sobre BBDD relacionales son lo más parecido conceptualmente a las BBDD multidimensionales, a los castigados cubos, e incluso a las tablas dinámicas del Excel, etc. Supongo que por ello, algunos defienden el Excel como una herramienta de BI, sin más comentarios.

■ Son más rápidos (por usar pocas tablas, pocos *joins*). Directamente, es una verdad a medias. Solamente podría ser así con poco volumen de información. Además, las políticas de claves, bajo este tipo de estructuras, no favorecen las búsquedas, aparte de ser francamente difícil planificar a priori la clave o claves más adecuadas, pues nunca podemos saber por dónde nos vendrá la consulta.

Otro problema más serio, y que solamente por ello **debería hacernos replantear toda la solución**, es que no es posible implementar integridad referencial. Por ello nuestro gestor nunca será capaz de garantizar



**Figura 2.** Modelo en estrella para un *Data Warehouse*.

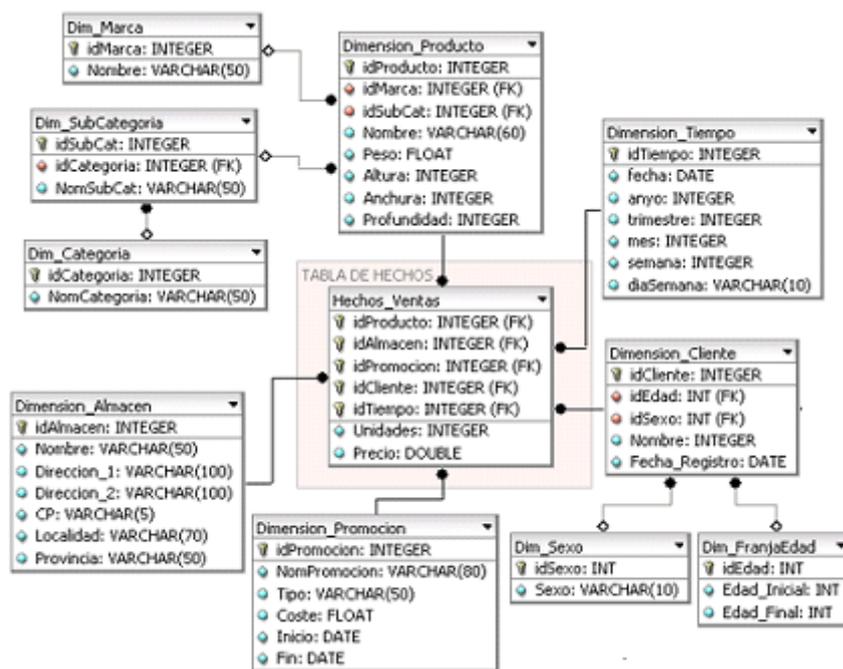


Figura 3. Modelo copo de nieve para un Data Warehouse.

la consistencia y la calidad del dato. Obligando a desarrollar unos procesos ETL muy estudiados y minuciosos.

En resumen, por ahorrar tiempo en el análisis, en el diseño y no complicarse la vida, algunas veces "te entregan unas estremitas muy monas". Nadie te explicará y te pondrá sobre aviso de que:

- Ese tipo de diseño puede implicar perder calidad y consistencia en los datos, pues no puedes definir integridad referencial entre campos de una misma tabla.
- Ello ocasiona pasar el control de la consistencia a los procesos ETL, lo cual implica un esfuerzo extra y un coste elevado.
- Dicha forma de diseñar delega la lógica de la navegación a la herramienta final y no está implícita en el modelo, esto es igual a problemas tarde o temprano.
- Dificultad para determinar índices para mejorar respuesta.
- Relaciones padres e hijos no identificados.
- La velocidad que supuestamente ganamos por no hacer *joins* por diversas tablas, la podemos perder por el gran volumen de la dimensión (en única tabla) y por la posible ausencia de índices adecuados, etc.

Creo que ello, unido a otros puntos difíciles de explicar sobre papel nos debería por lo menos cuestionar qué modelo establecer y avisarnos sobre todo lo que leemos por Internet. Lo más sensato: Haga una prueba con el escenario más parecido posible a la realidad, también respecto al volumen de información.

### 5. Conclusión

Los negocios están cambiando constante-

mente debido a cambios económicos, evoluciones tecnológicas, alteraciones en el mercado, impactados por diversos cambios culturales y sociales e incluso por fenómenos meteorológicos.

Todo ello obliga a replantearse las estrategias actuales y debería provocar una transformación en nuestro propio negocio. Así, un factor clave de éxito, e incluso de supervivencia, viene derivado de la capacidad de las organizaciones de gestionar de forma eficiente sus datos, y transformarlos en información útil y disponible para acertar en las decisiones. Esto y solo esto, es *Business Intelligence*.

**Business Intelligence no es tecnología, es negocio y es estrategia.** BI implica muchas cosas, pasando por la **vocación de medir para actuar en consecuencia**, gran problema pendiente en las organizaciones.

Actuar no es hacer un informe. Es la capacidad de controlar y gestionar las organizaciones, basada en datos e informaciones veraces y no en hipótesis. Es la capacidad de alinear la estrategia con las operaciones, es la capacidad de orientarse realmente hacia el cliente, es la capacidad de entender, es comprender y transmitir los objetivos empresariales y su desempeño, es la capacidad de crear consenso en la organización, derivando todo ello en un **cambio cultural**.

En caso contrario, nos quedaremos con unas decenas de informes, tras unas inversiones muy importantes.

$$BI = \text{Cambio Cultural}$$

$$= \text{Capacidad} + \text{agilidad} + \text{decisiones}$$

Es necesaria la redefinición de las competencias del BI dentro de las organizaciones. Pues siendo equipos de vital importancia y estratégicos, tan importantes o más que departamentos de marketing, financieros, recursos humanos, etc. suelen brillar por su ausencia. Delegando sus competencias en un equipo técnico más o menos formado pero con una visión alejada del verdadero sentido del BI. No podemos mostrar un cambio cultural donde no existe estrategia, aquellas empresas que han dado el salto y tiene una infraestructura real, tanto en recursos humanos como materiales, se están posicionando. Ustedes deciden...

Para terminar les dejo una frase que leí por Internet, la cual tomo prestada (gracias a su autor, aunque no recuerde quien fue): *"En la nueva económica, el grande ya no devorará al chico, sino el ágil le ganará al lento"*.

Bienvenido al apasionante mundo de los negocios, bienvenido a *Business Intelligence*.

### Bibliografía

- Ralph Kimball. *The data warehouse toolkit: Practical techniques for building dimensional data warehouse*, 1996.
- Bill Inmon. *Building the Data Warehouse*. 1st Edition. Wiley and Sons, 1992.

### Notas

<sup>1</sup> Un **Data Warehouse (DW)** es una base de datos usada para generación de informes. Los datos son cargados desde los sistemas operacionales para su consulta. Pueden pasar a través de un almacén de datos operacional para operaciones adicionales antes de que sean usados en el DW para la generación de informes (Traducción libre de la introducción al concepto que se encuentra en la Wikipedia en inglés el 24/6/2011: <[http://en.wikipedia.org/wiki/Data\\_warehouse](http://en.wikipedia.org/wiki/Data_warehouse)>).

<sup>2</sup> En las bases de datos, y más concretamente en un data warehouse, una **tabla de hechos** (o **tabla fact**) es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio. <[http://es.wikipedia.org/wiki/Tabla\\_de\\_hechos](http://es.wikipedia.org/wiki/Tabla_de_hechos)>

<sup>3</sup> Un **Data mart** es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades. <[http://es.wikipedia.org/wiki/Data\\_mart](http://es.wikipedia.org/wiki/Data_mart)>.

Óscar Alonso Llombart  
Analysis Manager, Penteo

<o.alonso@penteo.com>

## Data Governance: ¿qué?, ¿cómo?, ¿por qué?

### 1. Introducción

Llegar a obtener el valor real de los datos no es tarea sencilla. Recogemos y almacenamos datos provenientes de múltiples canales que a menudo se encuentran almacenados en diferentes sistemas de información y bases de datos sobre entornos tecnológicos y formatos heterogéneos. Aunque tengamos acceso directo a los datos, es difícil disponer de ellos dónde, cuándo y cómo los necesitamos, pero además los datos suelen estar "sucios", es decir, repletos de errores, omisiones e incoherencias.

Esta problemática es lo suficientemente importante como para hacer fracasar cualquier proyecto TIC (Tecnologías de la Información y de la Comunicación), iniciativa empresarial estratégica o incluso toda una compañía. La capa de datos de una organización es un componente crítico, sobre el que a menudo es fácil hacer suposiciones demasiado optimistas sobre su situación o bien ignorar la calidad real de los datos.

Por una parte existen datos que sólo se utilizan en un entorno tecnológico restringido para un proceso o una aplicación con impacto limitado, y por otra existen una serie de datos cuya importancia es fundamental porque definen las identidades más importantes (clientes, productos, empleados, proveedores...), y que deben ser compartidos por múltiples procesos, departamentos y líneas de negocio. Estos datos (los llamados "datos maestros") deben ser tratados como un activo estratégico.

Garantizar la calidad, integridad y exactitud de los datos es uno de nuestros principales retos. Obtener la visión única de los datos de manera transversal a través de los departamentos de una empresa, las distintas líneas de negocio o las distintas compañías de un grupo, es un factor crítico para facilitar la consecución de los objetivos de negocio.

Tener como objetivo unos datos de calidad es una filosofía que alinea la estrategia, la cultura empresarial, y la tecnología con el fin de gestionar los datos en beneficio propio. En pocas palabras, se trata de una auténtica estrategia competitiva, cada empresa tiene la oportunidad de diferenciarse mediante la calidad de sus datos.

¿Pero hasta qué punto afectan los datos

**Resumen:** La toma de decisiones está basada en la información que obtenemos de los datos empresariales. Toda toma de decisiones implica aceptar un riesgo, pero lo cierto es que no siempre es fácil disponer de datos rigurosos... Ante esta situación, ¿cómo podemos alcanzar el auténtico valor de los datos y ofrecer una visión consistente del rendimiento empresarial?, ¿cómo conseguir un adecuado análisis de la información teniendo en cuenta los cambios constantes que ocurren en nuestras organizaciones?

**Palabras clave:** Calidad de datos, Data Governance, data stewardship, gestión de los datos, gobierno de los datos, propiedad de los datos.

### Autor

Óscar Alonso Llombart es Ingeniero en Informática de Gestión por la Universidad Autónoma de Barcelona, Master en Ingeniería del Software por la Universidad Politécnica de Cataluña y Postgrado en Data Mining por la Universitat Oberta de Catalunya. Trabaja como Analysis Manager en Penteo. Cuenta con más de 15 años de experiencia en el ámbito de consultoría tecnológica en áreas como Business Intelligence, Datawarehousing, Corporate Performance Management, desarrollos a medida e implantación de metodologías de desarrollo. Es autor de numerosos artículos y estudios sobre la aplicación de los sistemas de inteligencia de negocio a las estrategias empresariales. Twitter: <@oalonsollombart>; LinkedIn: <<http://www.linkedin.com/in/oscaralonsollombart>>.

erróneos a nuestro negocio? Debido a la naturaleza dinámica de los datos, que típicamente se generan mediante numerosos procesos de negocio y fuentes de información que son combinadas, almacenadas y utilizadas en varios sistemas, es un importante reto establecer métodos para evaluar el impacto de los datos de poca calidad.

La mala calidad de los datos tiene un coste económico real, la eficiencia en los procesos se ve afectada debido a la escasez de datos de calidad, y no se alcanzan los beneficios potenciales de los sistemas tanto de los existentes como de nuevos proyectos.

De las investigaciones realizadas por Penteo se desprende que todavía existe un importante *gap* para conseguir una verdadera inteligencia de negocio. Si bien son muchas las compañías que han implantando sistemas de inteligencia de negocio en un porcentaje relevante, lo han hecho con proyectos aislados, dando respuesta a necesidades muy específicas. En la gran mayoría de las compañías las dificultades para encontrar y explotar adecuadamente datos e información respecto el estado y evolución del propio negocio son un denominador común (ver **figura 1**). Esta situación impacta invariablemente en el negocio en términos de aspectos económicos, confianza sobre los datos, cumplimiento de regulaciones, satisfacción y productividad.

### 2. La gestión de los datos

Los procesos de negocio se basan fuertemente en los sistemas de información, sistemas que interactúan entre ellos, que comparten la información y que deben ser capaces de comunicarse para poder prestar un servicio adecuado y eficiente a la organización. Además se toman decisiones estratégicas basadas en la información extraída de los sistemas, y hemos de disponer de información fiable para la buena gestión corporativa.

En esta situación hemos de ser conscientes que somos dependientes de la calidad de los datos que tenemos en nuestra organización. Los datos como entidad por sí misma no aportan valor añadido al negocio y las soluciones de inteligencia empresarial no son nada si no disponemos de datos fiables. Son aquellas compañías que han gestionado adecuadamente la calidad de los datos las que han evitado los problemas derivados de la toma de decisiones basada en información errónea.

La gestión de los datos es la primera pieza sobre la que sustentar una adecuada explotación de la información (ver **figura 2**), considerando los datos y posterior información inferida a partir de ellos como activos empresariales valiosos. Los datos y la información han de ser gestionados de manera cuidadosa, como cualquier otro activo, asegurando la

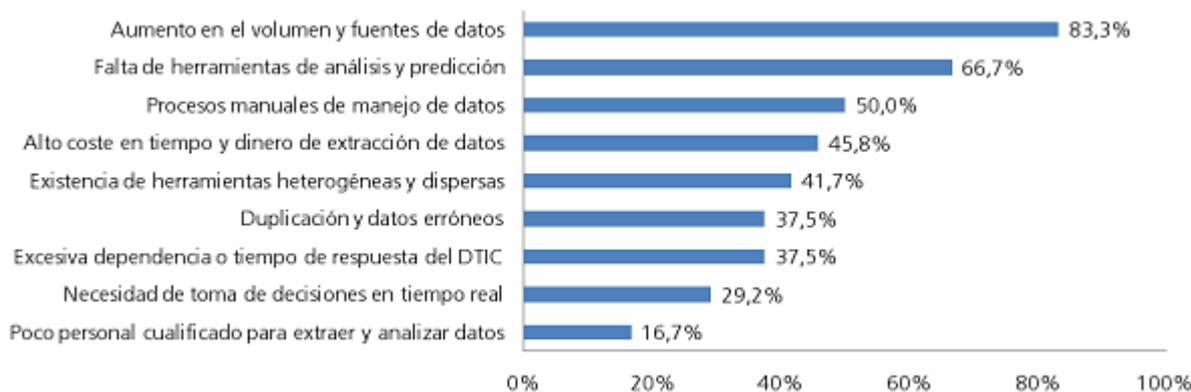


Figura 1. ¿Cuáles son los principales problemas en la toma de decisiones? Fuente: Penteo.

calidad, seguridad, integridad, disponibilidad y uso efectivo.

Los objetivos de la gestión de los datos son:

- Comprender las necesidades de información de la organización.
- Capturar, almacenar, proteger y asegurar la integridad de los activos de los datos.
- Mejorar de manera continua la calidad de los datos y de la información incluyendo la exactitud, integridad, integración, relevancia y utilidad de los datos.
- Asegurar la privacidad y la confidencialidad, y prevenir el uso no autorizado e inapropiado de los datos y la información.
- Maximizar el uso efectivo y el valor de los activos de los datos y la información.
- Controlar (y conocer) el coste de la gestión de los datos.
- Promocionar un uso y un conocimiento más amplio y profundo del valor de los activos de los datos.
- Gestionar la información de manera consistente a lo largo y ancho de la organización.
- Alinear la gestión de los datos y la tecnología necesaria con las necesidades del negocio.

La gestión de los datos debe ser vista como una función de negocio, únicamente compartiendo la responsabilidad de la gestión de los datos entre los usuarios propietarios de los datos y el departamento TIC llegaremos a obtener una auténtica ventaja competitiva mediante el adecuado uso de la información.

**3. Data Governance (la tecnología por sí sola no puede resolver el problema)**

*Data Governance...*<sup>1</sup> ¿qué es?, ¿por qué es importante?, ¿cuál es la relación entre gobierno y propiedad de los datos?, ¿incluye el concepto de gestión de los datos el gobierno de los datos?, ¿sabemos en qué costes está incurriendo nuestra organización por tener datos duplicados o por no disponer de definiciones estándares de datos comunes? ... Si no somos capaces de contestar a estas cuestiones, quizás debamos plantearnos una estrategia para hacer frente a la necesidad de comprender y utilizar los datos de manera más efectiva y eficiente.

Para alcanzar este objetivo las compañías han de implantar proyectos de *Data Governance*,

un conjunto de políticas y procedimientos que combinados establecen los procesos que supervisan el uso y gestión de los datos para transformarlos en un activo estratégico, con el objetivo de llevar a nuestra compañía a un nivel superior de "madurez en el uso de la información", mejorar la calidad de los datos y solucionar los posibles inconsistencias, gestionar el cambio en relación con el uso de los datos, y cumplir con regulaciones y estándares internos y externos.

*Data Governance* es la piedra angular sobre la que sustentan todas las prácticas relacionadas con la gestión de los datos, que interactúa e influencia con todas y cada una del resto de estas, como son los proyectos de calidad de los datos, integración de datos o *datawarehousing*<sup>2</sup>. El gobierno de los datos es el ejercicio de autoridad y control (planificación, monitorización y ejecución) sobre la gestión de los activos de datos, no gobierna los datos directamente sino que gobierna cómo los usuarios acceden a los datos a través de la tecnología.

El programa de *Data Governance* guía cómo han de actuar el resto de funciones de gestión de los datos, estableciendo los propietarios de los datos, tanto a nivel ejecutivo como operativo. Además ha de balancear adecuadamente objetivos contrapuestos como son el cumplimiento de regulaciones, que limitan el acceso a los datos, y los procesos de integración del negocio que amplían el acceso a estos. Las tareas que un programa de *Data Governance* ha de llevar a cabo son:

- Guiar a los gestores de la información en la toma de decisiones.
- Asegurar que la información se define de manera consistente y es comprendida por todos los actores implicados.
- Incrementar el uso y confianza de los datos como un activo de gran valor.
- Mejorar la consistencia de los proyectos a lo largo y ancho de la organización.
- Asegurar el cumplimiento de regulaciones internas y externas.

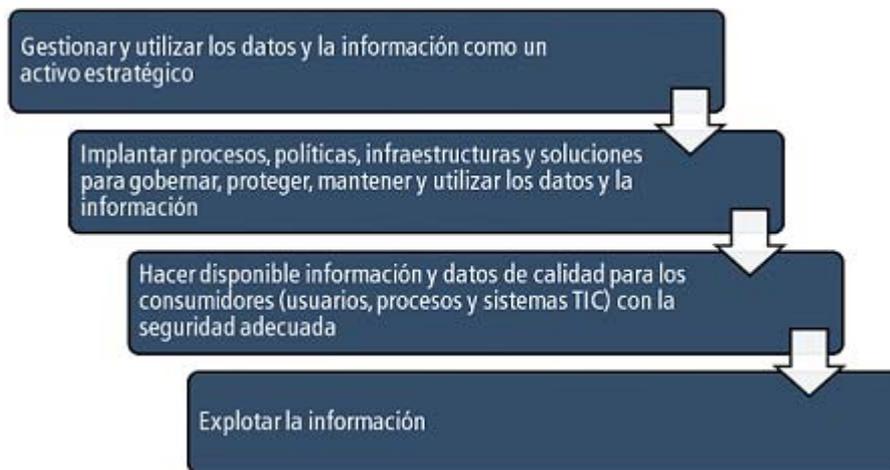


Figura 2. La gestión de los datos y la información. Fuente: Penteo

■ Eliminar posibles riesgos asociados al uso de los datos.

Los proyectos de implantación de programas de *Data Governance* son tan únicos como las compañías que los implantan. Sin embargo, los marcos estructurales que se han utilizado son en realidad bastante similares entre ellos. Existen componentes fundacionales comunes sobre los que construir la iniciativa:

■ Organización, estructura de recursos responsables de desplegar las capacidades de gobierno y administración de las actividades.

■ Políticas, principios y estándares, guías para la gestión de la información, y principios para asegurar los estándares de datos y los procedimientos de gobierno.

■ Procesos y prácticas, que establecen los principios que guían cómo las políticas y procesos son creados, modificados e implantados.

■ Métricas, medidas para monitorizar el rendimiento de la iniciativa de gobierno y acciones para mejorar de manera continúa la calidad de los datos.

■ Arquitectura de los datos, incluyendo estándares corporativos de los datos, diccionarios de metadatos, y además medidas de seguridad y privacidad.

■ Herramientas y tecnología, las tareas deben ser automatizadas con el uso de software siempre que sea posible, mediante herramientas de calidad de datos, *data profiling*,<sup>3</sup> herramientas de gestión de metadatos, *dashboards*<sup>4</sup>, etc.

#### 4. Organización de un equipo de *Data Governance*

Nos encontramos ante una iniciativa que no debe ser contemplada como un proyecto TIC, sino como un proceso continuo de cambio de

la cultura empresarial. Es negocio quien debe liderar la iniciativa, la implantación de *Data Governance* es un importante cambio de mentalidad que debe trascender a todas las áreas de la compañía.

La responsabilidad compartida es el sello distintivo del gobierno de datos ya que requiere de trabajo a través de fronteras organizativas y de sistemas, algunas decisiones son principalmente de negocio con aportaciones y guías del departamento de TIC, mientras que otras son decisiones técnicas con aportaciones y guías por parte de los usuarios a diferentes niveles.

Las distintas unidades del negocio se erigen en las "propietarias" de los datos, mientras que el departamento TIC proporciona la estructura y los procesos necesarios. Estos propietarios de los datos son expertos en determinadas áreas temáticas, se erigen en representantes de los intereses empresariales respecto a los datos y toman la responsabilidad acerca de la calidad y uso de estos.

Si con anterioridad a la implantación de la iniciativa de gobierno de datos, han existido proyectos de *Business Intelligence*, es muy posible que exista algún tipo de equipo de *Data Governance*. Este, si bien tendrá un carácter informal, permitirá mitigar los costes y cambios organizativos que suelen requerir este tipo de iniciativas, y seguramente nos permitirá disponer de personas que puedan ocupar los perfiles que se precisan.

El personal que forme parte del equipo de *Data Governance* debe saber cómo utilizar y analizar la información para facilitar la toma de decisiones disponiendo de una mezcla de habilidades técnicas, analíticas y de negocio:

■ Conocer el negocio, sus procesos, las capacidades analíticas de los sistemas, y la estrategia de la compañía para poder establecer un plan director de gobierno de datos.

■ Conocer la organización y canalizar la cultura en el acceso a la información.

■ Mantenerse al corriente de las nuevas capacidades que la tecnología pueda aportar a la organización.

Uno de los problemas históricos en los proyectos de implantación de iniciativas de *Data Governance* es la ausencia de un adecuado seguimiento; mientras que algunas organizaciones han definido correctamente políticas y procesos de gobiernos, en muchas ocasiones no se ha establecido la estructura organizativa necesaria para hacerlas funcionar adecuadamente.

El marco organizativo del programa de gobierno de los datos debe dar soporte a las necesidades de todos los participantes a lo largo y ancho de la compañía. Con el adecuado soporte ejecutivo, el programa de *Data Governance* se beneficiará de la participación de la empresa en las diferentes funciones necesarias, tanto tácticas como son las de los equipos de coordinación de datos y los propietarios de los datos, como estratégicas.

Los roles específicos incluyen (ver **figura 3**):

■ Director de *Data Governance*, responsable principal de gestionar la iniciativa y asegurar la máxima adopción en la organización. Este perfil da soporte a los patrocinadores ejecutivos y ofrece informes periódicos de rendimiento del proyecto, además de negociar con proveedores externos de datos los acuerdos de niveles de servicio asociados.

■ Comité de *Data Governance*, comité es-

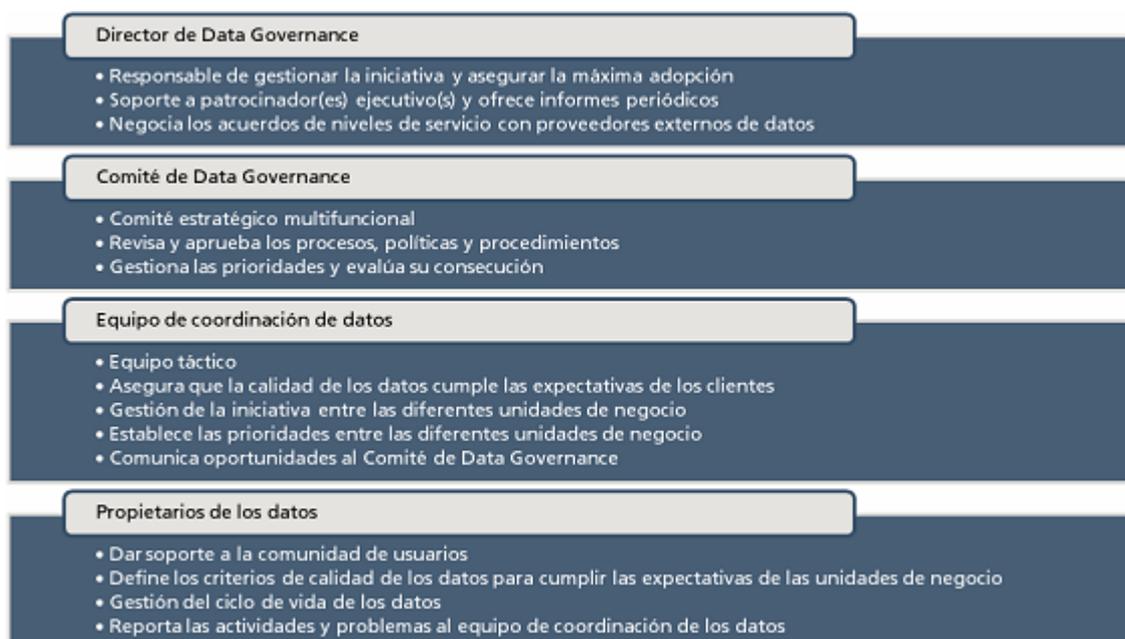


Figura 3. Diagrama organizativo del equipo de *Data Governance*. Fuente: Penteo.

tratégico multifuncional típicamente compuesto por el patrocinador ejecutivo, el director de la oficina de *Data Governance*, y por el CIO (*Chief Information Officer*, el líder o responsable de las Tecnologías de la Información) de la compañía. De manera ideal el patrocinio ejecutivo debería proceder de negocio y no del departamento TIC. Este comité revisa y aprueba los procesos, políticas y procedimientos, gestionando las prioridades y evaluando su adecuada consecución.

- Equipo de coordinación de datos, equipo táctico que asegura que la calidad de datos cumple las expectativas de los clientes y gestiona la iniciativa entre las diferentes unidades de negocio. Es responsabilidad de este equipo detectar y comunicar oportunidades al Comité de *Data Governance*.

- Los propietarios de los datos, que gestionan el ciclo de vida de los datos y dan soporte a la comunidad de usuarios. Estos propietarios definen los criterios de calidad de los datos para cumplir las expectativas de las unidades de negocio, y reportan las actividades y problemas al equipo de coordinación de los datos.

## 5. La necesidad de establecer la propiedad de los datos

Uno de los factores clave de éxito en implantaciones de iniciativas de *Data Governance* es el rol de *data stewardship* o "propiedad de los datos". La propiedad de los datos es la formalización de las responsabilidades que garantizan un control y un uso efectivo de los activos de los datos.

Los propietarios de los datos son usuarios de negocio, expertos en determinadas áreas temáticas designados como responsables para gestionar los datos en nombre de los demás usuarios. Estos representan los intereses de todas las partes involucradas, incluyendo pero no limitándose, a los intereses de sus propias áreas funcionales y departamentos protegiendo, administrando y reaprovechando los recursos de datos.

Estos perfiles deben tener una perspectiva de negocio para garantizar la calidad y el uso eficaz de los datos de la organización. El proceso de gobierno de los datos involucrará a los propietarios como participantes, pero además estos serán directamente responsables del éxito de la gestión de los datos en sus dominios.

En la práctica no existe un modelo "bala de plata" que encaje para todas las organizaciones. Básicamente existen cinco modelos de propiedad de los datos que las organizaciones pueden aplicar, siendo cada uno de estos modelos único, con sus propios pros y contras:

- Modelo 1: propiedad por áreas temáticas. En este modelo cada propietario de los datos gestiona un área temática determinada, así pues el responsable de los datos de los clientes

es diferente del responsable de los datos de productos, etc. En entornos grandes o complejos, puede existir más de un propietario para cada área temática. Este modelo funciona bien en compañías con múltiples departamentos que compartan los mismos datos.

- Modelo 2: propiedad por funciones de negocio. En este caso el propietario de los datos se centra en los datos que un departamento o línea de negocio utiliza, como pueden ser los datos relacionados con marketing, finanzas, ventas, etc. Dependiendo del tamaño de la organización y complejidad en la administración de los datos puede ser que existan otros propietarios de datos por áreas temáticas, resultando un modelo híbrido con el anterior.

- Modelo 3: propiedad por procesos de negocio. Para cada proceso de negocio se asigna un responsable de los datos, en este caso los propietarios de los datos son responsables sobre múltiples dominios de los datos o aplicaciones que participan sobre un determinado proceso de negocio. Nos encontramos ante un modelo muy efectivo para compañías con una orientación y una definición muy clara de sus procesos de negocio, en organizaciones en las que no existe una cultura de procesos o es inmadura este acercamiento no es la mejor elección.

- Modelo 4: propiedad por sistemas TIC. Los responsables de los datos son asignados a las aplicaciones que generan los datos que utilizan. Este modelo es una manera de evangelizar el concepto de propiedad de los datos desde el departamento TIC a las distintas unidades de negocio. Los propietarios de los datos pueden comunicar el progreso de la iniciativa y mostrar como los datos no únicamente van mejorando a lo largo del tiempo, sino que además van afectando a los resultados del negocio.

- Modelo 5: propiedad por proyectos. El asociar el concepto de propiedad de los datos a proyectos es una manera rápida y práctica de introducir la cultura de administración de los datos en la organización. De manera contraria al resto de modelos comentados anteriormente ésta es una medida temporal, que suele utilizarse como punto de partida para el establecimiento de otro modelo formal a largo plazo.

El decidir el modelo de propiedad de datos ideal para nuestra organización no es una tarea trivial en la que debemos plantearnos una serie de factores tales como:

- Los perfiles y habilidades disponibles en la organización para la gestión de los datos.

- La cultura de la compañía.

- La reputación de la calidad de los datos.

- La situación actual respecto a la propiedad de los datos.

- El uso actual de métricas asociadas a la calidad de los datos.

- Las necesidades de reutilización de los datos.

## 6. ¿Cómo abordar el proyecto de Data Governance?

Una implantación adecuada de *Data Governance* puede llegar a tener un impacto directo muy positivo en el rendimiento empresarial. No obstante, es un auténtico reto alcanzar la combinación idónea de personas, procesos y tecnologías para diseñar una iniciativa exitosa.

Para superar este reto debemos construir una estrategia de gobierno de datos efectiva, dirigida por los objetivos de negocio, dotando a los interesados con mejores capacidades para la toma de decisiones y ayudando a la compañía a alcanzar sus objetivos deseados. Una estrategia efectiva debe asegurar que los objetivos de la compañía, la estrategia del negocio, la inversión y los sistemas de *Data Governance* están alineados.

Una iniciativa de *Data Governance* no es nada si no está conducida por los objetivos de la compañía. Los requerimientos del negocio y los objetivos empresariales deben conducir las iteraciones del proyecto. Necesitamos el establecimiento de una estrategia antes de incluir a la tecnología en el proceso.

Incluso antes de empezar a trabajar con la estrategia de gobierno de datos, debemos comprender y documentar los objetivos generales para ayudar a formular la visión y misión del gobierno de los datos para el crecimiento del negocio. Después de documentar la lista inicial de objetivos se debe trabajar con los principales implicados para confirmar la validez de la lista de objetivos y su correcta priorización. Esto asegurará que empezamos a construir nuestra estrategia de *Data Governance* con una base adecuada alineada con el negocio y con los usuarios.

Del análisis del mercado y de las mejores prácticas extraídas de experiencias reales de adopción de *Data Governance*, se extraen las siguientes recomendaciones:

- Involucrar a unidades de negocio para que lideren la iniciativa. Porque *Data Governance* no es únicamente una tecnología sino además un importante cambio de mentalidad que debe trascender a todas las áreas de la compañía. La gestión del cambio y la comunicación desde el inicio del proyecto en este tipo de iniciativas resulta esencial para asegurar el éxito. El proyecto debe ser abordado desde la componente de organización y procesos, aunque tutelado desde cerca por el departamento TIC. La existencia histórica de la función de organización se perfila como un claro facilitador de la adopción de la iniciativa.

- Vender internamente el proceso. Las implantaciones de *Data Governance* suponen un impacto importante en la organización en muchos sentidos, por lo que los CIOs de las compañías deben iniciar sus proyectos de

*Data Governance* cuando han llegado a un consenso en la decisión con otros cargos directivos implicados en el proceso y cuando han conseguido vender internamente el proyecto. De esta forma, la implicación en el proyecto de las distintas áreas de negocio queda plenamente asegurada de antemano y por lo tanto el riesgo al abordar el proceso es mucho más controlado.

■ La adopción de *Data Governance* no se debe abordar como un proyecto finito. El cambio de mentalidad y de cultura, y la reorientación de la compañía a la calidad de la información son los indicadores que identifican el éxito de una iniciativa, por lo que no es usual abordarlo como un proyecto TIC típico.

■ Gestionar un portafolio de proveedores estratégicos. La situación del mercado nos obliga a evaluar, monitorizar y gestionar el ecosistema de nuestras aplicaciones y la hoja de ruta del portafolio de soluciones de los proveedores para estandarizar y reducir el riesgo, la redundancia y los costes. La selección de herramientas tiene menos que ver con las funcionalidades y más con el hecho que las herramientas seleccionadas puedan cumplir los requerimientos específicos de negocio.

■ Planificar y diseñar antes de implantar. Nos encontramos ante una iniciativa de complejidad importante por lo que debemos tomarnos nuestro tiempo para definir exactamente las bases de nuestro futuro sistema orientado a servicios. Debemos esbozar los planos de cómo serán nuestros sistemas de información objetivo y avanzar de forma gradual y progresiva en su consecución.

Finalmente, es importante destacar que una estrategia de *Data Governance* debe diseñarse para ser ágil y adaptativa. Ha de ser tratada como un ente vivo que evoluciona constantemente para alcanzar los objetivos empresariales. La estrategia debe focalizarse en comunicar qué estamos planificando implantar, cómo lo vamos a implantar y cuándo los usuarios verán reflejados sus requerimientos en el sistema. Empezamos con políticas y guías generales y con diagramas de alto nivel, a medida que el ecosistema madura en paralelo lo hará la documentación formal y el nivel de detalles identificados en la estrategia. Ha de ser nuestra intención evolucionar la estrategia de gobierno de datos como parte integrante de la visión de la compañía a medida que realizamos iteraciones y obtenemos más y más detalles al respecto. Planifiquemos para evaluar y reinventar continuamente a medida que las necesidades del negocio cambian, teniendo en cuenta las tendencias tecnológicas actuales y futuras para construir una estrategia de gobierno de datos exitosa.

## 7. Conclusiones

Los activos tangibles de las organizaciones tienen valor y son gestionados mediante sistemas de información y procesos empresaria-

les. Los datos, precisamente por su naturaleza intangible, no son percibidos en muchas ocasiones como activos estratégicos. No obstante, los datos de calidad, precisos y disponibles son un prerrequisito para que las operaciones de cualquier organización sean efectivas.

Las compañías que son capaces de reconocer el valor real de los datos, es decir, que han establecido procesos, políticas y procedimientos de calidad de datos, que son conscientes de cuáles son los datos realmente importantes o útiles para su negocio y, que en definitiva, confían en la calidad de sus datos, se transforman en "organizaciones basadas en los datos". Estas organizaciones se sitúan en una clara situación de ventaja respecto a su competencia gestionando los datos como un activo estratégico más, pero para alcanzar esta meta es necesaria una adecuada visión estratégica para mejorar la calidad de la información.

La implantación de un proyecto de *Data Governance* requiere del soporte de todas las áreas del negocio implicadas. Tomando el control de los datos podemos retener mejor a nuestros clientes, aumentar el éxito de estrategias de marketing, controlar mejor los riesgos y, en definitiva, permitir que la empresa se gestione de manera más eficaz y eficiente.

Una adecuada implantación de *Data Governance* elimina las discrepancias entre los silos de datos. Sin embargo aquellas compañías que han implantado estos proyectos se han dado cuenta en seguida de que los plazos de implantación varían mucho en función del alcance y que no son simples ejercicios tecnológicos.

Cuando se adopta correctamente, *Data Governance* es una disciplina que ayuda a alcanzar el verdadero valor de las aplicaciones analíticas y debe constituirse en los cimientos para todas las iniciativas de gestión de la información. Pero para alcanzar una adecuada gestión de estas entidades es necesaria una adecuada visión estratégica para mejorar la calidad de la información.

Son aquellos proyectos que se enfocan de manera iterativa, empezando con aquel conjunto de necesidades y datos que ofrecen el mayor valor al negocio en el menor tiempo posible los más exitosos. ¿Buscamos una mejor toma de decisiones mediante los sistemas de *Business Intelligence*? Entonces nuestro punto de partida deben ser los datos analíticos. ¿Buscamos conseguir una mayor eficiencia operacional o ganar consistencia en los procesos a lo largo de diferentes sistemas transaccionales? Entonces empezamos por los datos operacionales.

## Bibliografía

- Jill Dyché. *Five Models for Data Stewardship*. Baseline Consulting, 2009.  
David Loshin. *Data Governance for Master Data Management and Beyond*. DataFlux, 2008.  
Óscar Alonso. *Tendencias en el uso de BI en España 2009*. Penteo, 2009.  
Óscar Alonso. *El problema de la gestión de los datos*. Óscar Alonso, Penteo, 2010.

## Notas

<sup>1</sup> *Data Governance* es una disciplina emergente con una definición en proceso de evolución. Esta disciplina abarca una convergencia entre calidad de los datos, gestión de los datos, políticas de datos, gestión de procesos de negocio y gestión de riesgos alrededor del manejo de los datos en una organización (Traducción libre de la introducción al concepto que se encuentra en la Wikipedia en inglés el 24/6/2011: <[http://en.wikipedia.org/wiki/Data\\_governance](http://en.wikipedia.org/wiki/Data_governance)>).

<sup>2</sup> Un *Data Warehouse (DW)* es una base de datos usada para generación de informes. Los datos son cargados desde los sistemas operacionales para su consulta. Pueden pasar a través de un almacén de datos operacional para operaciones adicionales antes de que sean usados en el DW para la generación de informes (Traducción libre de la introducción al concepto que se encuentra en la Wikipedia en inglés el 24/6/2011: <[http://en.wikipedia.org/wiki/Data\\_warehouse](http://en.wikipedia.org/wiki/Data_warehouse)>). Se suele considerar que el término equivalente en castellano es "Almacén de datos": "En el contexto de la informática, un almacén de datos (del inglés data warehouse) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Se trata, sobre todo, de un expediente completo de una organización, más allá de la información transaccional y operacional, almacenado en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos (especialmente OLAP, procesamiento analítico en línea)." (Wikipedia en castellano 24/6/2011: <[http://es.wikipedia.org/wiki/Almac%C3%A9n\\_de\\_datos](http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos)>).

<sup>3</sup> *Data profiling* es el proceso de examinar los datos disponibles en las fuentes existentes (por ej, una base de datos o un fichero) recogiendo estadísticas e información sobre esos datos. (Traducción libre de la introducción al concepto que se encuentra en la Wikipedia en inglés el 24/6/2011: <[http://en.wikipedia.org/wiki/Data\\_profiling](http://en.wikipedia.org/wiki/Data_profiling)>). En castellano se suele considerar "perfilado de datos" como una traducción adecuada: "Por perfilado de datos se entiende el análisis de los datos de los sistemas para entender su contenido, estructura, calidad y dependencias": <<http://integraciony calidad.blogspot.com/2008/07/migraciones-fusiones-y-adquisiciones.html>>.

<sup>4</sup> Aunque la palabra inglesa "*dashboard*" puede ser usada en muchos contextos, en el que nos ocupa diríamos que "En gestión de sistemas de información, un *dashboard* es un sistema de información ejecutivo (similar al tablero de instrumentos de un coche) que se diseña para ser fácil de leer". (Traducción libre de la introducción al concepto que se encuentra en la Wikipedia en inglés el 24/6/2011: <[http://en.wikipedia.org/wiki/Dashboards\\_\(management\\_information\\_systems\)](http://en.wikipedia.org/wiki/Dashboards_(management_information_systems))>). Aunque en castellano abunda una diversidad de traducciones consecuente con la polisemia del término, "tablero de mandos" parecería la más adecuada en este contexto.

Carlos Luis Gómez  
Director de Ibertia Tecnología y Negocios

<cluisg@ibertia.es>

# Business Intelligence y pensamiento sistémico

## 1. Introducción

"¿Por qué fallan tantos proyectos de cambio en las organizaciones? Porque están enfocados en conseguir que las personas hagan las cosas mejor y más rápido y no en cambiar el **"sistema"**. Detrás de esta realidad está la raíz del limitado beneficio, en muchos casos muy lejos del potencial esperado, o incluso del fracaso, de muchas iniciativas de *Performance Management*, *Business Intelligence* (BI), TQM, BSC, CRM...

Afortunadamente, resulta sorprendente como el uso del concepto de "sistema" ha demostrado ser extraordinariamente útil para abordar con sencillez situaciones complejas de relaciones sociales, humanas y de negocio. No es de extrañar que esté anunciada por algunos la superación de la "era de la información" para dar paso a la "era de los sistemas".

Una metodología completa de BI debe adquirir el valor de la conjunción de dos corrientes metodológicas:

- Metodologías de BI enfocadas al negocio.
- Metodologías de Pensamiento Sistémico. Reduciremos al mínimo las cuestiones tecnológicas de BI, aunque éstas también son parte del sistema.

## 2. Metodologías de BI enfocadas al negocio

Las principales causas de fracaso en la implementación de soluciones de BI no se encuentran en cuestiones técnicas o tecnológicas, sino en aspectos organizacionales y de negocio. Esto puede corroborarse consultando numerosos informes de analistas. La necesidad de entendimiento y de creación de equipo entre unidades de negocio y Tecnologías de Información se hace aún más relevante en BI que en otras soluciones tecnológicas. Existen probadas metodologías (Kimball, Inmon, Imhof, híbridas), que bien implementadas resuelven con éxito los aspectos tecnológicos de BI. No ocurre lo mismo para los aspectos no tecnológicos.

Por este motivo en los últimos años han surgido metodologías de BI enfocadas al negocio. Estas metodologías hacen énfasis en la fase de arquitectura y han establecido la necesidad de organizar BI en torno a mapas de oportunidades de negocio, alineados con la estrategia de la organización. Un ejemplo

**Resumen:** Este artículo es una invitación a pensar en BI de una forma distinta, pensar sistémicamente como complemento a la aproximación analítica habitual. Pensar en sistemas nos permite añadir una visión holística, "ver el bosque además de ver los árboles". Business Intelligence (BI) necesita adaptarse al uso de metodologías de Systems Thinking (pensamiento sistémico) si persigue el propósito de construir un "sistema" real de toma de decisiones. El uso de herramientas y metodologías de Pensamiento Sistémico en BI no solo multiplica los beneficios potenciales de una solución de BI, complementando o mejorando la solución, sino que la aproximación metodológica más adecuada para construir soluciones de BI enfocadas a la obtención de valor de negocio debe ser en algunos casos necesariamente sistémica. No estamos hablando solo de futuro sino de presente. Algunos de los beneficios de esta simbiosis están probados. Por ejemplo, la superación del balanced scorecard (BSC) llegando al Dynamic Balanced ScoreCard (DBSC), o la creación de simuladores completos de negocio, sin los límites del "What If" de las herramientas habituales de BI.

**Palabras clave:** BI, dinámica, inteligencia, negocio, simulación, sistémico.

### Autor

**Carlos Luis Gómez** es Director de Ibertia Tecnología y Negocios, empresa especializada en soluciones y servicios de *Business Intelligence* (BI) <<http://www.ibertia.es/>>. Acumula una experiencia de más de 15 años como consultor en BI, *Data Warehousing*, y *Systems-Thinking*, cubriendo diferentes sectores de la industria y muy diversos entornos tecnológicos, y liderando proyectos de ámbito nacional e internacional. Es un experto en metodologías para el desarrollo de oportunidades de negocio sobre BI, y un prestigioso formador y colaborador de distintos medios de comunicación enfocados a BI, liderando el lanzamiento de la revista "Gestión del Rendimiento". Su capacitación técnica está acreditada por distintas certificaciones y programas: "DB2 Info Sphere Warehouse" de IBM, Teradata Professional, Netezza, "Microsoft Performance Point Server", consultor cualificado en España por BARC, analista líder europeo en BI para realizar servicios de selección y análisis de infraestructura tecnológica en BI.

de esta metodología lo encontramos descrito en [1].

Adicionalmente, desde el punto de vista organizacional se establece la necesidad de crear programas de *BI Governance* y estructuras organizacionales como Comités y Centros de Competencia de BI (BICC). Estos elementos han tenido cierta proliferación y han paliado muchos de los problemas que surgen en los proyectos complejos de BI.

Este mayor énfasis en aspectos organizacionales y de negocio ha sido una mejora importante en el desarrollo de iniciativas de BI. Cualquier metodología actual de BI debe incluir características como:

- Organización de BI como un portfolio de servicios enlazados a un mapa de oportunidades de mejoras de negocio.
- Establecimiento de un programa de *BI Governance*, incluyendo cuando sea necesario estructuras organizativas como Centros de Competencias de BI (BICC).
- Dirección de programas iterativos, no gestión de un proyecto.

- Diseño y arquitectura de información, no de aplicaciones.
- Incorporación de metodologías tecnológicas probadas (Kimball, Inmon, Inhoff...).

Pero todos estos componentes siguen siendo "hard systems", salvo en aspectos meramente organizacionales. Al introducir a las personas el sistema se nos convierte en un "soft system". El factor humano puede modificar completamente el funcionamiento de un sistema, por muy bien concebida que esté la metodología desde un punto de vista "hard system". Peter Checkland [2][4], introduce en su metodología de sistemas blandos, (*Soft Systems Methodology*, SSM), acciones de mejoras en tres frentes: organizacionales, procesos y de comportamiento. SSM ha sido utilizado con éxito en diferentes entornos de Sistemas de Información [3].

Algunas metodologías de BI [1] para abordar el "problema" humano incluyen métodos de "reingeniería de procesos" y de "gestión del cambio" para los usuarios, pero mantienen la tendencia generalizada en gestión y recursos

humanos, de usar una aproximación de resolución de problemas ("*problem-solving approach*"), hilo conductor de las metodologías de gestión técnica. "Si resuelvo los problemas de cada una de las partes, obtengo la solución completa" es la creencia errónea que sustenta este tipo de métodos. Esto, casi nunca sucede.

En contraposición, el pensamiento sistémico afirma que es necesario abordar la globalidad del sistema, contemplar la interrelación entre las partes. Por este motivo una de las reglas de SSM es evitar la palabra "problema" porque induce a buscar la "solución". Frente a esto hay que contemplar la realidad como "situaciones problemáticas" sobre las que actuar para obtener *mejoras deseables y viables*. Orientado a la acción, SSM busca "acciones de mejora" sobre el sistema global, no soluciones a problemas parciales.

Esta visión de SSM no es contradictoria, encaja de hecho a la perfección con las mencionadas metodologías de BI enfocadas al negocio. La diferencia estriba en la aproximación sistémica, no analítica. Las ventajas, sobre todo en cuanto al potencial de éxito y alcance de la solución son considerables.

Antes de definir el concepto de sistema para comprender esta diferencia, se muestra a continuación un ejemplo de algunas de las indeseadas consecuencias de la aproximación exclusivamente analítica en la toma de decisiones.

### 3. Mejorar para empeorar

La aproximación analítica, no sistémica, es causa de muchos de los problemas conocidos de BI: silos de información, soluciones departamentales no integradas, poco uso y participación del usuario de negocio, escaso valor estratégico. Realmente no son problemas propios de BI, sino heredados del modelo de organización, gestión y negocio.

Sobre todos ellos hay amplia literatura. Se expone aquí otro ejemplo de error, la sub-optimización, representativo de cómo buenas soluciones parciales generan errores sistémicos en la organización y en el negocio. En estos casos una buena solución

BI lo único que hace es mejorar la enfermedad, es decir empeorar los malos hábitos decisionales.

A diferencia de otros males, la organización víctima no es consciente del problema de sub-optimización, o no desea cambiarlo, incluso la fomenta. La sub-optimización es una de las causas que erosionan la competitividad de las empresas y atrofian la posibilidad de desarrollo de las mismas. Jamshid Gharajedaghi [5] cita las siguientes causas principales de erosión de la competitividad, de menor a mayor impacto: imitación, inercia, sub-optimización, cambio del juego (en el mercado por ejemplo), desplazamiento de paradigma.

El problema de la sub-optimización puede nacer de la diferencia de poder real entre unidades de negocio y de las personas en las empresas. Por ejemplo, una persona en una organización se convierte en un héroe al resolver un problema crítico de negocio. Esta persona recibe veneración y autoridad que acaban otorgándole mayor poder, convirtiéndole en un decisor clave. Cada vez más decisiones pasan por sus manos y la tendencia natural es fortalecer su posición. Su departamento se hace más poderoso y recibe más recursos. Los buenos vendedores venden a las personas con poder, no a las más necesitadas de solución. Todo contribuye a optimizar el departamento más optimizado de la empresa. El departamento se hipertrofia.

El círculo vicioso de la sub-optimización se agrava porque, en ambientes de competencia interna de recursos y poder, otros departamentos se atrofian siendo ignorados mientras los primeros se desarrollan. Las unidades de negocio atrofiadas ponen el límite al desarrollo del negocio, pero difícilmente recibirán los recursos que necesitan.

En cuanto a *Business Intelligence*, el departamento hipertrofiado es el máximo candidato para liderar una solución de BI y en consecuencia hacerse con el poder de la información y el conocimiento. BI es un contribuyente neto a la sub-optimización y a otros errores sistémicos de las organizaciones, salvo que las soluciones de BI se aborden con una visión

holística, lo que quiere decir entre otras cosas, con una visión corporativa y estratégica.

La sub-optimización para la organización es en cierto modo el equivalente a las burbujas del mercado, que hipertrofian temporalmente algunos sectores generando círculos viciosos hasta el colapso.

Es imposible romper un círculo vicioso de este tipo sin tomar conciencia del mismo. Desde el análisis, desde los datos, es imposible esta toma de conciencia. La única posibilidad de abordar estas situaciones es desde perspectivas sistémicas.

La sub-optimización es solamente un ejemplo, existen muchos más errores sistémicos, círculos viciosos y virtuosos transparentes, inaccesibles, a un aproximación analítica.

En consecuencia, tal como indicamos en la introducción, mejorar la infraestructura, la información, el conocimiento, las personas, pocas veces contribuye a una mejora global del sistema, e incluso tras una aparente mejora lo empeora en el largo plazo. Para mejorar el sistema hay cambiar el sistema.

### 4. El concepto de "sistema"

Nuestra cultura se basa en casi su totalidad en el pensamiento analítico. El éxito del método científico, basado en la aproximación analítica, ha contribuido a esta situación. La aproximación analítica hacia los problemas, también se denomina reduccionista, puesto que divide los problemas en las partes que los componen y establece que resolviendo cada una de las partes a las que ha sido reducido se resuelve el problema completo.

En contraposición al método tradicional surge la aproximación holística. Ésta utiliza el concepto de sistema, enunciado por primera vez por Ludwig von Bertalanffy en 1950, con la siguiente definición: "*un sistema es una entidad que mantiene su existencia a través de la mutua interacción de sus partes*". Una persona, un coche, una red social, la tierra, por supuesto una organización, son ejemplos de sistemas. En cambio, un montón de arena no es un sistema, no hay interacciones entre sus partes.

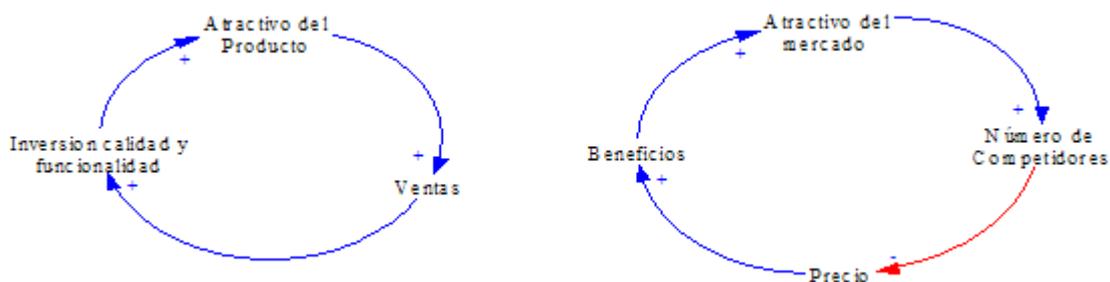


Figura 1. Comportamientos dinámicos en los sistemas.

Aproximación Analítica - Fases	Aproximación Sistémica - Fases
1 Toma una parte del todo	¿De qué forma parte? Identificar el todo en el cual está contenido
2 Tratar de entender que hace esa parte	Explicar el comportamiento del todo donde está contenido
3 El conocimiento del todo es el ensamblaje del conocimiento de todas las partes	Identificar el rol o función en el todo de lo que estoy tratando de explicar o estudiar

**Tabla 1.** Comparación entre las aproximaciones analítica y sistémica según Ackoff.

El lenguaje sistémico hace énfasis en las interacciones entre las partes, y no en las partes. En la **figura 1** usamos un ejemplo de diagramas causales de "dinámica de sistemas".

El gráfico nos ilustra la sencillez del lenguaje sistémico, una de sus ventajas. Los dos bucles mostrados prácticamente son auto-explicativos y nos muestran de forma cualitativa, el comportamiento dinámico de elementos del sistema. ¿Cuántas palabras y tiempo nos llevaría explicar lo mismo que expresan estos diagramas con lenguaje analítico? Mejor aún, pasar de modelos cualitativos a modelos cuantitativos que permiten la simulación es posible y existen herramientas que hacen fácil la tarea.

Se enuncian a continuación algunas características importantes de los sistemas.

**4.1. Propiedades emergentes**

Son propiedades exclusivas que emergen en el sistema global resultado de la interacción de las partes, que no se pueden encontrar en las partes. Por ejemplo, la interacción de cuerdas vocales, aire de los pulmones, posición de la lengua, órdenes del cerebro, un código de lenguaje producen la voz. La voz es una propiedad resultado de la interacción. No podemos llegar a la voz, por ejemplo su tono, analizando cada una de las partes que interaccionan para producirla.

**4.2. Los sistemas son abiertos**

Los sistemas se relacionan con otros sistemas y con su entorno, forman parte de sistemas más amplios que los contienen. Las interacciones más importantes del sistema son las que tiene con su entorno. Tampoco podemos entender el sistema sin este tipo de interacciones. Por ejemplo, en el sistema Tierra existen cuatro estaciones (primavera, verano, otoño, invierno). Podemos obtener el conocimiento de cómo suceden las estacio-

nes y cuáles son sus características analizando los datos registrados año a año. Pero solo estudiando la interacción de la Tierra con el sistema que la contiene, Sistema Solar, *entendemos por qué* se producen las estaciones, fruto de la interacción entre sol y tierra.

Afirma Russell Ackoff [6] que la aproximación analítica produce conocimiento, encuentra respuesta al "¿cómo?", mientras la sistémica produce entendimiento, responde al "¿por qué?". Son complementarias, por eso hoy día el pensamiento sistémico se extiende a todas las disciplinas científicas. Ackoff establece la comparación entre la aproximación analítica y sistémica que se muestra en la **tabla 1**.

El método empírico sigue la aproximación analítica, reproduce experimentos en laboratorios bajo determinadas condiciones. Los resultados obtenidos en el entorno aislado del laboratorio se extrapolan a la realidad completa.

Pero con la realidad no podemos experimentar. En un entorno de negocio no podemos hacer experimentos de prueba y error con distintas decisiones, ni podemos aislar una parte del entorno, del mercado, de los procesos para que funcionen aisladamente. El método analítico resulta inadecuado para abordar la mayoría de las problemáticas de negocio. Ante esto la aproximación sistémica nos ofrece la simulación como herramienta de estudio y prueba sin intervenir en la realidad, y la visión holística, no aislada, del sistema dentro de su entorno global, como la única válida en entornos sociales, económicos, de negocio...

**4.3. Dinámica compleja y retroalimentación (feedback)**

Si volvemos a la **figura 1** podemos extraer más conclusiones. En los dos bucles causales se produce un mecanismo de retroalimentación (*feedback*), importantísima característica de

los sistemas. En el primero esta retroalimentación es positiva produciendo un círculo virtuoso, si suben las ventas, o vicioso si éstas bajan. No hay un límite al crecimiento aparente. Por el contrario el segundo bucle muestra una retroalimentación negativa a través de la relación entre número de competidores y precio. El segundo bucle mostrará un límite en su crecimiento, tiende a la estabilidad.

En todos los procesos de negocio se producen mecanismos de *feedback* que afectan a la dinámica del sistema. El *feedback* no suele ser inmediato, sino que se producen retardos ("delays") en estos mecanismos que no solo complican la dinámica, sino que impiden su detección incluso con sofisticadas herramientas analíticas de BI. Esto provoca efectos de nuestras decisiones a largo plazo contrarias o distintas a las que esperadas en el corto plazo.

Cambios constantes, "*feedback*", "*delays*", no linealidades, efectos contra-intuitivos, acoplamientos... John Sterman en [7] describe qué hace compleja la dinámica de los sistemas y de los negocios. Esto, unido a las barreras de aprendizaje de nuestro cerebro, constituye la causa fundamental de que los expertos de negocio se equivoquen en sus análisis y tomen decisiones erróneas, incluso disponiendo de información de calidad proveniente de sistemas de BI [8].

La forma de superar estas barreras es la modelización y simulación del negocio. No hay otra vía. Las herramientas de BI no pueden abordar esta problemática, están concebidas para el análisis, no para la síntesis.

**5. Pensamiento sistémico en BI**

Existen distintas metodologías de pensamiento sistémico: Dinámica de Sistemas, Cibernética, Teoría del Caos, Planificación Interactiva, *Soft System Methodology*, *System Of Systems Methodology*... Michael C. Jackson [9] hace una detallada descripción y comparación de estas metodologías.

Describimos aquí la aplicación de dos de estas metodologías a BI. Por un lado como herramienta de modelización del negocio y simulación, "Dinámica de Sistemas" [10] es un complemento necesario probado que aporta mejoras sustanciales a los beneficios esperados típicos de una solución BI. Dos ejemplos probados de este maridaje son:

- Superación de las limitaciones de los cuadros de mando y un incremento sustancial de su valor de negocio y aplicación práctica. De hecho hoy día no deberíamos limitarnos a desarrollar *Balanced Scorecards* (BSC). Cualquier Organización que decida utilizar BSC debería directamente diseñar *Dynamic Balanced Scorecards* (DBSC) que incorporen pensamiento sistémico al BSC. Con muchas ventajas, más sencillo que el BSC tradicional, permite identificar, cuantificar y

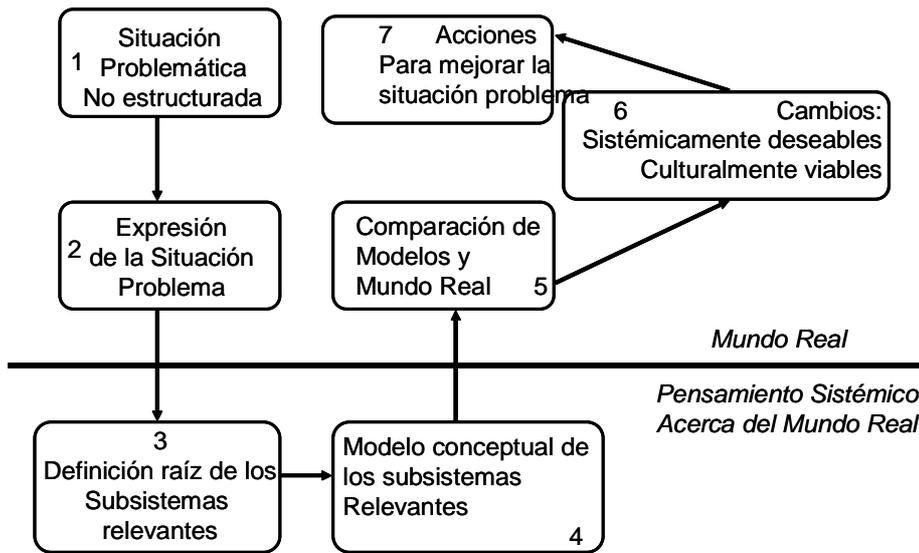


Figura 2. Diagrama de pasos de la Soft Systems Methodology (SSM).

ponderar métricas y KPIs (*Key Performance Indicators*), y estudiar el comportamiento dinámico del cuadro de mando.

■ Incorporación de modelos dinámicos de negocio, imprescindibles entre otras cosas para una correcta simulación. La predicción analítica, tanto a través de la Minería de Datos, o con el uso de las típicas simulaciones "What If" de las herramientas de BI, son claramente insuficientes para prever, influir o controlar el futuro del negocio o de la organización. Esto es así por razones de diversa índole que no es posible detallar aquí, tanto técnicas como metodológicas como muestra Mahesh Raisinghani [11].

Los dos ejemplos anteriores llegan a ser críticos en algunas organizaciones y tienen un futuro asegurado. La simulación está contemplada por Gartner y otros analistas como una tecnología clave y de mayor previsión de crecimiento. De la misma forma que antes de lanzar la fabricación un avión se usan los simuladores de vuelo para garantizar un correcto diseño y para el aprendizaje sin riesgo por parte de los pilotos, un simulador de negocio nos permite aprender sin riesgo y valorando el impacto de todo tipo de decisiones.

Además de la complementariedad técnica que proporciona la Dinámica de Sistemas, la otra vía para mejorar los programas de BI la encontramos en la mejora metodológica que proporciona el pensamiento sistémico. En este sentido enunciamos una propiedad importante adicional de los sistemas descrita por Peter Checkland en su SSM. **Los sistemas actúan con propósitos ("purposefully")** [2][4]. Las personas, por ejemplo, actúan con intención, con propósitos. Los sistemas complejos, como los seres humanos, son multipropósito.

SSM establece que las situaciones problemáticas contienen distintos *worldviews* (visiones) de los actores del sistema. Como se mencionó, SSM es una metodología orientada a la acción. La interacción entre las *worldviews* define las transformaciones deseables y viables en esta metodología. SSM acaba definiendo un conjunto de acciones para mejorar representadas en un "modelo de actividades con propósitos" con métricas de eficacia, eficiencia y efectividad. Modelo que sirve no solo para construir, sino como dispositivo de comunicación entre los actores del programa. La figura 2 muestra los pasos de la metodología SSM. El paso 4 define los modelos de actividades.

El enlace de la metodología de BI enfocada al negocio con SSM puede establecerse fácilmente. El mapa de oportunidades de negocio representa los *worldviews* y propósitos de SSM, el modelo de actividades con propósito representa el programa iterativo de BI, y la gobernanza se establece a través de las métricas de eficacia, eficiencia, efectividad.

### 6. Conclusión

El pensamiento sistémico se está extendiendo a multitud de disciplinas, mostrando una capacidad para afrontar situaciones complejas. BI debe beneficiarse de esta aproximación complementaria a la analítica.

Siguiendo los pasos de la metodología sistémica contemplamos la identificación del sistema superior en la que está contenida la parte en estudio. *Business Intelligence* es un elemento del sistema de decisiones. Un sistema de decisiones consta de dos subsistemas fundamentales, el de información, generada a través de los datos obtenidos de la realidad, y

el de los modelos mentales que generan la estrategia y reglas de negocio [9] [10]. BI adquiere el rol del subsistema de información en la toma de decisiones, mientras que pensamiento sistémico representa el subsistema de los modelos mentales o de negocio. La conjunción de la información de calidad y la modelización sistémica nos proporciona la capacidad de simulación cualitativa y cuantitativa del negocio. Este es el primer beneficio del uso de *Systems Thinking* en BI.

Como segundo beneficio BI puede beneficiarse de las mejoras metodológicas aportadas por el pensamiento sistémico para abordar situaciones o programas complejos, permitiendo priorizar oportunidades de mejora de negocio con sus métricas y KPIs. BI debe evolucionar de la misma forma que en otras ingenierías han evolucionado con éxito de las metodologías duras ("hard") a las metodologías suaves o blandas ("soft"). El beneficio esperado es un mayor potencial de éxito, al contemplar las acciones necesarias: tecnológicas, organizacionales, de proceso y de comportamiento.

Finalmente, destacar que el mayor beneficio del *pensamiento sistémico* es el de la creatividad y la innovación en cualquier área de actividad en la que se aplique. Creatividad con propósito es un requerimiento obligado en el entorno de gestión actual.

**Referencias**

[1] Steve Williams, Nancy Williams. *The profit Impact of Business Intelligence*. Elsevier, 2007.

[2] Peter Checkland. *Systems Thinking, Systems Practice*. Wiley, 1981 [rev 1999 1st ed.]. ISBN-10: 0471986062.

[3] Peter Checkland, Sue Holwell. *Information, Systems and Information Systems*. Wiley, 1990.

[4] Peter Checkland, John Poulter. *Learning For Action: A Short Definitive Account of Soft Systems Methodology, and its use Practitioners, Teachers and Students*. Wiley, 2006. ISBN-10: 0470025549.

[5] Jamshid Gharajedaghi. *Systems Thinking: Managing Chaos and Complexity. A Platform for Designing Business Architecture*. Elsevier, 2006.

[6] Russel L. Ackoff. *Re-Creating The Corporation*. Oxford University Press, 1999. ISBN-10: 9780195123876.

[7] John Sterman. *System Dynamics Modelling, Tools For Learning in a Complex World*. *California management review* 43 (4): pp. 8–25, 2001.

[8] Carlos Luis. *Los Hombres Que No Amaban a las Decisiones*. <<http://www.beyenetwork.es/channels/1576/view/11703>>, octubre 2009.

[9] Michael C. Jackson. *Systems Thinking, Creative Holism for Manager*. Wiley, 2003. ISBN-10: 0470845228.

[10] John D. Sterman. *Business Dynamics*. McGraw-Hill, 2000. ISBN-10: 007238915X.

[11] Mahesh Raisinghani. *Business Intelligence in the Digital Economy: Opportunities, Limitations, and Risks*. IGI Global, 2004. ISBN-10: 1591402069.

Diego Arenas Contreras  
 Jefe Soluciones de Business Intelligence en  
 Formulisa, Chile

<darenasc@gmail.com>

# Caso de estudio: Estrategia BI en una ONG

## 1. Introducción y definiciones

Este artículo describe cómo implementar una estrategia de *Business Intelligence* (BI) en una Organización No Gubernamental (ONG). Una estrategia se entiende como un plan de acción que permite obtener una ventaja competitiva respecto a un estado anterior de la misma organización o en comparación con una organización similar. "*Business Intelligence*" se define como el conjunto de herramientas, procesos, técnicas y algoritmos que permiten soportar el proceso de toma de decisiones en las organizaciones entregando la información correcta a quien corresponde y en el momento oportuno.

La ONG donde se aplica la teoría presentada en este artículo es VE-Global<sup>1</sup> (VE), una organización que recluta, entrena y organiza voluntarios para trabajar con niños en riesgo social en Chile. El trabajo realizado con VE es la esencia de este artículo e intenta difundir y compartir el conocimiento de esta experiencia. VE no contaba con una persona o equipo dedicado a temas de tecnologías de información, si bien estaban las ganas de utilizar mejor la información que tenían, había una oportunidad de mejora al contar con habilidades técnicas en tecnologías de información, que fue lo que resultó después de empezar el proyecto.

Una ONG tal como cualquier organización interactúa con personas y otras organizaciones generando relaciones y vínculos mutuos. Estas interacciones obligan a las organizaciones a tomar mejores decisiones pero sólo algunas lo hacen basadas en sus datos.

Los datos que las mismas organizaciones generan son un activo importante porque ninguna otra organización tiene acceso a ellos. Por esto, se deben identificar estos datos y basar la estrategia sobre ellos. Una adecuada estrategia de información para una ONG no sólo tiene importancia para la propia institución, tiene también un impacto social.

El primer paso es conocer las necesidades de información y los objetivos estratégicos de la organización. Para esto es necesario reunirse con los directores y generar acuerdos que finalmente serán la guía para definir la estrategia de información.

Una estrategia de información debe plantearse en función de las capacidades actuales de la organización. Asimismo, es necesario uni-

**Resumen:** En todas las organizaciones y en todos los niveles se toman decisiones. El soporte a la toma de decisiones permite mejorar y diferenciarse, facilita la gestión, mejora la eficiencia, entrega un apoyo y una guía para la continuidad de la organización. Las instituciones sin ánimo de lucro también deben tomar decisiones pero sus datos y su día a día son diferentes, sus problemáticas diversas, pero también pueden tomar decisiones basadas en sus datos y generar conocimiento a partir de su información. Este artículo muestra cómo planificar y aplicar una estrategia de Business Intelligence (BI) en organizaciones sin ánimo de lucro, partiendo desde el entendimiento de los procesos, identificar las necesidades de información, los datos disponibles y relevantes y entregar las soluciones precisas a los requerimientos de datos e información que la organización tiene. La presentación de las ideas está acompañado de un trabajo práctico con una ONG real en Chile.

**Palabras clave:** análisis de datos, CRM, factores clave de éxito, indicadores de gestión, ONG, reportes.

### Autor

**Diego Arenas Contreras** estudió Ingeniería Civil en Computación en la Universidad de Talca en Chile. Realizó un Diplomado en Gestión de Negocios con *Business Intelligence* y le apasiona el mundo de los sistemas de información y la visualización de información. Ha trabajado en diferentes empresas e industrias en temas y proyectos de Inteligencia de Negocios, *Performance Management* y Minería de Datos. Es actualmente Jefe de Soluciones de *Business Intelligence* en Formulisa, empresa consultora chilena dedicada a entender el comportamiento de los consumidores y a la Inteligencia de Clientes para sus clientes. <<http://www.formulisa.cl>>.

ficar la semántica con la que se trabajará y contar con una definición clara para los conceptos y términos de lenguaje utilizados por los participantes del proyecto, tener un entendimiento de cada concepto simplificará los problemas de comunicación y dará mayor agilidad al proyecto.

En VE se utilizó un documento compartido con acceso a los participantes para escribir las definiciones que conocían y agregar los conceptos que requerían definición, cada participante podía agregar su entendimiento en una nueva columna hasta llegar a un consenso en la definición. Con esto se aseguró la participación y conocimiento de cada integrante del proyecto y se facilitó la inducción para quienes se incorporaron al proyecto después de haber comenzado; es recomendable también el uso de wikis.

Gracias a los acuerdos semánticos podemos definir los términos utilizados en este artículo:

- **Leads:** Personas o instituciones que tienen un potencial de relacionamiento mutuo con VE.
- **Contacto:** Persona que tiene una relación mutua con la VE, puede ser de diferentes tipos.
- **Organizaciones:** Empresas u organizaciones que tienen relación de algún tipo con VE, permiten agrupar los contactos dentro de las organizaciones.

- **Instituciones:** Son los hogares con los cuales VE trabaja directamente a través de sus voluntarios y programas.

- **Voluntarios:** Persona natural que es capacitada y administrada por VE y presta servicios en instituciones y a VE.

- **Donaciones:** Entregas gratuitas de dinero, productos o servicios que recibe VE de parte de un contacto y/o organización.

- **Programa:** Plan de instrucción en un tema específico para apoyar el desarrollo de los beneficiarios de una Institución, por ejemplo: Programa de deportes para fomentar la vida sana y el deporte, programa de lectura para fomentar y potenciar las capacidades lectoras.

Así como en VE, habrá términos específicos en cada organización y será necesario definirlos para fomentar una comunicación efectiva.

Las herramientas utilizadas durante el proyecto fueron de tipo colaborativo y *open source*, para facilitar la comunicación:

- Google Docs, para trabajar documentos colaborativamente.
- Dropbox<sup>2</sup> para el intercambio de archivos.
- GanttProject<sup>3</sup>, herramienta *open source* para llevar la carta Gantt del proyecto.
- FreeMind<sup>4</sup> para diagramar mapas mentales y compartir ideas.

■ Sandbox de Salesforce, ambiente de pruebas proporcionado por Salesforce donde los cambios no afectan a los datos de producción.

Como sucede en la mayoría de las organizaciones, VE estaba manteniendo diferentes repositorios de información por lo que el primer objetivo fue llevar sus datos a un único sistema operacional que permitiera mantener los datos y ser usado como única fuente de información, por lo que se acordó potenciar y sacar el máximo provecho de su solución CRM (*Customer Relationship Management*) para organizaciones sin ánimos de lucro.

En primer lugar, la estrategia BI debe estar alineada con los objetivos estratégicos, segundo se debe trabajar en colaboración con los responsables de la información y comunicar los objetivos a todos los involucrados, en tercer lugar el análisis de los datos, procesos y flujos de información se hace desde la perspectiva de los objetivos estratégicos de cada área de la organización. Lo siguiente es asegurar la calidad de la información a utilizar asegurando la calidad de los datos, y luego aplicar la estrategia en función de las capacidades de la organización y anticipar los futuros requisitos de información basados en los datos almacenados. Finalmente monitorear el uso y evaluar mejoras.

## 2. Datos, información y procesos

Al inicio del proyecto con VE definimos reuniones con los directores para conocer los objetivos estratégicos y las necesidades de cada área de la organización. Al principio fueron reuniones tipo entrevista para obtener su visión y entender los datos que manejan, luego se realizaron sesiones de *brainstorming* (lluvia o tormenta de ideas) donde se les pidió imaginar el proyecto finalizado con éxito y que imaginaran la información de la que dispondrían para analizar los reportes que podrían consultar y las decisiones que serían capaces de tomar basadas en los datos, permitiendo identificar las entidades de datos presentes en VE y sus interacciones.

Se obtuvieron dos artefactos desde esta actividad que es el diagrama de alto nivel de entidades y sus relaciones y un documento con los reportes requeridos. Después de las reuniones y sesiones de *brainstorming*, se fijaron reuniones semanales de 30 minutos para comunicar los avances y fijar los objetivos semanales. En estas reuniones participaron los voluntarios involucrados y el director ejecutivo, que tiene el rol de sponsor del proyecto, y se fijaron reuniones adicionales para coordinar temas con más profundidad durante la semana si era necesario.

Durante la detección de las necesidades uno se interioriza en la cultura organizacional, con las entrevistas a los directivos se conocen los datos únicos de la organización, los prin-

cipales procesos y los actores responsables en los flujos de información y de datos, se palpa la realidad de datos que la organización vive. El rol de los expertos en BI es el de guiar y mostrar los beneficios de una solución de BI y recomendar la solución que se ajusta a las necesidades de información, definiendo una estrategia para alcanzar estos fines.

Una vez analizados los procesos y recolectada la información necesaria para planificar una estrategia de BI debemos dar el paso más importante de toda nuestra estrategia de BI, que no es otro que generar el entendimiento entre todos los implicados. Para ello debemos consensuar que todos entendemos lo mismo, y para ello generamos el primer entregable que será el pilar de todos los proyectos.

Este primer entregable es el que llamamos definiciones organizacionales. Consiste en unificar el vocabulario y que cada término tenga una definición semántica única y homogénea. La idea es que todo el mundo entienda lo mismo al escuchar términos como "voluntario" o "institución". Con esto se asegura el entendimiento entre los interesados y se define un vocabulario que facilitará la comunicación y el avance en la implementación exitosa. Si todos hablamos el mismo idioma seguro que no tendremos malos entendidos.

En VE registramos en un documento compartido los reportes que se necesitaban indicando dueño del reporte, nombre del reporte, resumen, razón para el reporte, periodicidad de consulta, campos requeridos en el reporte y observaciones que se quisieran agregar.

Se analizó este documento con alrededor de 45 definiciones de reportes llevándolo a un mapa mental donde el primer nivel son las entidades de datos como *Leads*, Contactos, Organizaciones, Donaciones, Eventos, Campañas y Programas y dentro de cada hoja, que representa una entidad, se agrupan los campos que se requieren para satisfacer el reporte agrupando por tipos y temática. Por ejemplo, para los *Leads* se requiere información básica de contacto donde se agrupan los campos de contacto y campos de metadatos para saber cómo llegó a conocer VE, tipo de *lead* (*lead* para voluntariado, socio potencial de VE, sponsor de VE, etc. (ver **figura 1**).

El mapa mental<sup>5</sup> con las entidades identificadas y los campos requeridos para satisfacer todos los reportes es la base sobre la cual se estructura el plan de calidad de datos y se asegura que los campos existan en la base de datos y sean correctamente llenados. Además, dentro del mapa mental se agrupan los reportes solicitados por temas mostrando así las grandes áreas de interés para VE las cuales son Contactos, una serie de informes para conocer las personas y organizaciones que tienen contacto con VE, Donaciones que

principalmente consiste en tener información oportuna acerca de las donaciones recibidas, dónde están, cómo se utilizan, quiénes las han hecho, etc., y el Rendimiento que es un tema atinente a las empresas para optimizar recursos, focalizar los esfuerzos, reducir costos y los beneficios de tomar decisiones basadas en los datos.

Hay que tener presente y recordar en todo momento que una estrategia de BI no es un proyecto puntual que entregará un producto final específico como un reporte o un *dashboard*, es un plan que aborda las necesidades de información y se encarga de generar las acciones necesarias para la satisfacción de éstas en función de las capacidades actuales de la organización.

Cada organización produce datos que son propios a la actividad y al día a día de la institución, estos datos que ninguna otra organización puede acceder se deben reconocer como un activo para la organización, y uno de importancia ya que son una fuente única de información y conocimiento. Estos datos únicos son inherentes a la actividad de la organización, se pueden levantar mediante entrevistas, mirando los reportes existentes y la calidad de datos del modelo de datos. También es posible identificar potenciales datos que no se están guardando que son únicos a la organización, identificar esto es clave para asegurar el registro de estos datos en la estrategia de BI.

En muchas ocasiones las ONG no cuentan con un experto en sistemas de información por lo que los procesos y datos han sido llevados de la mejor manera de acuerdo a cada persona encargada de la labor. Esto sumado a la rotación de voluntarios hace que la disponibilidad e integración de la información sea mínima. Un tema recurrente es la subutilización de los sistemas a los cuales tienen acceso debido al bajo nivel de configuración y adaptación a sus necesidades específicas.

En ONGs la comunicación con su red de contactos representa la continuidad, comunicar datos fácticos sobre el trabajo que se está realizando y medir estos contactos es valioso. Tener la información disponible para organizar eventos en lugares con alta densidad de contactos de la organización por ejemplo, poder entregar estadísticas actuales de las horas de voluntariado, cantidad de gente a la que se está llegando o dependiendo del tipo de organización poder presentar el trabajo en números actualizados y disponibles para la comunidad.

Desde el principio de la implementación, se debe enfocar en la información relevante de todos los datos y procesos que se levanten, se debe hacer un ranking ordenado por impor-

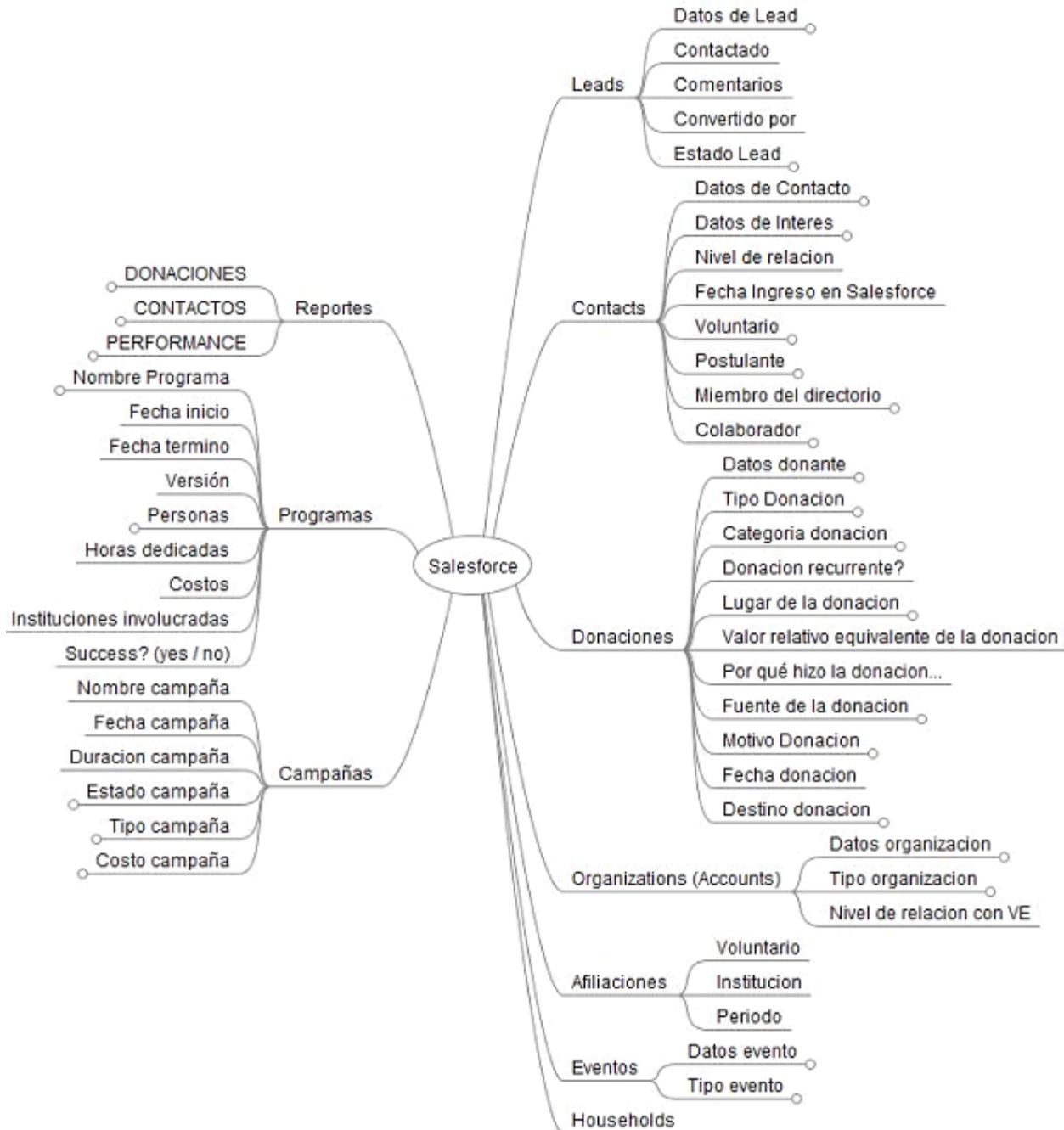


Figura 1. Primer nivel del mapa mental con la identificación de los campos mínimos y las temáticas de reportes.

tancia, y focalizar los esfuerzos en satisfacer las necesidades de información de acuerdo al ordenamiento.

**3. Planificación e implementación de una estrategia**

Contar con el apoyo directivo y de los actores principales en cada una de las etapas de la implementación es un factor clave de éxito del proyecto. Enfocarse en la información relevante para la organización y asegurarse de cubrir los ámbitos más importantes de acuerdo al ranking de importancia de la información requerida en el levantamiento. Asimismo, se debe pensar en cubrir la mayor cantidad

de necesidades de información y los procesos más importantes o de mayor impacto para la organización.

Una estrategia de *Business Intelligence* es independiente de las herramientas usadas y de los usuarios que participan en la iniciativa, se trata de entregar a todo nivel la información que la organización necesita. La información es como el oxígeno que mantiene a la organización respirando.

La arquitectura es dependiente a las necesidades de información y se define en torno al resultado del levantamiento. Es importante

que en organizaciones no gubernamentales se utilice un estilo de desarrollo flexible para alinearse a las necesidades cambiantes de información, es decir asegurar el alineamiento con los objetivos estratégicos de la organización pero también evaluar los nuevos requerimientos con prontitud. Por ejemplo, se evaluaron e implantaron funcionalidades nuevas del CRM de módulos que fueron lanzados después del inicio del proyecto, por lo que no fueron considerados en la planificación inicial, pero ofrecían mejoras que se alineaban con los objetivos del proyecto.

Los dueños de la información y de los datos

deben estar claramente definidos, y perfectamente podrían ser actores diferentes. Evaluar la calidad de datos actual y establecer un plan de mejoramiento continuo de la calidad de datos y de información; además se debe definir la seguridad y los accesos a datos sensibles. La planificación de la estrategia debe considerar una batería de reportes e indicadores de gestión y la medición de los procesos importantes y personas al interior de la organización.

En el caso de VE se establecieron las reuniones y entrevistas para conocer los procesos, los datos relevantes fueron obtenidos desde las reuniones y los informes registrados en el documento compartido. El mapa mental con los campos de información más la clasificación de los reportes sentaron la cobertura de datos mínima que se necesita en el CRM y cada entidad se evaluó en profundidad identificando claramente su campos y relaciones con otras entidades.

Luego se agregaron y validaron los campos en el sistema. Para una aceptable calidad de datos se parametrizó la mayor cantidad de campos posibles, tratando de minimizar los "datos abiertos" como por ejemplo listados de países, meses de estadía, preguntas con alternativas, etc. A partir de ahí se analizaron los cerca de 45 reportes solicitados, para reducir la cantidad de reportes y maximizar la información en cada uno de ellos.

Los reportes creados se entregaron a sus dueños, los cuales pueden conocer quiénes son los dueños de cada dato, resultando así que cuando un dato no está bien registrado se puede gestionar directamente con el responsable que se ingrese correctamente.

Durante el proceso se identificaron oportunidades de mejora y de optimización de procesos en VE, por ejemplo aprovechando la creación de formularios web que recolectan automáticamente datos y los ingresan al sistema sin intervención humana, esto reduce el tiempo de ingreso de datos y evita duplicar fuentes de información. Disponer de datos actualizados para consulta y para una mejor toma de decisiones permite a los directores de VE conocer mejor la organización a través de sus datos en forma oportuna y eficiente; también mejorar la comunicación con su red de contactos al segmentar y focalizar sus comunicaciones.

Lo recomendable es unificar las fuentes de datos, tener un solo repositorio y desde ahí concentrar los requerimientos de información, como fue el caso de VE que apuntó a llevar sus procesos y recolección de datos a su sistema CRM. Identificar los procesos permite el mejoramiento continuo de los flujos de información mediante la automatización de las tareas repetibles identificadas y la

optimización de procesos que faciliten el acceso a la información y faciliten el ingreso de los datos.

Al momento de la implementación hay que asegurarse de que los usuarios que interactúan con los sistemas operacionales o transaccionales donde se ingresan los datos requeridos, tengan conocimiento del motivo por qué los datos son importantes y que la calidad de datos y la calidad de la información depende de ellos. Es importante hacer capacitaciones en el uso del sistema y en las capacidades de información enfatizando que los usuarios deben usar correctamente el sistema. A este respecto, su involucramiento temprano en la implementación facilita su conocimiento del sistema.

Finalmente, los beneficios son entregar la información adecuada a los usuarios que la necesitan, identificar los datos que son necesarios para satisfacer todos los reportes, generar una trazabilidad en los datos en el proceso desde las fuentes de datos hasta los reportes finales, y hacer una reducción al nivel de área interesada, agrupando la mayor cantidad de información relacionada en la menor cantidad de reportes. Es decir, clasificar las necesidades de información similares y agruparlas en conjuntos, y que los conjuntos de necesidades de información sean lo más heterogéneos posible.

La estrategia de BI mejora la eficiencia y la accesibilidad de la información a quien la necesite. También mejora las relaciones con los socios y amigos de la ONG entregando los datos para la promoción de esta y el vínculo social a través de los hechos reflejados en los datos.

#### 4. Conclusiones y trabajo futuro

Las directrices presentadas en este artículo permiten comenzar a liderar una implementación de una estrategia BI en una organización no gubernamental y ajustarla a las necesidades específicas de la organización.

La idea de medir el rendimiento de programas y la efectividad de campañas y acciones de los voluntarios es extraer el conocimiento para hacerlo repetible en otras ONG que tengan los objetivos similares, ya que entonces se conoce la fórmula de éxito que funciona. Las oportunidades de gestionar con información son muchas de aquí en adelante, como por ejemplo conocer los perfiles de su red de apoyo en cuanto a personas y empresas; perfilar a los donantes, voluntarios, y contactos. Analizar los datos del pasado, poder generar un modelo de predicción de donaciones, comunicar a la comunidad los datos reales y actualizados de trabajo de la organización, tener el pulso de cuántos, quiénes, cómo, dónde se está trabajando, y lo que es más importante optimizar el trabajo en las instituciones donde los vo-

luntarios colaboran a través del análisis de datos que éstas mismas generan.

#### Agradecimientos

A VE Global por la oportunidad de desarrollar este proyecto con ellos, ha sido un enorme enriquecimiento profesional y personal. A los voluntarios Bushra Akram y Ben Richman que han trabajado directamente en la implementación y hecho realidad el proyecto. A Josh Pilz, Director Ejecutivo de VE, por su apoyo y entusiasmo, a los directores y colaboradores Annie Rondoni, Mariah Healy, Jamie Ensey, Faith Joseph por su trabajo y excelente disposición en todo momento.

#### Bibliografía

**Cindi Hudson.** *Successful Business Intelligence: Secrets to Making BI a Killer App.* McGraw-Hill Osborne Media, 1 edition. 2007. ISBN-10: 9780071498517.

**Douglas Wubbard.** *How to Measure Anything: Finding the Value of Intangibles in Business.* Wiley; 2 edition, 2010. ISBN-10: 9780470539392.

**Olivia Parr Rud.** *Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy.* Wiley, 2009. ISBN-10: 9780470392409.

**Henri Rouillé D' Orfeuil.** *La Diplomacia No Gubernamental.* Lom - Chile, 2008. ISBN: 9562829723.

#### Notas

<sup>1</sup> <<http://www.ve-global.org/>>.

<sup>2</sup> <<http://www.dropbox.com/>>.

<sup>3</sup> <<http://www.ganttproject.biz/>>.

<sup>4</sup> <[http://freemind.sourceforge.net/wiki/index.php/Main\\_Page](http://freemind.sourceforge.net/wiki/index.php/Main_Page)>.

<sup>5</sup> Desarrollado con FreeMind.

Ariel Sabiguero, Andrés Aguirre, Fabricio González, Daniel Pedraja, Agustín Van Rompaey

Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Montevideo (Uruguay)

<{asabiguc,aaguirre}@fing.edu.uy>;  
<{fabgonz,danigpc,fenix.uy}@gmail.com>

# Extensiones al núcleo de Linux para reducir los efectos del envejecimiento del software

Este artículo ha sido seleccionado de entre las mejores ponencias presentadas en la XXXVI Conferencia Latinoamericana de Informática (XXXVI CLEI), evento anual promovido por el Centro Latinoamericano de Estudios en Informática (CLEI), entidad en la cual ATI es el representante de España.

## 1. Introducción

Algunas de las razones por las que las imágenes del software en ejecución se corrompen son las pérdidas de memoria, bloqueos infinitos, bloqueos en memoria compartida, hilos sin finalizar, fragmentación del almacenamiento, corrupción de los datos, interferencia por rayos cósmicos, fallos en la disipación térmica, entre otros. Este fenómeno es conocido como *envejecimiento del software*<sup>1</sup>, y ha sido observado en *clusters* de procesamiento, sistemas de telecomunicaciones, servidores web, PC domésticos y otros sistemas. Las evidencias de sus efectos son más comunes en sistemas que operan de forma ininterrumpida, como servidores web y de correo electrónico, entre otros.

Una forma ingenua, intuitiva y común para combatir la degradación de la imagen de software en ejecución es el reinicio periódico y preventivo de los sistemas. A esta técnica se la refiere como *rejuvenecimiento del software*<sup>2</sup>. La técnica trae al centro de cómputos la receta *amateur* de apagar y encender un dispositivo que no responde. Diferentes técnicas de rejuvenecimiento sugieren estrategias para los re-inicios preventivos, brindándonos la sensación de que estamos combatiendo el fenómeno desde un punto de vista tecnológico.

El presente trabajo analiza y aísla fuentes específicas de degradación de la imagen del software en memoria y su efecto en la ejecución de sus respectivos programas. Un profundo análisis de las técnicas de manejo de memoria del sistema operativo nos permite proponer técnicas de rejuvenecimiento que no requieren el reinicio del sistema. Se presentan, asimismo, los resultados de la implementación de un prototipo, así como los datos experimentales.

La organización del trabajo se describe a continuación. La **sección 2** provee una introducción y análisis del fenómeno del envejecimiento del software. La **sección 3** brinda una introducción al manejo de memoria en los sistemas operativos modernos, poniendo foco en el manejo de la memoria de solo-lectura y las técnicas de detección de corrupción. Posteriormente, la **sección 4** brinda detalles del prototipo implementado, junto con resultados experimentales relevantes.

**Resumen:** Las técnicas actuales de rejuvenecimiento de software atacan diferentes errores transitorios de los sistemas de una forma poco específica: restauración del proceso o sistema completo. Nuestro equipo implementó un prototipo que ataca el problema con una aproximación más fina. El prototipo implementado detecta y corrige la corrupción de páginas de memoria de solo-lectura en un sistema GNU/Linux. El presente trabajo describe la clase de errores considerados, el mecanismo de corrección de errores implementado y su aplicación a la confiabilidad y disponibilidad del sistema. Para evaluar la implementación, realizamos estudios en diferentes contextos de aplicación bajo cargas de trabajo simuladas. Mecanismos de inyección de fallas fueron utilizados para generar y simular errores en el sistema. Los resultados relativos a desempeño y mejora en la disponibilidad se presentan también, mostrando que la utilización de esta técnica es adecuada en condiciones generales de un sistema.

**Palabras clave:** rejuvenecimiento, soft-errors, técnicas de corrección de memoria.

Diferentes líneas de trabajo para el área son sugeridas en la **sección 5**. El trabajo concluye en la **sección 6**.

## 2. Envejecimiento del software

No todas las fallas en nuestros sistemas se deben a software mal construido. De hecho, podríamos tener software perfecto ejecutándose en un sistema, y dicho sistema podría fallar debido a causas externas. Existe una alta probabilidad de que en muchos de estos fallos, causas externas afecten el hardware del sistema, y por consiguiente al software que allí se ejecuta. También existe la posibilidad de que ciertas condiciones de hardware o software transitorias afecten la ejecución esperada. El resto de esta sección introduce conceptos relevantes relacionados al envejecimiento de software.

### 2.1. Taxonomía

Jim Gray [1] clasifica los errores de software en dos categorías distintas, Bohrbugs y Heisenbugs. Esta distinción se basa en la facilidad de reproducir la falla causada por estos errores.

La primera categoría, los Bohrbugs, que deben su nombre al modelo atómico de Bohr, son esencialmente errores permanentes en el diseño, y son por naturaleza casi deterministas. Pueden ser identificados y corregidos durante la fase de pruebas del ciclo de vida del software.

Por otro lado, los Heisenbugs, así nombrados por el principio de incertidumbre de Heisenberg, incluyen aquellas fallas internas que son intermitentes. Fallas cuyas condiciones de activación ocurren raramente, y por lo

tanto son difíciles de reproducir. Un manejo de excepciones incorrecto o incompleto es un claro ejemplo de Heisenbugs. Estos errores son más difíciles de detectar durante la fase de pruebas que los Bohrbugs.

K. Vaidyanathan y Kishor S. Trivedi [2] agregan una tercera categoría a esta clasificación, la cual incluye aquellas fallas causadas por la ocurrencia de envejecimiento de software.

Esta categoría de fallas es similar en ciertos aspectos a los Heisenbugs, ya que su activación está determinada por ciertas condiciones como la falta de recursos del sistema operativo, algo que no es sencillo de reproducir.

Sin embargo, sus métodos de recuperación difieren significativamente. Mientras que los Heisenbugs se corrigen mediante técnicas puramente reactivas, las fallas causadas por el envejecimiento de software pueden ser prevenidas mediante la aplicación de técnicas proactivas, como puede ser el reinicio periódico del software afectado.

Estamos particularmente interesados en los errores de hardware, especialmente aquellos que ocurren en los chips de memoria de un sistema. Estos errores también son conocidos como *soft errors* [3] y pueden ser clasificados como errores causados por envejecimiento de software, aunque no son causados por el envejecimiento en sí mismo, sino por causas externas, como pueden ser los rayos cósmicos, temperatura y humedad, entre otros.

### 2.2. Soft Errors

Como se mencionó en la **sección 2.1.**, los

*soft errors* son errores en el hardware de memoria de un sistema. Es importante enfocarse en ellos, ya que normalmente no se toman en cuenta al momento de diseñar software, y pueden eventualmente causar fallas totales en un sistema.

Sería interesante contar con algún tipo de protección contra estos errores, que no esté construida directamente sobre el hardware, como los chips de memoria con código corrector de errores (memoria ECC).

Debido a la naturaleza de estos errores, es imposible predecir cuándo ocurrirán, y por lo tanto, necesitamos una técnica reactiva que permita al software recuperarse frente a la aparición de uno de ellos.

Los *soft errors* pueden ser causados por una variedad de eventos. James F. Ziegler, un investigador de IBM, publicó varios artículos probando que los rayos cósmicos podían ser otra causa de aparición de *soft errors* [4].

Durante estos experimentos, IBM descubrió que la frecuencia de aparición de *soft errors* se incrementaba con la altura, duplicándose a 800 metros por encima del nivel del mar. En la ciudad de Denver, a 1.600 metros por encima del nivel del mar, la frecuencia de aparición de *soft errors* es 10 veces la del nivel del mar.

En otro experimento, descubrieron que los sistemas ubicados bajo tierra experimentaban una frecuencia de *soft errors* mucho menor, y dado que 20 metros de roca pueden bloquear casi la totalidad de rayos cósmicos, concluyeron que dichos rayos cósmicos eran efectivamente una posible causa de aparición de *soft errors*.

Otra posible causa de aparición de *soft errors* es la interferencia electromagnética (EMI). Esta interacción puede causar alteraciones en los transistores o buses y, con el tiempo, puede causar que su valor lógico cambie. Este tipo de interferencia es común en lugares donde operan motores eléctricos de alta frecuencia [5].

En entornos industriales donde se utilizan este tipo de motores y muchos sistemas empotrados controlan la maquinaria, la probabilidad de ocurrencia de un *soft error* es mucho más alta, lo cual puede llevar a daños en la maquinaria o peor aún, lesiones en los operarios humanos. Estos factores son cruciales a la hora de considerar aspectos de seguridad.

### 2.3. Soluciones y técnicas existentes

Las técnicas más comunes para combatir los errores causados por el envejecimiento de software se conocen como "técnicas de rejuvenecimiento de software".

Existen dos estrategias distintas que pueden utilizarse para determinar el momento adecuado para aplicar el rejuvenecimiento.

La primera es conocida como enfoque *open-loop* [2]. Este método consiste en aplicar el proceso de rejuvenecimiento sin utilizar ninguna clase de información sobre el desempeño del sistema. El rejuvenecimiento se puede aplicar luego de que una cierta cantidad de tiempo ha pasado desde la última aplicación, o cuando el número de trabajos concurrentes del sistema llega a un cierto número crítico.

Por otro lado, tenemos el enfoque *closed-loop* [2]. En este método, la salud del sistema es continuamente monitoreada para detectar la posibilidad de que ocurra un error causado por el envejecimiento, como la falta de un recurso particular del sistema, que pueda llevar a una disminución del rendimiento o a una falla total del sistema. En este enfoque podemos clasificar nuestras técnicas según la forma de analizar los datos recolectados.

El *análisis offline* utiliza datos sobre el rendimiento del sistema, recolectados durante un extenso período de tiempo (semanas o meses) para determinar la frecuencia óptima de ejecución del rejuvenecimiento, siendo adecuado para sistemas cuyo comportamiento es determinístico.

El *análisis online* se basa en juegos de datos recolectados mientras el sistema está en funcionamiento, que se combinan con juegos de datos anteriores, para decidir si es un buen momento para aplicar rejuvenecimiento, siendo aplicable para sistemas cuyo comportamiento es difícil de predecir.

Las técnicas de rejuvenecimiento de software se pueden aplicar en distintos niveles de granularidad.

Un ejemplo de rejuvenecimiento aplicado a nivel de sistema es un reinicio total de hardware, y a nivel de aplicación el reinicio de procesos afectados es otro ejemplo de rejuvenecimiento.

La aplicación de rejuvenecimiento tiene un impacto negativo en la disponibilidad del sistema rejuvenecido. Por este motivo es muy útil contar con una arquitectura de alta disponibilidad, donde podemos aplicar rejuvenecimiento a ciertos nodos individuales sin impactar en la disponibilidad del sistema globalmente.

Si diseñamos nuestros sistemas teniendo en cuenta todos estos factores, aplicar rejuvenecimiento ha probado ser una técnica proactiva muy efectiva y de muy bajo costo [6] para prevenir errores. Con este enfoque no estamos mejorando la confiabilidad del sistema pero sí la disponibilidad de los servicios.

### 3. Considerando los *soft errors* y la memoria de solo lectura

Las arquitecturas de hardware y sistemas operativos modernos proveen mecanismos de protección sobre la memoria disponible, esencialmente, asignando privilegios de acceso a cada una de las páginas de memoria virtual. La asignación de privilegios de acceso a las páginas de memoria está directamente soportada por la mayoría de las arquitecturas, debido al soporte provisto por los microprocesadores actuales.

A lo largo del presente trabajo nos referimos con el término *memoria de solo-lectura* (*RO por sus iniciales en inglés*) a las partes de la memoria de un sistema a las cuales los mecanismos de protección del sistema operativo prohíben la modificación de su contenido.

De las múltiples facilidades ofrecidas por el sistema de memoria, cabe destacar que en cualquier instante del tiempo podemos clasificar las páginas de un proceso en dos categorías: solo-lectura y modificables. Las páginas de solo-lectura son constantes a lo largo de la ejecución de un programa.

En base a la definición anterior de memoria de solo-lectura, es directo ver como los cambios observados en dichas regiones son interpretados como una falla.

Assumiendo que el mecanismo de protección brindado por el sistema operativo está carente de errores, el único motivo por el que podemos encontrar un cambio en la memoria de solo lectura es debido a un *soft error*. Bajo esta hipótesis asumimos que el uso de técnicas de detección de cambios en memoria de solo-lectura es equivalente a la detección de *soft errors*.

#### 3.1. Memoria RO en software estándar

Cualquier programa corriendo en un sistema operativo estándar actual está conformado por un conjunto de archivos-objeto, conteniendo bibliotecas y ejecutables. Compiladores y enlazadores trabajan en conjunto con el SO para convertir código fuente en código objeto que puede ser cargado y ejecutado en su plataforma objetivo. Existen muchos formatos diferentes para archivos ejecutables, como ELF (*Executable and Linking Format*), normalmente utilizado en sistemas GNU/Linux, o PE (*Portable Executable*) utilizado en sistemas basados en Windows NT.

La evolución en los lenguajes de programación, verificaciones y restricciones impuestas por los nuevos compiladores, están convirtiendo el código ejecutable en un código más seguro. Los ejecutables construidos con compiladores actuales fuerzan la separación del código objeto en páginas de solo-lectura y otras con privilegios de escritura.

La proporción entre éstas es dependiente de los programas particulares en sí mismos. Programas compactos que manipulan grandes volúmenes de datos (p.e. software de *streaming*), presentan una cantidad reducida de páginas de memoria de solo-lectura. Programas de alta complejidad que manipulan juegos de datos pequeños (p.e. paquetes de oficina) presentan una tasa mucho mayor de páginas de solo-lectura. Es por esta razón que la cobertura ofrecida por la presente técnica no es la misma en todos los escenarios. Algunos resultados experimentales son presentados en la **sección 4.3**.

### 3.2. Memoria RO en Linux

El formato estándar de objetos, ejecutables y bibliotecas en GNU/Linux es el formato ELF.

ELF soporta el concepto de secciones, que son colecciones de información de tipos similares. Cada sección representa una porción del archivo. A modo de ejemplo mencionamos que el código ejecutable es colocado en una sección conocida como `.text`. Las variables inicializadas por el usuario son colocadas en una sección denominada `.data` y los datos sin inicializar en la sección `.bss`.

El gestor de memoria puede marcar porciones de memoria como RO de forma tal que los intentos de modificación de las mismas resultan en la finalización de la ejecución del programa por parte del sistema operativo. Es por ello que, en vez de no esperar que una posición de memoria RO cambie como consecuencia de un error en el programa, tenemos la certeza de que cualquier intento de modificación de una posición RO es un error fatal, indicando un fallo en el software. El sistema operativo se encarga de finalizar el programa que ofendió la protección.

Dado que deseamos que las porciones ejecutables estén en páginas de memoria RO y las posiciones de memoria modificables en memoria de lectura-escritura, resulta más eficiente agrupar todas las porciones ejecutables en la sección `.text`, y todas las áreas que pueden ser modificadas en la sección `.data`. Los datos experimentales relativos a un servidor web estándar (LAMP) son presentados en las **secciones 4.2. y 4.3**.

Los sistemas empotrados representan otro entorno de aplicación en el que GNU/Linux está ganando terreno. Estos sistemas empotrados corren generalmente en plataformas en las que no se dispone de mecanismos de detección y corrección de errores en memoria.

Los beneficios de una solución estándar, basada en servicios del sistema operativo, son de gran interés en estos sistemas que corren Linux, en los que las fallas tienen un impacto directo en características como la confiabilidad de los mismos.

## 4. Implementación de un prototipo en GNU/Linux

Describiremos la implementación de un prototipo que hace uso de la hipótesis presentada en la **sección 3** para combatir los efectos del envejecimiento en el software causado por errores en la memoria. La herramienta se desarrolló como una extensión al núcleo de GNU/Linux, extendiendo la funcionalidad del subsistema de memoria (*Memory Manager*).

En la **sección 4.1** describimos la funcionalidad del prototipo y los detalles más relevantes de su implementación. Dado que su función se basa en la detección de *soft errors*, las pruebas de validación requirieron la alteración de la memoria, de forma que pudieran ser consideradas equivalentes a los errores que buscamos detectar. Como se describe en la **sección 4.2**, fue necesario hacer uso de técnicas de inyección de fallos, lo que nos permitió simular errores de hardware de una forma repetible, económica y segura. Por último, en la **sección 4.3** presentamos los resultados de algunas pruebas funcionales y de desempeño del prototipo ejecutándose en una instancia concreta de GNU/Linux. Las pruebas de desempeño se basan en pruebas estándar de algunos aplicativos en Linux.

### 4.1. Detalles de la implementación

Como parte del trabajo de tesis de González, Pedraja y Van Rompaey [7], se construyó un prototipo basado en Linux. La primera meta de este prototipo fue detectar la ocurrencia de *soft errors* en la memoria física del sistema, sacando provecho de la existencia de regiones de memoria de solo lectura.

Como se vio en la **sección 3**, Linux administra los recursos de memoria con granularidad de páginas, por lo que el primer paso para lograr el objetivo mencionado es identificar el subconjunto páginas para los que el S.O. nos garantiza acceso de solo lectura. Este esquema de protección es representado por medio de los permisos de acceso de las áreas de memoria virtual que forman parte del espacio de direcciones de un proceso y se hace cumplir por medio de las tablas de páginas del proceso. El mecanismo de memoria virtual de Linux hace posible que las páginas de memoria sean compartidas entre los espacios de memoria de varios procesos con diferentes permisos de acceso en cada caso. Tomando esto en cuenta, se puede definir el subconjunto de páginas de solo-lectura como aquellas que están asignadas al espacio de direcciones de uno o más procesos del espacio de usuario, con permisos de solo-lectura en todos los casos.

Todos los eventos de cambios en la asignación, y permisos de acceso de las páginas de memoria, fueron analizados. Manejando estos eventos y manteniendo información de

estado extra, fue posible determinar en cualquier momento la pertenencia o no de una página dada al subconjunto deseado. Los cambios realizados al núcleo se limitaron lo más posible y la mayoría de la funcionalidad fue encapsulada e implementada en un módulo del núcleo (ver **figura 1**).

Una vez identificado el subconjunto de páginas de solo-lectura, el segundo paso es contar con un mecanismo de detección de cambios a nivel de bits sobre sus elementos. Cualquier cambio en un bit en este contexto se puede interpretar como la ocurrencia de un *soft error*.

La solución general para proveer esta clase de funcionalidad es mantener, para cada página en el conjunto, un código de redundancia sobre los bytes pertenecientes al rango de memoria que este representa. En el prototipo Linux, esto fue implementado agregando un código de redundancia al estado mantenido para cada página de memoria. El código tenía como requerimiento proveer detección de errores en múltiples bits de una forma eficiente, es decir, usando una pequeña cantidad de bits de redundancia para los volúmenes de datos objetivo (el tamaño de una página de memoria va desde 4KB a 4MB).

Luego de evaluar diferentes opciones, el algoritmo de detección de errores seleccionado para esta tarea fue CRC32. La redundancia mantenida debe ser actualizada en los eventos mencionados anteriormente, de forma que sea la correspondiente a los contenidos de la página mientras la misma pertenece al conjunto de solo-lectura. La estrategia seguida para garantizar esto consiste en recalcular y almacenar el código cada vez que la página ingresa a un estado que lo posiciona dentro del conjunto (por ejemplo cuando la primer asignación de solo-lectura es realizada sobre una página que previamente estaba libre), y borrar (o ignorar) la redundancia en transiciones a estados fuera del conjunto (como puede ser cuándo una asignación de escritura es agregada a una página que era previamente de solo-lectura). Gracias al estado almacenado, verificar la existencia de errores es tan simple como recalcular el CRC32 y compararlo con el almacenado.

La herramienta puede ser configurada para seguir una de tres estrategias de búsqueda de errores. Para comprender las diferentes estrategias, es importante tener en cuenta que, debido a que estamos tratando de reaccionar a eventos que ya habrán sucedido, la detección de errores será asíncrona a la ocurrencia de los mismos y siempre habrá un retraso entre ambos sucesos. Hacer ese intervalo de tiempo lo más pequeño posible fue la meta al seleccionar las estrategias de búsqueda, puesto que esto reduce las probabilidades de que un proceso acceda a la sección de memoria

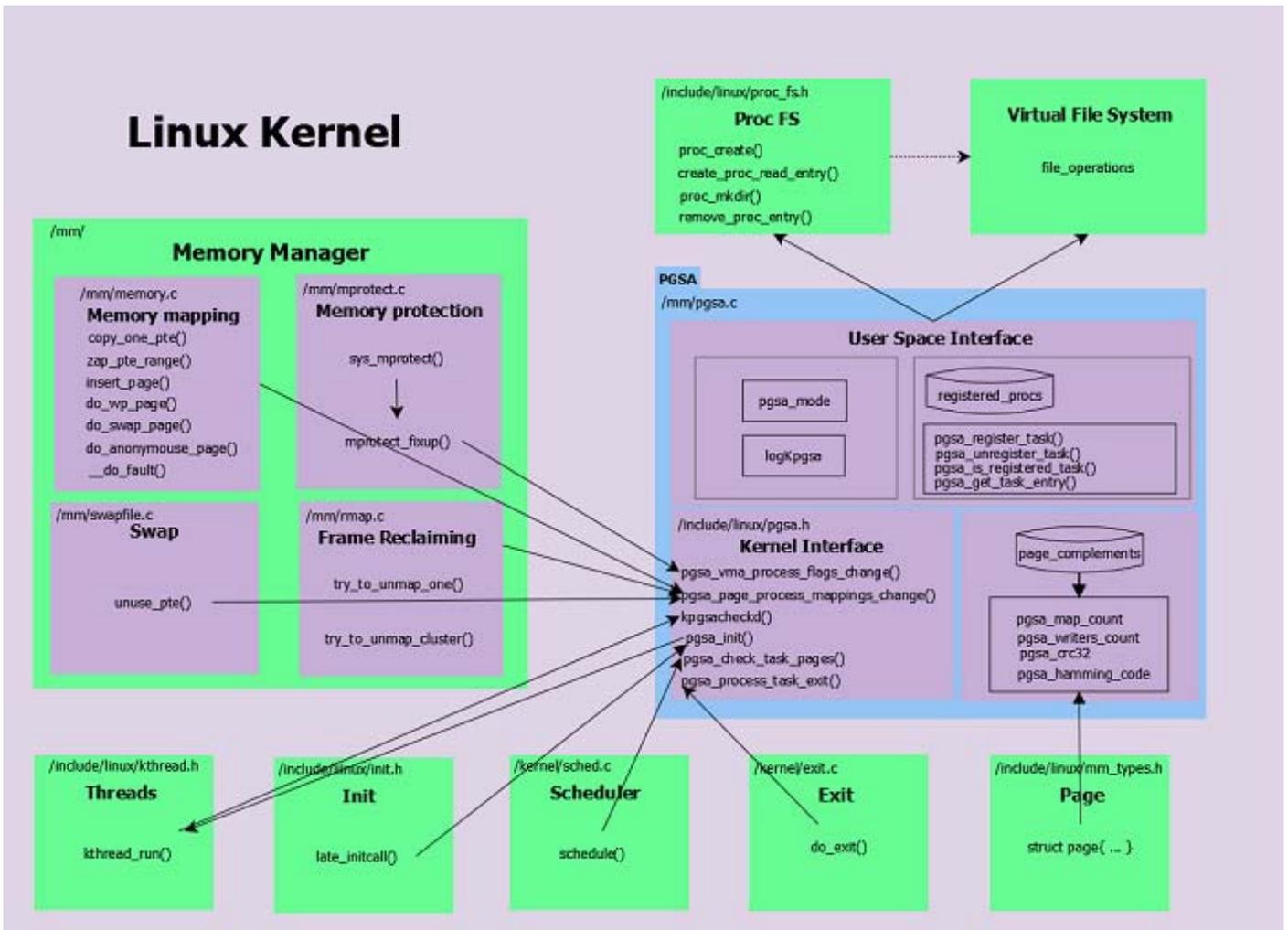


Figura 1. Cambios realizados al núcleo de Linux.

corrupta antes de que se puedan tomar acciones.

El primer y más simple de los enfoques implementado consiste en un hilo del núcleo que chequea continuamente todo el conjunto de páginas de solo-lectura en un ciclo constante. Aunque esta estrategia puede sugerir una pérdida de desempeño, en la práctica ha probado ser bastante efectiva.

El prototipo brinda algunas funcionalidades de configuración, las que permiten a un administrador del espacio de usuarios ajustarla de acuerdo a cada escenario y sus requerimientos. Trataremos sobre las mismas luego, pero se da aquí un adelanto sobre el registro de procesos, que es esencial para la existencia de las dos estrategias de búsqueda restantes. El registro de procesos básicamente permite al administrador configurar la herramienta de forma que ésta enfoque sus esfuerzos en un grupo indicado de procesos.

Cuando el módulo es configurado para trabajar solo con tareas registradas, puede realizar verificaciones de errores de dos maneras. La primera es similar a la estrategia ya explicada, solo que en este caso el hilo itera sobre

la lista de tareas registradas chequeando solamente aquellas páginas de solo lectura asignadas al espacio de memoria de la tarea actual. Esta estrategia ha probado ser considerablemente más eficiente que la anterior.

Los experimentos realizados para comparar los retrasos de detección de errores entre ambas técnicas mostraron un intervalo de tiempo promedio de 275 milisegundos para la básica. Por otro lado, las pruebas indicaron un retraso promedio de 1 milisegundo para la estrategia que usa los procesos registrados, aún cuando se estaba simulando carga al sistema al registrar las tareas del *stackLAMP*.

La última estrategia de búsqueda de errores implementada también trabaja solo con las tareas registradas. La misma fue desarrollada modificando el planificador de tareas del núcleo, para chequear las páginas de solo-lectura asignadas al espacio de memoria de la próxima tarea a obtener la UCP, justo antes de que realmente obtenga el recurso. La meta de esta estrategia es permitir la toma de acciones en todas las secciones de memoria corruptas pertenecientes al espacio de memoria de un proceso, antes de que el mismo obtenga el control de la UCP y logre acceder

a dichas secciones. A diferencia de los dos enfoques presentados previamente, en este caso el objetivo no es reducir el retraso de detección de los errores. Esta estrategia ha probado ser efectiva, pero puede también significar una importante sobrecarga en la planificación de procesos, por lo que debería ser usada con extrema cautela.

El solo hecho de la detección no es suficiente, por lo que se definió una secuencia de acciones para el manejo de errores como parte del prototipo. El primer paso consiste en emplear un código corrector de errores, para tratar de arreglar la memoria afectada de forma automática. Existen varios tipos de códigos ampliamente utilizados para este tipo de tarea, entre los cuales se seleccionó una adaptación de los *Códigos de Hamming*, debido a su fácil implementación. Los Códigos de Hamming son en realidad el método más comúnmente usado por los chips de memoria ECC, pero en este caso la corrección de un solo bit que proveen puede no ser suficiente para toda una página de memoria. Con diferentes características de cubrimiento y desempeño, las técnicas Reed-Solomon podrían haberse aplicado, pero las mismas fueron descartadas debido a la complejidad de su implementación.

La segunda acción tomada para tratar de corregir errores se basó en el concepto de rejuvenecimiento, concebido como técnica pro-activa de manejo de fallos, orientada a limpiar el estado interno del sistema para prevenir la ocurrencia de fallos más severos que provoquen caídas en el futuro. Una posible solución para corregir páginas de memoria de solo-lectura corruptas usando rejuvenecimiento, es mantener un respaldo de sus contenidos y recargar el mismo ante errores. La mayoría de las páginas de solo-lectura son la imagen de un archivo en disco, por lo que el propio archivo se convierte en el respaldo requerido para estas páginas. El prototipo provee una funcionalidad que recarga automáticamente la imagen de una página desde disco, cuando se detecta un *soft error* dentro del mismo. Para el caso de páginas no vinculadas a archivos (también llamadas anónimas) se debería proveer una implementación de almacén de respaldo, pero su desarrollo excedió el alcance del prototipo.

Se encontró una forma de aplicar rejuvenecimiento cuando se ejecuta el núcleo, pero en realidad la mayoría de la teoría sobre esta técnica apunta al espacio de usuario. Dado que muchas estrategias de rejuvenecimiento se basan en estadísticas de los recursos para decidir cuándo aplicarlo, se entendió que en este caso el rol del SO era proveer información sobre la detección de errores y notificaciones, de forma que asista a esas decisiones.

Primero, se decidió brindar notificaciones sincrónicas de los errores detectados a los agentes de rejuvenecimiento del espacio de usuario. Esta funcionalidad se implementó por medio de señales de Linux. Cuando un proceso es registrado en la herramienta, se especifica también otro proceso que toma el rol de agente de rejuvenecimiento de la tarea registrada. En caso de detección de errores en el espacio de memoria de la tarea, una señal específica es enviada a su agente para que éste pueda tomar acciones.

A veces el rejuvenecimiento simplemente consiste en el reinicio de procesos o incluso de todo el sistema. Sin embargo, si se cuenta con la información adecuada, se pueden tomar acciones de mayor granularidad (por ejemplo, volver a asignar solo un archivo). Para brindar a los agentes la oportunidad de reconocer tales acciones, el prototipo publica al espacio de usuario información detallada sobre cada error detectado. Estos detalles incluyen para cada proceso (registrado o no), las direcciones física y virtual exactas de los bytes cambiados, el área virtual de memoria a la cual pertenece la dirección en el espacio de direcciones del proceso y a qué tipo de vínculo (archivo, anónimo) corresponde. En caso de que el error ocurra en la asignación de un archivo, se provee también el nombre del archivo y el rango asignado.

Finalmente, el estado resultante del error (corregido o no) también se muestra. La interfaz con el espacio de usuarios seleccionada para presentar esta información fue sistema de archivos virtual/*proc*, usando carpetas y archivos a nivel de tareas.

Se mencionó que el prototipo también provee algunas funcionalidades de configuración, como los ya explicados procesos registrados. El módulo soporta además el concepto de modo global el cual está compuesto por un conjunto de banderas. Cada una de estas banderas permite al administrador activar o desactivar diferentes funcionalidades. Se crearon banderas adicionales para poder elegir entre las posibles estrategias de búsqueda de errores y para encender y apagar las acciones de manejo de errores o incluso todo el módulo. La interfaz con el espacio de usuarios para la configuración del modo y las tareas registradas fue implementada también usando el sistema virtual de archivos/*proc*, mediante la escritura en archivos especiales.

Se ha usado el término *módulo* a lo largo de esta sección para referirse a la implementación del prototipo, pero no se ha explicado como la misma se integra realmente con el núcleo de Linux.

Como el término lo implica, todas las funcionalidades presentadas se agrupan en un único módulo en el núcleo. El mismo provee un grupo de interfaces con un conjunto de *callbacks* que le permiten manejar los eventos internos del núcleo y las interfaces con el espacio de usuarios. Estas *callbacks* se diseñaron para reducir lo más posible el acoplamiento del núcleo con el prototipo. Cuando se dice módulo, no se refiere al concepto de módulo dinámico que soporta el núcleo. Dichos módulos son capaces de agregar funcionalidad a una instancia activa del núcleo y son mayormente usados para desarrollar controladores de dispositivos. En este caso, debido a las secciones del núcleo donde las interfaces del módulo son invocadas, necesitamos agregar funcionalidad al núcleo en tiempo de compilación. Debido al tipo de funcionalidad que provee, se decidió incluir los fuentes del módulo como parte del subsistema de manejo de memoria del núcleo de Linux.

#### 4.2. Pruebas funcionales del prototipo

La verificación de la correctitud del prototipo fue importante por varias razones. Por un lado, siempre está la necesidad de agregar calidad a la implementación que está siendo validada, pero, por otra parte, se requería separar los defectos del prototipo de los *soft errors* en la memoria. Los tests consistieron en alterar valores de memoria localizados en páginas de solo lectura y verificar que los cambios fueron detectados y corregidos.

Para el diseño de los casos de prueba no se contaba con medios para inyectar errores por hardware en los subsistemas de memoria, por lo que se recurrió a la inyección de errores por software. Los detalles tecnológicos implicados están fuera del alcance de este trabajo, pero simplemente podemos mencionar que una API de alto nivel del núcleo no puede ser usada para modificar las páginas de memoria de solo-lectura. Alterar los valores de memoria de solo-lectura desde el espacio de usuarios genera fallos de segmentación. Alterar los valores de memoria de solo-lectura desde una API portable de alto nivel (como *ptrace*) genera efectos no deseados; el núcleo convierte la página de solo-lectura en una de lectura-escritura antes de efectivizar el cambio. Tan pronto como la página se convierte en una de lectura-escritura, es removida del conjunto de páginas monitoreadas por el prototipo.

La inyección de errores por software debe ser hecha en modo supervisor, donde se obtiene acceso sin restricciones a la memoria y las modificaciones directas a la misma pueden saltar el esquema de protección de memoria del núcleo. Se implementó la inyección direccionando la memoria directamente desde el espacio virtual del núcleo, evitando las limitaciones de las tablas de páginas de los procesos de espacio de usuarios. Se expuso esta funcionalidad al espacio de usuarios por medio de una nueva llamada al sistema (nombrada *sys\_kpgsa\_inject()*) y se desarrolló una variedad de herramientas de alto nivel. Las opciones incluyen desde alterar posiciones de memoria de un solo byte, hasta simular fallos de memoria periódicos y aleatorios (lluvia de rayos cósmicos).

#### 4.3. Pruebas de desempeño

Un sistema con un buen rendimiento es importante, a pesar de las positivas consecuencias sobre la confiabilidad, disponibilidad y seguridad física (*safety*) de las técnicas de rejuvenecimiento automático y de software de propósito general. Con el fin de monitorear los cambios, la memoria de solo-lectura del sistema debe ser explorada completamente de forma periódica, transformando el prototipo en una aplicación con uso intensivo de UCP, sin impacto en el uso de entrada/salida. Esto introduce competencia por el acceso a memoria, y además, pérdida en la localidad.

Se analizaron dos escenarios diferentes de aplicaciones: aplicaciones con uso intensivo de UCP y las de uso intensivo de entrada/salida. Como aplicación intensiva en uso de UCP se seleccionó POV-Ray [8] (ver **figura 2**), y la de uso intensivo en entrada/salida es un servidor LAMP<sup>3</sup>.

Para medir el impacto sobre el rendimiento del sistema utilizamos las propias herramientas de *benchmarking* que incluyen estos paquetes de software, comparando su rendimiento en

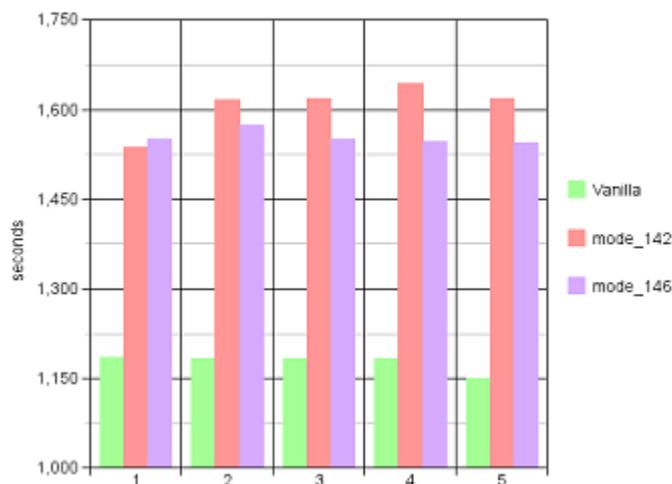


Figura 2. POV-Ray: tiempo de render.

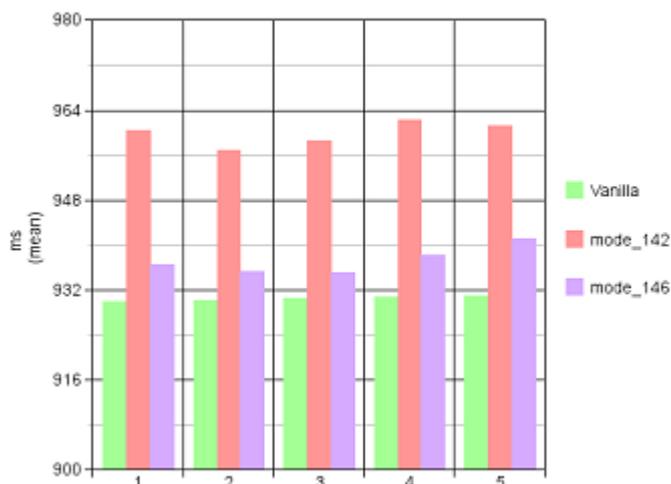


Figura 3. LAMP: tiempo de servicio.

el mismo sistema, con y sin uso de las rutinas de corrección de corrupción de memoria. Los tests fueron realizados en ambos modos.

Las mediciones de desempeño fueron consistentes con la intuición. El impacto en el rendimiento en un servidor LAMP es marginal, como puede verse en la figura 3, mientras el impacto en la aplicación de técnicas de raytracing es significativo, como se muestra en la figura 2.

En otras palabras, las rutinas de corrección de corrupción de memoria prácticamente no compiten con tareas que hacen un uso intensivo de IO mientras si lo hacen con las tareas que hacen uso intensivo de UCP.

Como última observación sobre el impacto en el funcionamiento del software, no todos los sistemas tienen la misma proporción de páginas de solo-lectura sobre páginas de lectura-escritura. En base a los resultados anteriores analizamos el mecanismo de funcionamiento de Apache y MySQL, y ambos comparten el paradigma de atención basado en fork. Los procesos padre tienen una tasa superior a la de los hijos, como se puede ver en la tabla 1. Los resultados observados son consistentes con el hecho que el proceso padre no almacena datos de ejecución requeridos para contestar a clientes particulares, sino información de gestión global del servicio.

Los procesos que tratan con los pedidos son los clientes. Los procesos padres están en ejecución durante más tiempo que los hijos en este escenario, y son estos los procesos que más se benefician de los algoritmos de corrección de corrupción.

El impacto de un *soft error* no corregido en un proceso hijo podría afectar solamente a una única sesión, mientras que un error similar en un proceso padre podría afectar a la totalidad del servicio de allí en más. La detección de corrupción toma más relevancia en los procesos padre, donde además la tasa RO/RW es también mayor.

### 5. Resultados y trabajo futuro

Quedan todavía diferentes parámetros de ajuste ya identificados que deberían ser agregados a la implementación, de forma que se vuelva más adaptable a diferentes escenarios de aplicación. *freqd* y otros demonios de compensación de UCP (o *CPU equalization*) tienen que ser considerados para alcanzar también las plataformas móviles. La implementación actual mantendría el UCP de un *notebook* al 100%, consumiendo la batería demasiado rápido.

Portar el prototipo a diferentes plataformas es necesario para acceder al mercado de sistemas empotrados. Su uso en sistemas empotrados, donde comúnmente no se cuenta con

memoria ECC, es un área donde el impacto puede ser significativo.

La implementación actual probó su capacidad de mejorar la disponibilidad de un servidor LAMP, bajo carga pesada y bajo lluvia pesada de rayos cósmicos simulada. Un sistema sin una implementación de corrección de corrupción de memoria de solo-lectura, falla luego de unos pocos segundos de modificaciones aleatorias de su memoria. La corrección de corrupción de memoria mostró que es posible hacer que un sistema recargue imágenes frescas del software en la RAM cuando ocurre la corrupción, sin impacto en los servicios que este provee.

Como observación final, todavía tenemos que asegurarnos que la implementación es capaz de detectar y corregir *soft errors reales*. Nuestras estimaciones [7] indican que en sistemas individuales de vanguardia (4GB DDR2 memory), habrá un error corregible, manejable para nuestro prototipo, cada 22 años en promedio por sistema. Todavía no hemos encontrado ningún error en núcleos modificados ejecutándose en ambientes de producción. Es deseable obtener acceso a ambientes donde la tasa de error es mayor.

### 6. Conclusiones

La industria ya sabe acerca del *software aging* o envejecimiento del software: cuando un

	MySQL				Apache httpd			
	Padre		Hijo		Padre		Hijo	
Páginas de solo lectura	287	47%	614	13%	930	38%	952	23%
Páginas de lectura/escritura	330	53%	4235	87%	1487	62%	3245	77%

Tabla 1. Tasa de páginas RO/RW en el software utilizado.

sistema comienza a tener un comportamiento errático, lo primero es reiniciarlo.

Las técnicas de rejuvenecimiento de software han ganado un lugar más significativo en la administración de sistemas en los últimos años y distintos productos las introducen, con diferentes niveles de granularidad, como parte de sus soluciones potenciadas para la disponibilidad. Hasta donde llega nuestro conocimiento, solo el rejuvenecimiento a nivel de todo el sistema o todo el proceso han sido presentados hasta el momento.

Nuestro prototipo ha probado que el rejuvenecimiento de granularidad fina es posible, como una herramienta del sistema operativo, en ambientes de aplicación de propósito general. Aunque la solución implementada solo cubre las páginas de memoria de solo lectura, la misma no requiere ningún cambio a nivel de despliegue, como es el caso de técnicas basadas en réplicas y balance de carga.

Debido al hecho de que es aplicable a cualquier pieza de software sin necesidad de realizar cambios, ofrece una forma simple y no intrusiva de reducir los costos de manejo de un ambiente de producción.

Todavía no podemos reemplazar la memoria ECC con una solución por software, pero nuestra implementación de corrección de la corrupción de memoria es capaz de mejorar la disponibilidad de los sistemas donde la tecnología ECC no está disponible.

Aunque todavía se requiere más investigación, mostramos que es posible mejorar la disponibilidad y confiabilidad de aplicaciones arbitrarias con técnicas generales de rejuvenecimiento de granularidad fina, implementadas a nivel del sistema operativo, con un impacto razonable en el desempeño.

#### Referencias

- [1] **Jim Gray.** Why do computers stop and what can be done about it? *Tandem Computers. Technical Report 85.7, PN 87614*, <<http://www.hpl.hp.com/techreports/tandem/TR-85.7.pdf>>, 1985.
- [2] **Kishor S. Trivedi, Kalyanaraman Vaidyanathan, Katerina Goseva-popstojanova.** Modeling and analysis of software aging and rejuvenation. *Proceedings of the IEEE Annual Simulation Symposium*, pages 270–279, 2000.
- [3] **Actel Corporation.** Understanding soft and firm

errors in semiconductor devices. *Actel Corporation. 51700002-1/12.02*, <[http://www.actel.com/documents/SER\\_FAQ.pdf](http://www.actel.com/documents/SER_FAQ.pdf)>, 2002.

[4] **J. F. Ziegler.** Terrestrial cosmic rays. *IBM J. Res. Dev.*, 40(1): pp.19–39, 1996.

[5] **G.L. Skibinski, R.J. Kerkman, D. Schlegel.** Emi emissions of modern pwm ac drives. *Industry Applications Magazine, IEEE*, 5(6): pp. 47–80, nov/dec 1999.

[6] **Artur Andrezejak, Monika Moser, Luis Silva.** Managing Performance of Aging Applications Via Synchronized Replica Rejuvenation. En Alexander Clemm, Lisandro Zambenedetti, and Rolf Stadler, editors, (DSOM 2007) *Managing Virtualization of Networks and Services*, ISBN 978-3-540-75693-4, pages 98–109. Springer, 2007.

[7] **Daniel Pedraja, Fabricio González, Agustín Van Rompaey.** *Herramientas del sistema operativo para combatir el software aging*. Proyecto de grado, <<http://www.fing.edu.uy/~asabigue/prgrado/softwareAging.pdf>>.

[8] **Persistence of Vision Raytracer.** *POV-Ray*, <<http://www.povray.org>>.

#### Notas

- <sup>1</sup> Del inglés *software aging*.
- <sup>2</sup> Del inglés, *software rejuvenation*.
- <sup>3</sup> Linux, Apache, MySQL, PHP.



# JAIIO

Las **JAIIOs**, Jornadas Argentinas de Informática, se realizan desde 1961 y son organizadas por **SADIO**.

En sesiones paralelas se presentan trabajos que se publican en Anales, se discuten resultados de investigaciones y actividades sobre diferentes tópicos, desarrollándose también conferencias y reuniones con la asistencia de profesionales argentinos y extranjeros. Las JAIIOs se organizan como un conjunto de simposios separados, cada uno dedicado a un tema específico, de uno o dos días de duración, de tal forma de permitir la interacción de sus participantes.

Una lista completa de todos los simposios de las 40 JAIIO se pueden encontrar en: <<http://www.40jaiio.org.ar>>.

**Fecha:** del 29 de agosto al 02 de septiembre de 2011

**Lugar:** Ciudad de Córdoba, Argentina (UTN - Facultad Regional de Córdoba).

Edmundo Sáez Peña  
Junta de Andalucía

<edmundo.saez@juntadeandalucia.es>

# La protección de datos personales en el desarrollo de software

## 1. Introducción

Una de las primeras etapas que se llevan a cabo durante el proceso de desarrollo de software es la que se conoce como análisis del sistema de información, donde se realiza, entre otras tareas, la definición detallada de los requisitos del sistema con el objetivo de obtener un catálogo de requisitos que permita definir con precisión el sistema que se pretende desarrollar. La participación de los usuarios en esta etapa es fundamental, pues sus especificaciones sobre lo que esperan que sea el comportamiento del sistema serán las que posteriormente guíen el proceso de desarrollo software.

Si bien los requisitos del sistema son, en muchos desarrollos, fijados únicamente por los usuarios en base a sus necesidades, en los sistemas que tratan datos de carácter personal los requerimientos de éstos tienen que ser considerados conjuntamente con las restricciones e imposiciones que exigen tanto la Ley Orgánica de Protección de Datos (LOPD) [1], como el Real Decreto que la desarrolla (RDLOPD) [2]. Por tanto, la libertad del cliente para decidir sobre la funcionalidad de su nuevo sistema encontrará en este caso ciertos límites.

Más aún, el RDLOPD incorpora una única disposición adicional sobre los productos de software, la cual estipula que deberá incorporarse, en la descripción técnica de los productos software destinados al tratamiento de datos de carácter personal, el nivel de seguridad, básico, medio o alto, descrito en dicha norma, que permitan alcanzar. Esto implica que, no sólo es necesario incorporar al sistema la funcionalidad necesaria para hacer cumplir con las medidas de seguridad del nivel al que vaya dirigido, sino que, además, debe quedar constancia en su descripción técnica de que el sistema permite el tratamiento de datos personales cumpliendo con las medidas de seguridad del nivel exigido impuestas por el RDLOPD. Así, en el caso de una hipotética sanción al responsable del fichero por parte de la Agencia Española de Protección de Datos (AEPD) [3], en la que se demostrara que la infracción cometida ha sido originada por un desarrollo software que no cumple con los preceptos legales en materia de protección de datos, éste podría actuar judicialmente contra quien hubiera desarrollado dicho sistema, dado que la descripción técnica del mismo establece que cumple con

**Resumen:** La Ley Orgánica de Protección de Datos de Carácter Personal y su Reglamento de desarrollo imponen una serie de restricciones y medidas de seguridad a los sistemas informáticos que traten datos personales de forma que el desarrollo de un sistema de estas características debe necesariamente observar dicha normativa. A lo largo de este artículo se realiza un análisis de las implicaciones que la normativa vigente en materia de protección de datos tiene en el desarrollo de software, así como de las medidas que dicho software debe implantar con objeto de cumplir con las exigencias legales.

**Palabras clave:** desarrollo de software, LOPD, protección de datos personales.

los requisitos necesarios para tratar datos personales a determinado nivel de seguridad. Por tanto, es imprescindible tener presente las implicaciones de la normativa vigente en materia de protección de datos personales a la hora de desarrollar un sistema de información.

Este artículo analiza dichas implicaciones y muestra como deben trasladarse al sistema de información. Para ello, realiza en primer lugar una descripción de los distintos niveles de seguridad en que se categorizan los ficheros que contienen datos personales. Seguidamente, analiza las implicaciones que las medidas de seguridad conllevan en el desarrollo de sistemas de información, en función de cada uno de los tres niveles de seguridad existentes. A continuación, se realizan otras consideraciones que deben tenerse en cuenta a la hora de diseñar un sistema de información, no derivadas directamente de las medidas de seguridad pero sí de otros aspectos de la normativa. Finalmente, se ofrecen unas breves conclusiones.

## 2. Niveles de seguridad

La protección de los datos de carácter personal en los sistemas informáticos viene garantizada a través de las medidas de seguridad impuestas por el RDLOPD. La aplicación de las medidas de seguridad se organiza en base al concepto de nivel de seguridad, de forma que se establecen tres niveles de seguridad: básico, medio y alto. A los ficheros con datos de carácter personal en una organización les corresponderá siempre uno de estos niveles de seguridad. A la hora de desarrollar un sistema de información es necesario tener claro qué tipo de datos van a manejarse y cuál es la finalidad de estos datos, pues en base a esta información al fichero resultante le corresponderá un determinado nivel de seguridad y, por tanto, el sistema software deberá incorporar las medidas de seguridad correspondientes. La **tabla 1** ofrece una relación de los

niveles de seguridad que corresponden a cada fichero con datos personales en función de los datos que contiene y la finalidad de los mismos, y constituye una síntesis del artículo 81 del RDLOPD.

Una inspección detenida de la **tabla 1** refleja que, en principio, cualquier fichero con datos de carácter personal recibe la consideración de fichero de nivel básico, y por tanto le corresponde adoptar las medidas de seguridad propias de este nivel. Sin embargo, existen determinados tipos de datos que se consideran más sensibles y que, por tanto, merecen unas medidas de protección más elevadas.

En primer lugar, se encuentran aquellos datos que elevan el nivel de seguridad de un fichero hasta el nivel medio. Como puede verse, son numerosos los ficheros a los que les corresponde este nivel de seguridad, aunque generalmente su descripción en la tabla permite identificarlos con facilidad. Quizás el tipo de ficheros que pueden suscitar algún tipo de duda son aquellos que "ofrezcan una definición de las características o de la personalidad de los ciudadanos y que permitan evaluar determinados aspectos de la personalidad o del comportamiento de los mismos". A este respecto, el Informe Jurídico 0590/2008 de la AEPD establece que se encuentran dentro de la categoría de este tipo de ficheros todos aquellos que contengan datos "a partir de los cuales puedan deducirse los hábitos, preferencias, aficiones, actividades o posición económica de los afectados", o bien "datos curriculares que incluyan información muy detallada sobre el sujeto que permitan deducir un perfil de estudios o de trabajador".

Y en segundo lugar, se encuentran aquellos ficheros a los que les corresponde el nivel de seguridad alto. Estos ficheros son aquellos que contienen datos que se denominan especialmente protegidos, recogidos en el artículo 7 de la LOPD. Se trata de datos especialmente

Nivel básico	Nivel medio	Nivel alto
<ul style="list-style-type: none"> <li>● Cualquier fichero con datos de carácter personal.</li> <li>Fichero con datos de ideología, afiliación sindical, religión, creencias, salud, origen racial o vida sexual cuando             <ol style="list-style-type: none"> <li>1. Los datos se utilicen con la única finalidad de realizar una transferencia dineraria a las entidades de las que los afectados sean asociados o miembros.</li> <li>2. Se trate de ficheros o tratamientos en los que de forma incidental o accesorio se contengan aquellos datos sin guardar relación con su finalidad.</li> </ol> </li> <li>● Ficheros o tratamientos que contengan datos relativos a la salud, referentes exclusivamente al grado de discapacidad o la simple declaración de la condición de discapacidad o invalidez del afectado, con motivo del cumplimiento de deberes públicos.</li> </ul>	<p>Ficheros relativos a la comisión de infracciones administrativas o penales.</p> <p>Ficheros de prestación de servicios de solvencia patrimonial y crédito.</p> <p>Ficheros de las Administraciones tributarias relacionados con el ejercicio de sus potestades tributarias.</p> <p>Ficheros de las entidades financieras para las finalidades relacionadas con la prestación de servicios financieros.</p> <p>Ficheros de las Entidades Gestoras y Servicios Comunes de la Seguridad Social relacionados con el ejercicio de sus competencias.</p> <p>Ficheros de las mutuas de accidentes de trabajo y enfermedades profesionales de la Seguridad Social relacionados con el ejercicio de sus competencias.</p> <p>Ficheros que ofrezcan una definición de las características o de la personalidad de los ciudadanos y que permitan evaluar determinados aspectos de la personalidad o del comportamiento de los mismos.</p> <p>Ficheros de los operadores de comunicaciones electrónicas, respecto de los datos de tráfico y localización.</p>	<p>Ficheros con datos sobre ideología, afiliación sindical, religión, creencias, origen racial, salud o vida sexual, cuando no proceda adoptar el nivel básico.</p> <p>Ficheros con datos recabados para fines policiales sin consentimiento de los afectados.</p> <p>Ficheros con datos derivados de actos de violencia de género.</p>

Tabla 1. Niveles de seguridad de los ficheros con datos de carácter personal.

sensibles, sobre los cuales hay que extremar el control pues su tratamiento no autorizado podría suponer perjuicios importantes a los afectados. Estos datos son los relativos a la ideología, afiliación sindical, religión, creencias, origen racial, salud o vida sexual. Por tanto, en principio, cualquier fichero que incluyera datos de estas categorías debería considerarse un fichero de nivel alto. Aquí el problema suele aparecer cuando no está claro si un dato determinado se considera o no perteneciente a alguna de las categorías que lo convierten en dato especialmente protegido. En particular, los datos relativos a la salud suelen ofrecer bastante confusión, motivo por el cual el RDLLOPD incluye una definición específica de datos de carácter personal relacionados con la salud, configurándolos como "las informaciones concernientes a la salud pasada, presente y futura, física o mental, de un individuo. En particular, se consideran datos relacionados con la salud de las personas los referidos a su porcentaje de discapacidad y a su información genética". Aún así, son numerosos los informes jurídicos que la AEPD ha tenido que emitir para aclarar situaciones relacionadas con este tipo de datos, como por ej. el Informe Jurídico 0445/2009, que viene a establecer que los test psicotécnicos suponen el tratamiento de datos psicológicos, los cuales quedan encuadrados dentro de la categoría de datos de salud, o como el Informe Jurídico 0129/2005, que concluye que el mero hecho de ser o no fumador no puede considerarse dato relativo

a la salud, pues "debería considerarse dato directamente vinculado con la salud aquel que reflejase, en relación con las sustancias estupefactivas en general, su mero consumo. Sin embargo, en el caso del consumo de alcohol o tabaco el dato referido al mero consumo, sin especificación de la cantidad consumida, no sería en principio un dato vinculado con la salud, revistiendo tal naturaleza el dato que reflejase la cantidad consumida, en caso de que el mismo significase un consumo abusivo".

No son los datos de salud los únicos que presentan esta problemática, aunque sí los que la presentan más frecuentemente. Sin embargo, otros tipos de datos como la profesión pueden convertirse en datos especialmente protegidos en determinadas circunstancias, como la que analiza el Informe Jurídico 0044/2004 de la AEPD en la que un fichero que contiene el dato de profesión es susceptible de revelar, en algún momento, las creencias de una persona, como ocurriría en el caso de que alguien desempeñara la profesión de sacerdote.

El citado informe, concluye que "en caso de que los datos (...) puedan revelar la ideología, afiliación sindical, religión o creencias de los afectados, los mismos tendrán en todo caso la condición de especialmente protegidos (...) Por todo ello, ha de concluirse que será precisa la implantación sobre los ficheros que contengan datos profesionales de los

afectados entre los que pueda encontrarse el de sacerdote de las medidas de seguridad de nivel alto".

La interpretación del carácter de dato especialmente protegido puede resultar muy controvertida en algunas ocasiones, como la derivada del Informe Jurídico 0524/2009 de la AEPD, en el que se analiza el nivel de seguridad que procede aplicar a los ficheros que almacenan datos relacionados con la actividad de asesoramiento fiscal. Concretamente, establece que la decisión de un contribuyente de efectuar un aporte a la Iglesia Católica en su declaración de IRPF no revela necesariamente sus creencias, dado que "la contribución puede traer causa de las relaciones personales o familiares del sujeto o de su conocimiento de la obra llevada a cabo por la Iglesia, denotando la marcación de la casilla únicamente la preferencia del contribuyente por una determinada obra social, sin que ello implique necesariamente que aquél profesa unas determinadas creencias". Aunque la conclusión más llamativa del informe es la relacionada con el dato que revela el matrimonio entre personas del mismo sexo, afirmando que "si hasta la Ley 13/2005, de 1 de julio, por la que se modifica el Código Civil en materia de derecho a contraer matrimonio, los hechos relativos al matrimonio inscritos en el Registro Civil no se consideraban especialmente protegidos, la misma conclusión debe mantenerse con posterioridad a aquella. En consecuencia el tratamiento de este dato

*no requiere la adopción de medidas de seguridad de nivel alto".*

Además de los datos especialmente protegidos, existen otros tipos de datos que también obligan a elevar el nivel de seguridad a alto para el fichero que los contenga, como son los datos recabados por las Fuerzas y Cuerpos de Seguridad para fines policiales sin consentimiento de los afectados, o aquéllos derivados de actos relacionados con la violencia de género, por el evidente riesgo que para la seguridad de las víctimas supondría su revelación no autorizada.

No obstante, el RDLOPD ha venido a rebajar el nivel de seguridad para los ficheros que contengan ciertos datos en determinadas circunstancias, con objeto de simplificar la gestión de la seguridad en determinados ficheros muy habituales en las organizaciones. Así, los ficheros que contengan datos especialmente protegidos se considerarán de nivel básico en dos circunstancias:

a) Cuando los datos se utilicen con la única finalidad de realizar una transferencia dineraria a las entidades de las que los afectados sean asociados o miembros. Éste es el caso de los ficheros de nóminas, en el que el dato de afiliación sindical se emplea, únicamente, para deducir al empleado el importe de la cuota sindical y transferir dicho importe al sindicato del que éste forma parte.

b) Cuando se trate de ficheros o tratamientos en los que, de forma incidental o accesorio, se contengan aquellos datos sin guardar relación con su finalidad. Éste es el caso de los ficheros de control de presencia, pues suelen ser ficheros mixtos que, en su parte automatizada, almacenan datos no especialmente protegidos sobre la ausencia de los trabajadores, pero que sin embargo, en la parte no automatizada, incluyen copias de los justificantes de ausencia que, en ocasiones, son partes de asistencia médica en los que se refleja la patología por la que el trabajador ha sido asistido.

Por otra parte, también se considerarán de nivel básico los ficheros que contengan datos relativos a la salud, referentes exclusivamente al grado de discapacidad o la simple declaración de la condición de discapacidad o invalidez del afectado, con motivo del cumplimiento de deberes públicos. Para clarificar este precepto, la AEPD ha dictado, entre otros, el Informe Jurídico 0179/2008, que establece que *"serán únicamente exigibles las medidas de seguridad de nivel básico en aquellos ficheros que contengan uno o varios de los siguientes datos:*

1. La mera indicación del grado o porcentaje de minusvalía del afectado o de los miembros de su unidad familiar a los efectos previstos para el cálculo de las retenciones en la legislación reguladora del IRPF.

2. La indicación del dato "apto" o "no apto" de un trabajador a los efectos previstos en la Ley de Prevención de Riesgos Laborales.

3. Los datos relacionados con las obligaciones impuestas al empresario por la legislación vigente en materia de seguridad social que se limiten a señalar únicamente la existencia o no de enfermedad común, enfermedad profesional o accidente laboral o no laboral, así como la incapacidad laboral del trabajador".

### 3. Implicaciones de las medidas de seguridad

Las medidas de seguridad de implantación obligatoria en los sistemas que traten datos de carácter personal pretenden garantizar la confidencialidad, integridad y disponibilidad de éstos. Su aplicación se realiza en función del nivel de seguridad que corresponde al fichero, pero de forma acumulativa. Esto es, a un fichero de nivel básico corresponderá la aplicación de las medidas de seguridad de nivel básico, pero a un fichero de nivel medio corresponderá la aplicación de las medidas de nivel básico y de nivel medio. Finalmente, a los ficheros de nivel alto les corresponde la aplicación de las medidas de nivel básico, medio y alto. Por otra parte, las medidas de seguridad tienen la condición de mínimos exigibles, por lo que nada impide que se adopten medidas más rigurosas, siempre y cuando se garantice la adopción de las medidas exigidas en el RDLOPD.

Para la efectiva implantación de estas medidas es preciso que los sistemas de información que tratan datos de carácter personal hayan sido diseñados incorporando como requisitos, además de los propios de la aplicación, aquellos que permitirán el cumplimiento de las citadas medidas de seguridad. Si bien no todas las medidas de seguridad son susceptibles de ser aplicadas en el momento del desarrollo del sistema de información, algunas de ellas solo alcanzarán toda su efectividad mediante su consideración en esta fase.

A continuación, se relacionan las medidas cuya incorporación al sistema como requisito durante su desarrollo es necesario, o al menos recomendable. Las medidas se presentan organizadas en base al nivel de seguridad que corresponde.

#### 3.1. Nivel básico

Las medidas de seguridad aplicables a los ficheros de nivel básico se relacionan en los artículos 89 a 94 del RDLOPD. En relación al control de acceso, es necesario tener en cuenta las siguientes consideraciones:

**1. Los usuarios deberán poder acceder únicamente a los recursos necesarios para el desarrollo de sus funciones. Deberán establecerse mecanismos para evitar que un usuario pueda acceder a recursos con derechos dis-**

**tintos de los autorizados.** Será preciso establecer perfiles en la aplicación de forma que cada usuario únicamente tenga acceso a aquella información que le vaya a ser necesaria para el desarrollo de sus funciones. Por ejemplo, los operadores del sistema informático de un hospital, que asignan citas médicas a los pacientes, no necesitan tener acceso a la historia clínica de éstos, a diferencia de los facultativos. Por ello, la aplicación debe estar diseñada de forma que a cada usuario le corresponda un perfil determinado, de manera que sea imposible el acceso a aquellos datos de carácter personal que no tengan relación con el perfil del puesto desempeñado.

**2. Exclusivamente el personal autorizado podrá conceder, alterar o anular el acceso autorizado sobre los recursos.** Deberán existir usuarios con el perfil de administradores que sean los únicos habilitados para modificar los permisos de acceso del resto de los usuarios. Los usuarios de la aplicación que no tengan este perfil no podrán modificar los accesos autorizados de los demás usuarios.

#### **Deberá existir una relación actualizada de usuarios y perfiles de cada uno, así como de sus accesos autorizados.**

Se trata de una imposición de la normativa que, si bien puede ser perfectamente alcanzada al margen del propio sistema informático, facilitaría enormemente su cumplimiento por parte del responsable del fichero que la propia aplicación pudiera generar dicha relación, por lo que su inclusión como requisito es una opción a tener en cuenta.

En lo que respecta a la identificación y autenticación de los usuarios, deben considerarse los siguientes aspectos:

**3. Se establecerá un mecanismo que permita la identificación de forma inequívoca y personalizada de todo aquel usuario que intente acceder al sistema de información y la verificación de que está autorizado.** Las cuentas genéricas están taxativamente prohibidas, dado que posibilitarían el acceso de una multiplicidad de personas al sistema sin que puedan quedar identificadas de forma inequívoca y personalizada. El Informe Jurídico 0021/2009 de la AEPD apunta en este sentido ante la consulta de si es posible la utilización de la misma cuenta por parte de varias personas para el acceso a datos personales en una Administración pública. La plena consecución de esta medida no es responsabilidad exclusiva de la aplicación, puesto que nada impide al administrador de la misma la creación de cuentas cuyas credenciales de acceso sean posteriormente distribuidas a varias personas. Sin embargo, sí es posible realizar un diseño del módulo de gestión de usuarios que minimice en lo posible estas acciones. Por ejemplo, previo a la creación de un nuevo

usuario en el sistema, podría mostrarse un mensaje indicando la prohibición de crear cuentas genéricas. Además, podría incluirse como dato asociado a la cuenta de usuario el nombre y apellidos de la persona que lo utilizará, de forma que fuera requisito indispensable completar esta información para tramitar el alta del nuevo usuario. Mención especial merece la cuenta de administrador del sistema, pues conceptualmente se trata de una cuenta genérica que, además, mantiene los máximos privilegios. Si bien no se tiene constancia de un dictamen jurídico a este respecto, parece evidente que una cuenta con tales características no debería ser permitida, pues a través de la misma se podría acceder a todos los datos del sistema sin que pudiera determinarse qué persona en concreto ha realizado el acceso. Por tanto, y salvo que se pudiera garantizar que en cada instante una y sólo una persona va a contar con las credenciales de administrador, lo recomendable sería no contar con cuentas de administración del sistema, sino con perfiles de administrador que fueran asignados a los usuarios individuales del sistema. De esa forma podrían existir varias cuentas de administrador, garantizando que en cada acceso al sistema el titular de la cuenta queda perfectamente identificado de forma inequívoca.

**4. Las contraseñas deberán ser cambiadas con una periodicidad no superior a un año.** Deberá existir un mecanismo en el sistema que informe al usuario de que su contraseña está próxima a caducar, impidiéndole el acceso una vez que haya transcurrido más de un año sin que la contraseña haya sido cambiada.

**Las contraseñas se almacenarán de forma ininteligible.** El procedimiento de almacenamiento de las contraseñas debe garantizar que no sea posible, de ninguna manera, la recuperación de éstas. Para conseguirlo, el procedimiento más sencillo es el cifrado de las mismas con algún algoritmo que no permita reconstruir la contraseña original, como MD5 o SHA-1. De esta forma se consigue que ningún atacante que consiguiera tener acceso a las contraseñas cifradas pudiera descifrarlas para suplantar posteriormente a los usuarios mediante la utilización de sus credenciales.

**Existirá un procedimiento de asignación, distribución y almacenamiento de contraseñas que garantice su confidencialidad e integridad.** Se deberá garantizar que sólo el usuario conoce su contraseña de acceso al sistema. Para ello, es recomendable que la aplicación ofrezca al usuario la posibilidad de cambiar su contraseña, sin que sea necesaria la intervención de ningún administrador u operador. Más aún, puesto que la primera contraseña de un usuario en el sistema suele ser fijada por el administrador, sería conveniente dotar al sistema

de la posibilidad de forzar al usuario el cambio de contraseña en el primer acceso al sistema.

Finalmente, las medidas de seguridad de nivel básico contemplan actuaciones en materia de copias de respaldo. Sería interesante considerar la siguiente:

**5. Deberá realizarse una copia de respaldo semanal como mínimo, salvo que en dicho periodo no se hubiera producido ninguna actualización de los datos.** Por lo general, para alcanzar esta medida de forma satisfactoria no es necesario plantearla durante el desarrollo del sistema de información. Lo habitual es disponer de una política de copia de seguridad global que se encargue de realizar la salvaguarda de toda la información corporativa. Sin embargo, es conveniente tener en cuenta esta medida para el caso en que se desarrolle un sistema de información para una organización en la que no exista definida una política de copia de seguridad previa. En este caso, podría incorporarse la funcionalidad consistente en la realización de una copia de seguridad semanal de los datos de forma automatizada, liberando así al administrador del sistema de la tarea de realizar la copia de seguridad manualmente cada semana, o de tener que definir y poner en marcha una política de copia de seguridad global para todos los activos de información, lo cual, por otra parte, sería lo recomendable.

### 3.2. Nivel medio

Las medidas de seguridad aplicables a los ficheros de nivel medio se relacionan en los artículos 95 a 100 del RDLOPD. La única consideración aplicable al desarrollo de sistemas de información que traten datos personales almacenados en ficheros de nivel medio está relacionada con la identificación y autenticación de los usuarios, y es la siguiente:

**6. Se establecerá un mecanismo que limite la posibilidad de intentar reiteradamente el acceso no autorizado al sistema de información.** Deberá implementarse un mecanismo que, ante un número prefijado de intentos fallidos de acceso al sistema, bloquee la cuenta de usuario. Para el desbloqueo de la misma, el usuario deberá personarse ante el administrador del sistema, que comprobará su identidad y las causas que originaron el bloqueo.

### 3.3. Nivel alto

Las medidas de seguridad aplicables a los ficheros de nivel alto se relacionan en los artículos 101 a 104 del RDLOPD. El control de acceso en los sistemas de información que tratan datos personales almacenados en ficheros de nivel alto impone la siguiente consideración:

**7. Existirá un registro de acceso que almacenará, como mínimo, la identificación del usuario, la fecha y hora**

**del acceso, el fichero accedido, el tipo de acceso y si ha sido autorizado o denegado. En caso de que el acceso haya sido autorizado, será preciso guardar la información que permita identificar el registro accedido. El periodo mínimo de conservación de los datos registrados será de dos años.**

Se trata de una medida que influye de manera significativa en el desarrollo del sistema de información, pues será preciso crear como estructura en la base de datos el registro de acceso como tal, y diseñar todas las transacciones del sistema que operen sobre datos personales de forma que dejen constancia de la operación realizada en el registro. En el caso de consultas masivas de información, teniendo en cuenta que deben almacenarse los datos que permitan identificar los registros accedidos, habrá que tener en cuenta el posible impacto sobre el rendimiento que pudiera tener esta medida.

En relación a las comunicaciones electrónicas, es preciso tener en cuenta lo siguiente:

**8. La transmisión de datos de carácter personal a través de redes públicas o redes inalámbricas de comunicaciones electrónicas se realizará cifrando dichos datos, o bien utilizando cualquier otro mecanismo que garantice que la información no sea inteligible ni manipulada por terceros.** Esta exigencia viene a imponer el uso de protocolos cifrados (como puede ser HTTPS en el caso de aplicaciones web) para las aplicaciones que transmitan datos a través de redes inseguras. La elección del mecanismo de cifrado no es tema baladí, como se desprende del Informe jurídico 0494/2009 de la AEPD, que establece que *"no sólo es necesario cifrar, sino cifrar de forma que la información no sea inteligible ni manipulada por terceros. Sin esta última condición, no se cumplirá lo estipulado en el citado artículo 104. Esto implica (...) que el sistema de cifrado a emplear no esté comprometido, es decir, que no se conozca forma de romperlo. (...) tanto el cifrado que ofrecen los productos que generan archivos PDF o el realizado por WinZip tienen vulnerabilidades conocidas y se disponen de herramientas de libre distribución que aprovechan dichas vulnerabilidades. Más concretamente, no sólo se pueden obtener en Internet fácilmente utilidades que rompen las protecciones de los archivos PDF o ZIP, sino que el propio algoritmo en el que descansa la cifra de documentos PDF, el algoritmo RC4, es manifiestamente vulnerable. Aunque para el uso particular pudieran considerarse adecuadas, no así para el intercambio de información con las garantías que se precisan en el Reglamento"*.

Finalmente, en relación con la gestión y distribución de soportes, es conveniente considerar las siguientes medidas:

**9. La distribución de los soportes**

que contengan datos de carácter personal se realizará cifrando dichos datos o bien utilizando otro mecanismo que garantice que dicha información no sea accesible o manipulada durante su transporte. Si el sistema de información ofrece la posibilidad de exportar datos para su almacenamiento en soportes, sería conveniente dotarlo de una opción que permita realizar la exportación de datos de forma cifrada, evitando así tener que efectuar un posterior procesamiento para el cifrado de dichos datos. En el proceso de cifrado deberá tenerse en cuenta lo anteriormente indicado por el Informe jurídico 0494/2009 de la AEPD respecto a la seguridad del algoritmo de cifrado escogido. Es más, otorgando a la aplicación la funcionalidad de exportación cifrada de datos se elimina la posibilidad de que el responsable del fichero utilice mecanismos de cifrado inseguros.

**10. Se cifrarán los datos que contengan los dispositivos portátiles cuando éstos se encuentren fuera de las instalaciones que están bajo control del responsable del fichero.** Si se prevé que la aplicación y los datos personales que maneja pudieran ser instalados en un dispositivo portátil, sería una opción a tener en cuenta la creación de una versión de la misma que fuera capaz de trabajar con la base de datos cifrada, de forma que en caso de robo o pérdida del dispositivo no pudiera accederse a los datos. Al igual que el punto anterior, esta medida puede conseguirse fácilmente a posteriori, sin necesidad de incorporarla en el propio sistema de información. No obstante, es recomendable considerar su inclusión para evitar, nuevamente, que el responsable del fichero utilice mecanismos de cifrado inseguros del dispositivo portátil.

#### 4. Otras consideraciones

Si bien las medidas de seguridad condicionan sensiblemente el desarrollo de un sistema de información, hay otros aspectos de la normativa que también tienen una influencia decisiva sobre el proceso de desarrollo. A continuación se analizan los más importantes.

##### 4.1. Recogida de datos mediante formularios web

La LOPD exige, por norma general, el consentimiento inequívoco del afectado para el tratamiento de sus datos de carácter personal, salvo en una serie de excepciones que se encuentran reguladas en el artículo 6.2 de la norma. El consentimiento deberá ser libre (obtenido sin intervención de vicio alguno del consentimiento), específico (referido a una determinada operación de tratamiento y para una finalidad determinada, explícita y legítima del responsable del fichero), informado (el usuario debe conocer, antes de su tratamiento, la existencia y las finalidades para las que se recogen los datos) e inequívoco (debe existir expresamente una acción u omisión

que implique la existencia del consentimiento; no se admite el consentimiento presunto). El artículo 5 de la LOPD es el que establece la información que ha de facilitarse al afectado para que el consentimiento se pueda considerar informado y, por tanto, válido.

Cuando el propio afectado facilita sus datos personales a una aplicación web a través de un formulario deben tenerse en cuenta las consideraciones del Informe jurídico 0093/2008 de la AEPD, que establece lo siguiente: *"En cuanto al consentimiento informado, éste habrá de recabarse de tal forma que resulte imposible la introducción de dato alguno sin que previamente el afectado haya conocido la advertencia que contenga las menciones a las que nos hemos referido, pudiendo servir como prueba del consentimiento la acreditación de que el programa impide introducir los datos sin antes haber aceptado el aviso legal al que hemos hecho referencia. Todo ello tiene por objeto asegurar que el consentimiento de los afectados sea efectivamente específico e inequívoco tal y como exige la Ley"*. Esto es, no será suficiente con mostrar una cláusula al pie del formulario informando al usuario de los preceptos que establece el artículo 5 de la LOPD y solicitando el consentimiento para el tratamiento de sus datos, sino que deberá articularse un mecanismo que impida al usuario la introducción de datos en el formulario hasta que éste haya aceptado expresamente que es consciente de que se van a recoger sus datos, a qué finalidad se van a destinar, los derechos que le amparan, y que consiente en el tratamiento.

##### 4.2. Cancelación de datos

Es práctica habitual por parte de las organizaciones la conservación indefinida de los datos a modo de histórico o para la realización de estudios estadísticos o de mercado. Sin embargo, cuando estos datos son de carácter personal, esta política debe ser revisada, pues el mantenimiento de los mismos se encuentra regulado por la normativa de protección de datos. En particular, el artículo 4.5 de la LOPD establece que *"Los datos de carácter personal serán cancelados cuando hayan dejado de ser necesarios o pertinentes para la finalidad para la cual hubieran sido recabados o registrados"*. El concepto de cancelación se clarifica en el artículo 16.3 de la LOPD, que determina que *"La cancelación dará lugar al bloqueo de los datos, conservándose únicamente a disposición de las Administraciones públicas, Jueces y Tribunales, para la atención de las posibles responsabilidades nacidas del tratamiento, durante el plazo de prescripción de éstas. Cumplido el citado plazo deberá procederse a la supresión"*.

Por tanto, una vez que los datos hayan dejado de ser necesarios para la finalidad que originó su recogida, el sistema debe ofrecer la posibilidad de cancelarlos, lo que implica su blo-

queo. El Informe jurídico 0127/2006 de la AEPD viene a aclarar el significado del término bloqueo, de la siguiente forma: *"en cuanto al modo de llevar a cabo el bloqueo, deberá efectuarse de forma tal que no sea posible el acceso a los datos por parte del personal que tuviera habitualmente tal acceso, por ejemplo, el personal que preste sus servicios en el centro consultante, limitándose el acceso a una persona con la máxima responsabilidad y en virtud de la existencia de un requerimiento judicial o administrativo a tal efecto. De este modo, pese a permanecer el tratamiento de los datos, el acceso a los mismos quedaría enteramente restringido a las personas a las que se ha hecho referencia"*.

Es decir, el sistema debe estar dotado con una funcionalidad que permita la cancelación de datos mediante su bloqueo. De esta forma, únicamente aquellos usuarios que tengan asignado un perfil de máxima responsabilidad deberían poder acceder a los mismos, no siendo posible el acceso para el resto de usuarios que, hasta el momento de la cancelación, podían acceder a los datos sin más restricciones que las impuestas por las medidas de seguridad.

Asimismo, el sistema debe permitir la supresión definitiva de los datos una vez que el plazo de prescripción de las responsabilidades nacidas del tratamiento se haya cumplido. La determinación del citado plazo de prescripción es una cuestión puramente normativa, que queda fuera del ámbito de este trabajo. No obstante, y a modo de ejemplo, el citado Informe jurídico 0127/2006 cita algunos de estos plazos: *"podría considerarse que el bloqueo habrá de efectuarse durante los plazos de prescripción de las acciones derivadas de la relación jurídica que funda el tratamiento, en los términos previstos por la legislación civil o mercantil que resulte de aplicación, así como el plazo de cuatro años de prescripción de las deudas tributarias, en cuanto los datos puedan revestir trascendencia desde el punto de vista tributario (habida cuenta de la obligación de conservación que impone el artículo 111 de la Ley General Tributaria y el plazo legal de prescripción de cuatro años previsto en el artículo 24 de la Ley de Derechos y Garantías de los Contribuyentes) (...) o el plazo de prescripción de tres años, previsto en el artículo 47.1 de la propia Ley Orgánica 15/1999 en relación con las conductas constitutivas de infracción muy grave"*.

En caso de que la organización desee conservar los datos una vez que hayan dejado de ser necesarios para la finalidad que motivó su recogida, y transcurrido el plazo de prescripción de las responsabilidades nacidas del tratamiento, sólo podrá conservarlos previa disociación de los mismos, salvo la excepción prevista en el artículo 9 del RDLOPD, consistiendo el procedimiento de disociación en

"todo tratamiento de datos personales de modo que la información que se obtenga no pueda asociarse a persona identificada o identificable".

**4.3. Pruebas con datos reales**

Una etapa de vital importancia en el desarrollo de software es la que corresponde a la realización de pruebas para verificar el correcto funcionamiento del sistema desarrollado. Los sistemas que gestionan datos de carácter personal no son distintos al resto de los sistemas, por lo que para garantizar la corrección de dichos sistemas antes de su puesta en producción es necesario realizar las pruebas pertinentes.

Sin embargo, la realización de pruebas en estos sistemas se encuentra limitada por el artículo 94.4 del RDLOPD, que establece que "Las pruebas anteriores a la implantación o modificación de los sistemas de información que traten ficheros con datos de carácter personal no se realizarán con datos reales, salvo que se asegure el nivel de seguridad correspondiente al tratamiento realizado (...). Si está previsto realizar pruebas con datos reales deberá haberse realizado una copia de seguridad". Es decir, que en la realización de pruebas deberán emplearse datos ficticios preferentemente. Si las características del sistema desarrollado exigen la realización de pruebas con datos reales, la normativa impone, por una parte, que se asegure el nivel de seguridad correspondiente (acceso a los datos por parte del personal autorizado únicamente, transferencia cifrada de información en redes públicas, etc.), y, por otra parte, la realización de una copia de seguridad previa a la ejecución de las pruebas.

**5. Conclusiones**

La normativa de protección de datos de carácter personal ha venido a introducir una serie de consideraciones en el desarrollo de sistemas software, en forma de requisitos funcionales impuestos al sistema, o bien en forma de limitaciones a los requerimientos solicitados por los usuarios. En este artículo se ha efectuado una revisión de los distintos niveles

de seguridad definidos en la norma, así como las implicaciones que sobre el desarrollo de sistemas de información que gestionan datos de carácter personal tienen las medidas de seguridad impuestas por dicha norma.

Si bien, la categorización del nivel de seguridad que corresponde a cada fichero está en principio bien definida en el RDLOPD, es necesario poner en común los preceptos de tal disposición con las aclaraciones e interpretaciones que la AEPD realiza, contribuyendo así a aclarar algunas dudas que pueden surgir fácilmente cuando se pretende poner en marcha un sistema para la gestión de datos personales.

Por otra parte, ha quedado establecido que para el cumplimiento de las medidas de seguridad impuestas por la normativa es necesario incorporar a los sistemas de información, desde su desarrollo, la funcionalidad que les permita alcanzar el nivel de seguridad al que están obligados. En función del nivel de seguridad, las medidas de seguridad a implantar supondrán un impacto mayor o menor para el sistema. En concreto, las medidas impuestas por el nivel de seguridad básico, consistentes principalmente en medidas para el control de acceso, identificación y autenticación de los usuarios, son, en buena parte, medidas que vienen aplicándose habitualmente en el desarrollo de nuevos sistemas de información, y que suponen un bajo impacto en el rendimiento del sistema. Sin embargo, las medidas impuestas por el nivel de seguridad alto, tales como el registro de accesos o el cifrado de la información en la transferencia a través de redes inseguras, pueden suponer un impacto mayor en el rendimiento, por lo que deben estudiarse con mayor detenimiento. Finalmente, no sólo es necesario tener en cuenta las medidas de seguridad impuestas por el RDLOPD en el desarrollo de sistemas de información. La propia LOPD contiene preceptos que, para ser alcanzados, deben incorporarse a los sistemas durante su desarrollo, como los relativos a la obtención de datos personales desde formularios web o los que afectan a la cancelación de datos.

Sin duda, el conocimiento de la normativa en materia de protección de datos personales es un deber que todo aquel que asuma la responsabilidad sobre el desarrollo de un sistema de información para el tratamiento de datos personales debe tener presente, de cara a conseguir sistemas cuyo funcionamiento se adecue a las exigencias legales que afectan a este tipo de software.

**Referencias**

[1] BOE. Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, <<http://www.boe.es/boe/dias/1999/12/14/pdfs/A43088-43099.pdf>>.

[2] BOE. Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal, <<http://www.boe.es/boe/dias/2008/01/19/pdfs/A04103-04136.pdf>>.

[3] AEPD. Agencia Española de Protección de Datos, <<http://www.agpd.es>>.

**¿Estudiante de Ingeniería Técnica o Ingeniería Superior de Informática?**

**Puedes aprovecharte de las condiciones especiales para hacerte socio estudiante de ATI**

**y gozar de los servicios que te ofrece nuestra asociación, según el acuerdo firmado con la Asociación RITSI**

Infórmate en <[www.ati.es](http://www.ati.es)>  
o ponte en contacto con la Secretaría de ATI Madrid




Jaime Urquiza Fuentes, Francisco J. Almeida Martínez, Antonio Pérez Carrasco

Grupo LITE (Laboratorio de Tecnologías de la Información en la Educación), Universidad Rey Juan Carlos

<jaime.urquiza@urjc.es>, <francisco.almeida@urjc.es>, <antonio.perez.carrasco@urjc.es>

# Reorganización de las prácticas de compiladores para mejorar el aprendizaje de los estudiantes

Este artículo fue seleccionado para su publicación en **Novática** entre las ponencias presentadas a las XVI Jornadas de Enseñanza Universitaria de la Informática (JENU 2010) celebradas en Santiago de Compostela en julio del pasado año y de las que ATI fue entidad colaboradora.

## 1. Introducción

Compiladores es una de las asignaturas más difíciles de las titulaciones de Informática. Aparte de basarse en la teoría de lenguajes formales, cuyo grado de abstracción añade algo de complejidad, la parte práctica suele ser bastante compleja y costosa. Esta práctica suele durar la mayor parte del curso requiriendo cierto grado de planificación y continuidad en el trabajo. Por otro lado ocurre que las herramientas utilizadas en la práctica no tienen una conexión clara con los fundamentos teóricos. En esta comunicación presentamos una reorganización de las prácticas que ayuda a los estudiantes a superar estos problemas.

En la **sección 2** describimos los distintos enfoques que se han utilizado en la enseñanza de la asignatura. A continuación detallamos el contexto educativo de la propuesta. En la **sección 4** especificamos las nuevas prácticas integradas en el curso, su tipología y contenidos. En las **secciones 5 y 6** detallamos los resultados de esta propuesta, y finalmente exponemos nuestras conclusiones.

## 2. Trabajos relacionados

La enseñanza de estas asignaturas se realiza de formas muy variadas. Desde el punto de vista de los contenidos, hay enfoques centrados en la teoría de construcción con una cobertura variada de cada concepto (por ejemplo en cuanto a algoritmos de análisis sintáctico o detalle en la implementación de la tabla de símbolos). Otros enfoques se centran en el diseño de lenguajes de programación [1], o incluso en la ingeniería del software. W.M. Waite [16] describe varios de estos enfoques.

Otro punto importante es la práctica necesaria en estas asignaturas. El enfoque típico es una práctica que dura todo el curso, pero se han planteado distintas alternativas evitando tener que crear un compilador entero, por ejemplo, desarrollar diferentes partes de uno [4] cuya estructura ya está desarrollada, hacer pequeñas prácticas que no tienen porque estar relacionadas entre sí [14] o incluso explorar el comportamiento de un compilador real con una herramienta de depuración [17]. Aún así, la construcción de un compilador

**Resumen:** La parte práctica de asignaturas como *Compiladores o Procesadores de Lenguajes* (las trataremos como la misma en el resto de la comunicación) suele ser bastante costosa, ya que requiere cierto grado de planificación y continuidad en el trabajo de los estudiantes y las herramientas utilizadas no tienen una conexión clara con los fundamentos teóricos. Nuestra propuesta estructura estas sesiones prácticas en tres tipos: las que se encargan de enlazar teoría y práctica, las que introducen a los estudiantes las herramientas de generación de compiladores y la final donde se desarrolla un compilador de cierta complejidad. Con este enfoque hemos mejorado el porcentaje de éxito en la parte práctica hasta un 86%.

**Palabras clave:** *Compiladores, enseñanza de informática, Procesadores de Lenguajes, prácticas.*

completo sigue siendo interesante [7]. Para esta opción también existen distintas variantes, aunque todas siguen un desarrollo incremental similar: análisis léxico, sintáctico, semántico y generación, incluyendo a veces el desarrollo de una máquina virtual. H.L. Sathi [14] disminuye la complejidad del compilador a cambio de no usar herramientas generadoras. A.V. Aho [1] pide a los estudiantes que diseñen su propio lenguaje para el que construirán el compilador. Aycock [3] se centra en que los estudiantes puedan probar sus compiladores con programas complejos y para ellos desarrollan lenguajes especiales. Finalmente, otra variante utiliza lenguajes fuente de ámbitos distintos a la programación [6][8][13][18].

Nosotros hemos decidido implantar las prácticas basadas en el desarrollo de una parte significativa de un compilador a lo largo de todo el curso. El trabajo lo realizan en grupo y así pueden afrontar la complejidad de un lenguaje más realista. Además, estas prácticas aportan una visión completa del compilador, y de la relación entre las distintas etapas.

## 3. Contexto educativo

En esta sección describimos la organización de nuestra asignatura y revisamos el rendimiento de los estudiantes en la práctica del curso completo.

Nuestra asignatura es anual y troncal (todos los estudiantes de la titulación deben aprobarla). Sus contenidos se estructuran de la forma típica. Una introducción general y el análisis léxico y sintáctico en la primera mitad del curso, y la traducción dirigida por sintaxis, el análisis semántico y la generación de código intermedio y final en la segunda.

En nuestra universidad existen dos periodos de evaluación oficiales (junio y septiembre), por lo que la práctica tiene 4 entregas: dos parciales para el analizador léxico y sintáctico y dos finales coincidiendo con los periodos de evaluación oficiales donde se debe completar el resto de la práctica. Antes de realizar ninguna entrega, los estudiantes deben organizarse en grupos como máximo de 3 personas.

Dado que en nuestra titulación no se puede asegurar que los estudiantes posean conocimientos suficientes de lenguaje ensamblador, la práctica no exige la generación de código objeto. En su lugar usamos lenguajes finales para los que el esfuerzo de traducción a realizar sea significativo, por ejemplo lenguajes de medio-bajo nivel con restricciones similares a las existentes en los lenguajes ensambladores (expresiones aritméticas con un máximo de dos operandos, sin bucles).

### 3.1. Retrospectiva de las prácticas

Hemos analizado el rendimiento de los estudiantes en las distintas entregas. En las gráficas correspondientes (**figuras 1, 2 y 3**), las cantidades reflejan el porcentaje de los grupos registrados que realizaron de forma satisfactoria cada entrega: AL-análisis léxico, AS-sintáctico, TDS(J)-resto de la práctica en junio y TDS(S)-resto de la práctica en junio y septiembre acumulados.

No hemos tenido en cuenta el primer año de impartición de la asignatura, curso 2000-01, por la poca cantidad de estudiantes. Tampoco hemos tenido en cuenta los cursos 2004-05 y 2005-06, ya que carecemos de los datos necesarios para el análisis. Así pues esta retrospectiva contempla 5 cursos académicos en dos periodos diferentes, 2001-04 y

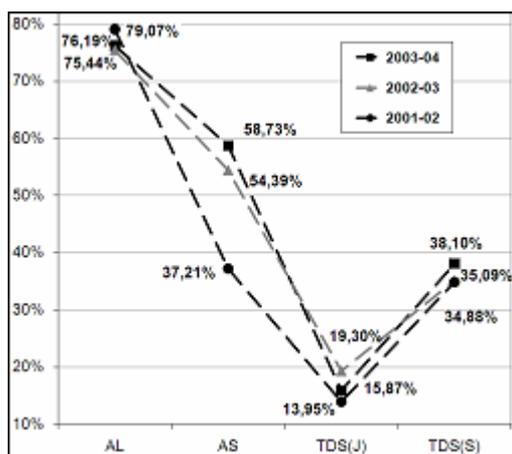


Figura 1. Gráfica de los años 2001-04.

2006-08. Aplicamos la propuesta detallada en esta comunicación en el curso 2008-09.

Durante los 3 primeros años, el enunciado de la práctica mantenía una parte básica para todos y a su vez permitía numerosas variantes usando opciones en cuanto a técnicas de compilación (analizadores sintácticos utilizados y representaciones de código intermedio) y características del lenguaje fuente a compilar: tipos de datos complejos, sentencias de control de flujo. La asignación de las opciones a cada grupo era aleatoria.

Para aprobar la práctica el grupo debía implementar las características que se le había asignado, los analizadores léxico y sintáctico junto con al menos un tipo de datos y una sentencia de control de flujo. Los incrementos de la calificación se obtenían a través de la calidad de la memoria y la ampliación de las características implementadas.

El lenguaje fuente que utilizamos fue cambiando, el primer año en que se impartió la asignatura, 2000-01, utilizamos *Java*. Después de observar los problemas que tuvieron los estudiantes decidimos cambiarlo. Así, durante los tres primeros años del periodo analizado, 2001-04, los lenguajes fuentes fueron: un lenguaje de consultas a bases de datos estilo SQL que había que traducir a operaciones de álgebra relacional, y dos variantes léxico-sintácticas del lenguaje C con restricciones semánticas como la eliminación de operaciones de punteros.

Como se puede ver en la figura 1, en los tres años la cantidad de grupos que completó con éxito la entrega del analizador léxico fue bastante importante, una media de 76,9%.

La cantidad de grupos que abandonaron la práctica después de esa entrega fue muy alta, especialmente el primer año donde entregaron el analizador sintáctico sólo un 37,21%. En los otros dos años estuvo alrededor del 56%.

Aún así la cantidad de grupos que continuaban con la práctica siguió disminuyendo, terminaron con éxito la entrega de junio una media del 16,37%. A final de curso, terminó la práctica una media del 36,02% de los grupos registrados.

La tasa de abandono fue demasiado alta. Entre las principales causas encontramos una típica, el proyecto de compiladores es el primero de larga duración que afrontan los estudiantes de nuestra titulación [14][15]. A pesar de la estructuración en entregas a los estudiantes les costaba establecer objetivos claros. Además, el hecho de trabajar con enunciados distintos les impedía compartir ideas sobre soluciones a problemas que pudieran encontrar.

En el siguiente periodo analizado tratamos de solventar estos problemas. En primer lugar permitimos la selección voluntaria de las opciones de compilación y características del lenguaje, e introdujimos la defensa presencial para evitar el plagio entre grupos. En segundo lugar definimos unos criterios totalmente detallados que permitían establecer objetivos

claros para alcanzar las calificaciones de aprobado, notable o sobresaliente.

Esta vez los lenguajes fuente eran orientados a objetos. El primer año se pidió traducir un subconjunto significativo de *Java*, llamado *SimpleJava*, a un subconjunto de *SmallTalk*, llamado *TinyTalk*. El segundo año se invirtió el enunciado. La complejidad de los lenguajes fuente seguía siendo similar a otros años requiriendo el uso de análisis semántico, tablas de símbolos, ámbitos, etc.

En la figura 2 mostramos los datos de este periodo, 2006-08, junto con el periodo anterior, 2001-04, resumido por sus medias. Aunque en un principio pudiera parecer que empeoramos resultados, ya que los porcentajes de la primera entrega eran menores que en años anteriores, los resultados en la segunda entrega se acercaron al periodo anterior y mejoraron en las entregas de junio y julio aproximadamente en un 20%. Creemos que mejoramos la capacidad de los grupos de valorar sus posibilidades a la hora de terminar con éxito la práctica, ya que la tasa de abandono entre entregas es mucho menor, y en cómputo global la cantidad de grupos que hicieron la primera entrega es prácticamente igual a la de grupos que terminaron la práctica en septiembre.

Sin embargo, el porcentaje de grupos que termina la práctica a final de curso sigue siendo bajo, una media de 58,73% en este periodo. En términos de porcentajes de alumnos que aprueban la práctica con respecto a los matriculados, en el primer periodo tuvimos una media de 33,2% y en el segundo 52,43%. Comparando estas cifras con los porcentajes de estudiantes que aprueban la parte teórica, una media del 85,05%, la diferencia salta a la vista.

A la luz de estos datos creemos que debemos establecer mejor el vínculo entre los conceptos teóricos, que los estudiantes sí dominan,

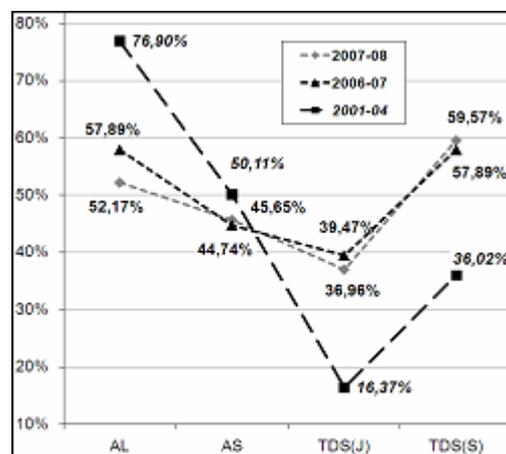


Figura 2. Gráfica de los años 2006-08 junto al periodo 2001-04 resumido por sus medias.

y los prácticos. Para ello proponemos una reorganización de las sesiones prácticas de la asignatura que refuerce ese vínculo. Aplicamos la organización en el curso 2008-09.

### 4. Reorganización de las prácticas

Nuestra propuesta es salvar el salto entre la teoría (algoritmos y estructuras formales) y la práctica (herramientas de generación e implementación de estructuras de datos complejas) con prácticas más sencillas y cercanas entre ambos extremos. Vamos a integrar en la asignatura otros dos tipos de prácticas: básicas y aplicativas, a parte de la práctica final.

Las *prácticas básicas* tienen carácter voluntario y tratan de acercar los conceptos teóricos a la práctica de dos formas diferentes: mostrando el funcionamiento de los algoritmos vistos en clase mediante herramientas de simulación/visualización o relacionando el resultado producido por las herramientas de generación con sus fundamentos teóricos. La duración será de 1 o 2 horas.

El resultado final de este tipo de prácticas es que el estudiante ha visto en funcionamiento los algoritmos, los ha relacionado con su implementación por parte de las herramientas generadoras y ha utilizado éstas una forma básica y tutorizada.

Las *prácticas aplicativas* son voluntarias e incentivadas. Pueden aumentar 1 punto (sobre 10) la nota final de la asignatura, siempre y cuando se apruebe la parte teórica. Estas prácticas tratan de involucrar a los estudiantes en el uso de las herramientas de generación resolviendo problemas algo más complejos. La duración será de 3 a 15 días. El resultado final de este tipo de prácticas es que los estudiantes aprenden cuándo y cómo aplicar las funcionalidades de estas herramientas, lo que les ayudará a afrontar la realización de la práctica final.

A continuación describimos las prácticas agrupadas en las tres entregas de la práctica final: análisis léxico, sintáctico y traducción dirigida por la sintaxis (análisis semántico y generación de código).

#### 4.1. Análisis léxico

Durante la fase de análisis léxico planificamos tres prácticas, dos básicas y una aplicativa. La primera práctica básica es de simulación y trata sobre *autómatas y expresiones regulares*. El objetivo de esta práctica es mostrar el funcionamiento de los fundamentos teóricos del análisis léxico. Usamos la herramienta JFlap [12] para que los estudiantes generen autómatas finitos y expresiones regulares y experimenten con el proceso de reconocimiento.

La segunda práctica básica es sobre herramientas generadoras y trata sobre la *herramienta JFlex* [10]. Su objetivo es introducir

a los estudiantes en la utilización de los generadores de analizadores léxicos. Lo enfocamos como un tutorial sobre la herramienta con ejercicios simples, así los estudiantes terminan conociendo las distintas posibilidades de JFlex. También se les muestra el código generado identificando las tablas de transición del autómata finito. De esta forma mostramos la aplicación directa de los conceptos teóricos en la práctica.

Finalmente, la práctica aplicativa consiste en implementar un *traductor de código Morse*. El objetivo de esta práctica es que los estudiantes apliquen las posibilidades de traducción de JFlex a un enunciado pseudo-realista. Se les proporciona una descripción del código Morse y se les pide que construyan un traductor a/desde caracteres del alfabeto occidental. El plazo de entrega son dos días, debiendo proporcionar la especificación JFlex del traductor. El peso en la nota extra de prácticas es de un 10%.

Con estas prácticas conseguimos que los estudiantes se ejerciten en el diseño de autómatas y expresiones regulares, vean la conexión con un generador de analizadores léxicos y practiquen con él antes de afrontar la parte del analizador léxico correspondiente a la práctica final.

#### 4.2. Análisis sintáctico

Durante la fase de análisis sintáctico planificamos ocho prácticas, seis básicas y dos aplicativas. Utilizamos la herramienta JFlap [12] para el trabajo con gramáticas y autómatas, así como con analizadores LL(1) y SLR(1). Las herramientas generadoras que usamos eran ANTLR [11] para analizadores descendentes y CUP [9] para analizadores ascendentes.

La primera práctica básica era de simulación y trataba de *autómatas con pila (AP) y gramáticas independientes del contexto (GIC)*. El objetivo de esta práctica es mostrar el funcionamiento de los fundamentos teóricos del análisis sintáctico. Los ejercicios son las típicas descripciones de lenguajes para los que hay que diseñar el AP y la GIC que lo reconozca. En el caso de los APs la comprobación de su funcionamiento es inmediata.

Sin embargo, en el caso de las GICs la herramienta debe construir el analizador correspondiente, en esta práctica utilizamos el analizador con retroceso. JFlap muestra la ejecución de este algoritmo informando sobre la cantidad de nodos creados durante la ejecución, consiguiendo así que los estudiantes sean conscientes de la necesidad de encontrar otros algoritmos de construcción de analizadores sintácticos más eficientes.

Las dos siguientes prácticas básicas tratan de analizadores descendentes, la primera es de

simulación y la segunda de generador. En primer lugar usamos JFlap para experimentar con los conceptos teóricos relacionados con los *analizadores LL(1)*. Esta herramienta permite a los estudiantes explorar de forma activa el proceso de construcción de los conjuntos *cabecera* y *siguientes* de cada símbolo no terminal, así como la tabla de análisis LL(1). La exploración activa consiste en pedir al estudiante que especifique el contenido de los conjuntos y las celdas de la tabla, indicando los errores cometidos y proporcionando las soluciones correctas si así lo necesita el estudiante. Finalmente, JFlap construye el analizador y permite que el estudiante visualice su comportamiento usando sus propias cadenas de entrada.

La *práctica básica de generación* trata sobre ANTLR [11] y su objetivo es doble. Por un lado queremos introducir a los estudiantes en el uso de generadores de analizadores sintácticos descendentes de carácter profesional como ANTLR [11]. Por otro lado los estudiantes podrán ver la aplicación práctica de los conceptos teóricos relacionados. De nuevo utilizamos un enfoque de tutorial sobre la herramienta, construyendo analizadores para lenguajes simples como los usados en la práctica sobre APs y GICs. Finalmente los estudiantes explorarán el código fuente generado por ANTLR identificando conceptos teóricos como las reglas borradoras, los símbolos de anticipación (y su construcción con los conjuntos *cabecera* y *siguientes*), los métodos asociados a no terminales o la detección de terminales en la cadena de entrada.

Para terminar con los analizadores descendentes usamos una *práctica aplicativa sobre ANTLR*. El objetivo de esta práctica es que los estudiantes diseñen gramáticas no triviales teniendo en cuenta las restricciones LL(1). Proponemos a los estudiantes una gramática de expresiones aritméticas sin precedencia alguna en los operadores que deben modificar para que los operadores aritméticos cumplan unas normas de precedencia específicas. Además les pedimos que documenten la solución con visualizaciones que justifiquen la existencia de estas precedencias en los árboles sintácticos. Las herramientas de visualización utilizadas son VAST [2] y ANTLRWorks [4]. En concreto, los estudiantes deben modificar la gramática, producir visualizaciones seleccionando ellos mismos las cadenas de entrada y explicar dichas visualizaciones. Además se les pide generar visualizaciones de ejemplo para las situaciones erróneas LL(1): fallo con el símbolo director y terminal no encontrado en la cadena de entrada. El plazo de entrega fue de tres días, debiendo proporcionar la especificación ANTLR de la gramática con las precedencias implementadas, así como la documentación basada en visualizaciones. El peso en la nota extra de prácticas era de un 20%.

Semana	Práctica	Tipo
2	AFDs y Expresiones reg. JFlap	B
3	Herramienta JFlex	B
4	Traductores con JFlex	A
4	APs y GICs con JFlap	B
8	Analizadores LL(1) con JFlap	B
9	Herramienta ANTLR	B
10	Analizadores LL(1) con ANTLR	A
11	Analizadores SLR(1) con JFlap	B
12	Analizadores LR(1) con CUP	B
13	Herramienta CUP	B
14-15	Recuperación de errores sintácticos en CUP	A
17-20	TDS con ANTLR y CUP	A
23	Análisis semántico	A

**Tabla 1.** Distribución temporal de las prácticas básicas (B) y aplicativas (A).

A continuación pasamos a describir las prácticas relacionadas con los analizadores ascendentes. Tres de ellas básicas, una de simulación y dos de generadores, y otra aplicativa. La práctica básica de simulación trata sobre *analizadores SLR(1) con JFlap*. Al igual que en la práctica básica de simulación LL(1), JFlap permite la exploración activa del proceso de construcción de un analizador SLR(1), desde los cálculos de cierres de ítems, hasta la construcción de las tablas del autómata. El estudiante también podrá visualizar el comportamiento del analizador con sus propias cadenas de entrada.

Las dos *prácticas básicas de generación de analizadores ascendentes* usan la herramienta CUP [9]. Con ambas prácticas introducimos a los estudiantes en el uso de herramientas profesionales para la generación de analizadores sintácticos y les mostramos la aplicación de los conceptos teóricos relativos a analizadores ascendentes. Así, la primera práctica básica, *Analizadores LR(1) con CUP*, muestra en el generador CUP los conceptos vistos en la construcción de analizadores LR: conjuntos primeros y siguientes, producción de ítems, generación de estados del autómata LR, producción de las tablas acción e ir-a y finalmente conflictos reducción-reducción y reducción-desplazamiento. La segunda práctica básica, *Herramienta generadora CUP*, trata sobre el enlace entre los analizadores léxico y sintáctico. Los estudiantes aprenden cómo usar analizadores léxicos construidos con JFlex, así como otras funcionalidades auxiliares que ellos mismos hayan desarrollado e integrarlas en la especificación CUP.

La práctica aplicativa sobre generadores de analizadores ascendentes, *Recuperación de errores con CUP*, como su propio nombre indica trata sobre la recuperación de errores sintácticos. CUP utiliza una estrategia de

recuperación en modo pánico similar a YACC, mediante la inserción de puntos de sincronización en la gramática con el símbolo terminal "error". En esta práctica se pide a los estudiantes que diseñen recuperaciones para tres errores concretos en una gramática simple (5 no terminales con 9 producciones) procesando la mayor cantidad posible de la entrada. Y finalmente, en una gramática más compleja (9 no terminales, NT, con 20 reglas) en la que se pide diseñar la mejor recuperación de errores posible usando el menor número de puntos de sincronización. El plazo de entrega fueron cinco días, debiendo proporcionar las especificaciones CUP con las recuperaciones de errores. El peso en la nota extra de prácticas era de un 20%.

En resumen, hemos conseguido que los estudiantes se ejerciten en el diseño de APs y GICs, vean la conexión con dos generadores de analizadores sintácticos y practiquen con ellos antes de afrontar la parte del analizador sintáctico correspondiente a la práctica final.

**4.3. Traducción dirigida por la sintaxis**  
En la fase de traducción dirigida por la sintaxis (TDS) planificamos una práctica aplicativa, que cubre dos conceptos importantes: *TDS con definiciones dirigidas por la sintaxis con CUP* y *TDS con esquemas de traducción con ANTLR*.

En ambas partes se comienza con un enfoque tutorial explicando cómo añadir acciones semánticas a las producciones y cómo usar atributos asociados a los símbolos no terminales. En la parte dedicada a definiciones dirigidas por la sintaxis con CUP se proporciona a los estudiantes los analizadores léxico y sintáctico correspondiente a una gramática simple de cálculo de expresiones lógicas (4 NTs con 10 reglas). En la parte dedicada a esquemas de traducción con ANTLR se pro-

porciona a los estudiantes la especificación ANTLR léxica y sintáctica de una gramática simple de cálculo de expresiones aritméticas (6 NTs con 12 reglas).

En ambos casos la complejidad no se encuentra en el tamaño de la gramática sino en la selección de atributos a usar y la inserción de las acciones semánticas pertinentes. También se pidió a los estudiantes que documentaran su práctica con animaciones que mostraran el funcionamiento de sus traductores. Finalmente distribuimos estas animaciones para que, además del profesor de la asignatura, los estudiantes evaluaran de forma anónima las animaciones de otros estudiantes. El plazo de entrega fue de 13 días para la entrega de los dos traductores con su documentación y otros 15 para la evaluación de las animaciones asignadas, 6 por cada estudiante. El peso en la nota extra de prácticas era de un 35%.

En resumen, hemos conseguido que los estudiantes se ejerciten en el diseño de traductores dirigidos por la sintaxis con ANTLR y con CUP, y que además hagan un esfuerzo de reflexión sobre las soluciones aportadas al evaluar la documentación de otros estudiantes. Esto les ayudará a la hora de afrontar la parte de traducción dirigida por la sintaxis correspondiente a la práctica final.

**4.4. Análisis semántico**

En la fase de análisis semántico planificamos una práctica aplicativa sobre *comprobación de tipos y manejo de tabla de símbolos*. En esta práctica se pide a los estudiantes que diseñen una acción semántica asociada a una producción que representa una llamada a un método de una clase.

En concreto, los estudiantes deben completar la acción semántica dada para que compruebe que la llamada al método efectuada es correcta. También deben implementar dentro de dicha acción la propagación del resultado de la acción (devolución de tipo), aportando además un mensaje concreto para cada tipo de error que pudiera darse. También deben aportar la interfaz de todas las clases adicionales que se necesiten (por ejemplo la que implemente la tabla de símbolos). Estas interfaces deben contener para cada método un detallado comentario que indique la funcionalidad del método, qué entrada recibe y qué salida devuelve.

El plazo de entrega fueron 7 días y el peso en la nota extra de prácticas era de un 15%.

**4.5. Distribución temporal de las prácticas**

La planificación de estas prácticas se sincronizó en la medida de lo posible con los contenidos de las clases de teoría. Así las prácticas básicas de simulación estaban perfectamente sincronizadas con las sesiones de

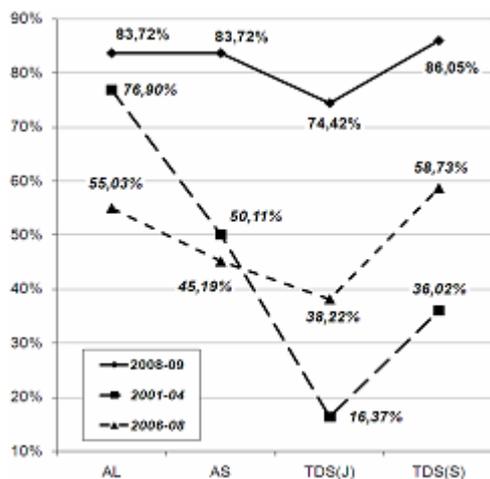


Figura 3. Gráfica del año 2008-09. Los datos de 2001-04 y 2006-08, se han resumido con las medias de cada entrega.

teoría donde se explicaron los algoritmos correspondientes.

En la **tabla 1** se puede ver la planificación temporal de las prácticas básicas y aplicativas. Los huecos existentes en las semanas corresponden a sesiones dedicadas a la práctica obligatoria. Se puede observar que las primeras 16 semanas (primer cuatrimestre) tiene muchas más prácticas que el segundo. Esto se debe a la necesidad de que los alumnos lleguen al segundo cuatrimestre con los conceptos relativos al análisis sintáctico bien asentados.

Las dos primeras entregas de la práctica final consisten en generar especificaciones relativamente sencillas del analizador léxico y sintáctico. Sin embargo, la tercera entrega consistía en el analizador semántico y generador de código. Esta entrega es bastante más compleja y requiere de un esfuerzo de diseño y programación mucho mayor. Por ello decidimos liberar el segundo cuatrimestre de prácticas básicas centrándonos en las aplicativas.

**5. Resultados de los estudiantes**

Analizamos los resultados tras utilizar este enfoque de práctica final a lo largo del curso junto con prácticas pequeñas. Como se puede ver en la **figura 3**, resultados han mejorado significativamente. Más del 83% de los grupos registrados entregaron con éxito el analizador léxico y ninguno abandonó en el sintáctico. La entrega final de junio sí sufrió un poco de abandono, pero al final del curso más de un 86% de los grupos registrados superaron la práctica final con éxito.

En cuanto al porcentaje de estudiantes que han aprobado la práctica, representa un 72,27% de los matriculados.

**6. Conclusiones**

Hemos presentado nuestra experiencia con la

asignatura de Procesadores de Lenguajes. Uno de los puntos importantes son las prácticas. Elegimos la práctica larga porque permite a los estudiantes afrontar un problema más realista y tomar conciencia de las relaciones existentes entre las distintas fases de análisis y síntesis.

Al principio los resultados no fueron buenos. Tratamos de mejorarlos modificando ciertos aspectos burocráticos de la práctica. Esto ayudó a mejorar la tasa de abandono en casi un 22%. Aún así, en junio los grupos no pasaban del 38,22% y en septiembre del 58,7%.

Llegamos a la conclusión de que la práctica única puede desconectar la teoría y la práctica en esta asignatura. Por ello la reforzamos con prácticas pequeñas de dos tipos, las que trabajan los conceptos teóricos y las que enganchan la teoría con los generadores de analizadores.

El efecto ha sido claramente positivo. Hemos conseguido anular la tasa de abandono entre el léxico y el sintáctico, en junio más de un 74% de grupos registrados completaron la práctica con éxito y hasta septiembre más de un 86%. Lo que en términos de estudiantes, significa un 72% de los matriculados.

**Agradecimientos**

Este trabajo se ha financiado con el proyecto TIN2008-04103 del Ministerio de Ciencia y Tecnología del Reino de España.

**Referencias**

[1] A.V. Aho. *Teaching the compilers course*. SIGCSE Bulletin 40, 4 (Nov. 2008), pp. 6-8.  
 [2] F.J. Almeida-Martínez, J. Urquiza-Fuentes, J.A. Velázquez-Iturbide. *Visualization of Syntax Trees for Language Processing Courses*. Journal of Universal Computer Science, 15(7) (Abr. 2009), pp. 1546-1561.  
 [3] J. Aycock. *The ART of compiler construction projects*. SIGPLAN Not. 38, 12 (Dec. 2003), pp. 28-32.  
 [4] D. Baldwin. *A compiler for teaching about compilers*. SIGCSE Bull. 35, 1 (Jan. 2003), pp. 220-223.  
 [5] J. Bovet. *ANTLRWorks: The ANTLR GUI Development Environment*, <<http://www.antlr.org/works>>, 2010.  
 [6] S. Debray. *Making compiler design relevant for students who will (most likely) never design a compiler*. SIGCSE Bull. 34, 1 (Mar. 2002), pp. 341-345.  
 [7] A. Demaille. *Making compiler construction projects relevant to core curriculums*. SIGCSE Bull. 37, 3 (Sep. 2005), pp. 266-270.  
 [8] T. Henry. *Teaching compiler construction using a domain specific language*. SIGCSE Bull. 37, 1 (Feb. 2005), pp. 7-11.  
 [9] S. Hudson. *LALR Parser Generator in Java*, <<http://www.cs.tum.edu/projects/cup>>, 2010.  
 [10] G. Klein, S. Rowe, R. Décamps. *JFlex - The Fast Scanner Generator for Java*, <<http://jflex.de>>, 2010.  
 [11] T. Parr. *ANTLR Parser Generator*, <<http://www.antlr.org>>, 2010.  
 [12] S.H. Rodger. *Learning automata and formal languages interactively with JFLAP*. SIGCSE Bull. 38, 3 (Sep. 2006), p. 360.  
 [13] M. Ruckert. *Teaching compiler construction and language design: making the case for unusual compiler projects with postscript as the target language*. SIGCSE Bull. 39, 1 (Mar. 2007), pp. 435-439.  
 [14] H.L. Sathi. *A project-based course in compiler construction*. SIGCSE Bull. 18, 1 (Feb. 1986), pp. 114-119.  
 [15] H.D. Shapiro, M.D. Mickunas. *A new approach to teaching a first course in compiler construction*. SIGCSE Bull. 8, 1 (Feb. 1976), pp. 158-166.  
 [16] W.M. Waite. *The compiler course in today's curriculum: three strategies*. SIGCSE Bull. 38, 1 (Mar. 2006), pp. 87-91.  
 [17] E. White, R. Sen, N. Stewart. *Hide and show: using real compiler for teaching*. SIGCSE Bull. 37, 1 (Feb. 2005), pp. 12-16.  
 [18] L. Xu, F.G. Martin. *Chirp on crickets: teaching compilers using an embedded robot controller*. SIGCSE Bull. 38, 1 (Mar. 2006), pp. 82-86.

Rubén Casado Tejedor<sup>1</sup>,  
Javier Tuya González<sup>1</sup>,  
Muhammad Younas<sup>2</sup>

<sup>1</sup>Departamento de Informática, Universidad de Oviedo, <sup>2</sup>Department of Computing and Electronics, Oxford Brookes University, Oxford (Reino Unido)

<rcasado@lsi.uniovi.es>,  
<tuya@lsi.uniovi.es>,  
<m.younas@brookes.ac.uk>

# Especificación y prueba de requisitos de recuperabilidad en transacciones WS-BusinessActivity

Este artículo fue seleccionado para su publicación en *Novática* entre las ponencias presentadas a las VI Jornadas Científico-Técnicas en Servicios Web y SOA (JSWEB2010) celebradas en Valencia y de las que ATI fue entidad colaboradora

## 1. Introducción

Las transacciones son un concepto fundamental para conseguir fiabilidad en aplicaciones distribuidas. Una transacción es un mecanismo que asegura que los participantes en un proceso finalizan de manera consensuada. Tradicionalmente las transacciones siguieron el modelo ACID: Atomicidad, Consistencia, Aislamiento y Durabilidad.

En los Servicios Web (SW), las transacciones son complejas, involucran múltiples participantes, atraviesan diferentes dominios y pueden tener una larga duración. El estricto modelo ACID no es apropiado para un entorno de servicios autónomos e independientes dado que la naturaleza del proceso, (principalmente su incrementada duración) hace inviable el bloqueo de recursos imposibilitando las acciones de *rollback*. Este tipo de transacciones con nuevas características son denominadas de *larga duración o larga ejecución* [14] diferenciándolas de las anteriores, denominadas *transacciones atómicas*. La comunidad investigadora ha propuesto nuevos modelos de gestión de transacciones que han sido adaptados para los SW. Estos modelos principalmente relajan las *políticas de atomicidad y aislamiento* de manera que los resultados intermedios de las transacciones activas son visibles para otras [24].

Se han propuesto diferentes estándares para su gestión en SW. *Business Transaction Protocol* [5], aun pudiendo ser utilizado en SW, no fue originariamente definido para ello. *Web Services Composite Application Framework* [21] es un estándar para la composición de aplicaciones conteniendo un protocolo para las transacciones. *Web Services Business Process Execution Language (WS-BPEL)* [23] es un lenguaje ejecutable para la especificación de interacciones entre SW que incorpora un modelo básico para la gestión de transacciones. *WS-Coordination* (WS-COOR) [22], *WS-AtomicTransactions* (WS-AT) [19] y *WS-BusinessActivity* (WS-BA) [20] son un conjunto de protocolos para la coordinación de transacciones atómicas y de larga duración. Como indica Curbera [11], WS-COOR, WS-BA y WS-AT complementan WS-BPEL proveyéndole de un verdadero mecanismo transaccional para la coordinación de SW.

**Resumen:** Las transacciones son un concepto clave en la fiabilidad de aplicaciones basadas en servicios web, siendo WS-Coordination y WS-BusinessActivity los más recientes y aceptados estándares para su manejo. Este artículo aborda la prueba de transacciones en servicios web, tema al que la investigación actual ha prestado escasa atención. Se define un modelo para la especificación de requisitos funcionales de una transacción según el estándar WS-BusinessActivity, así como se propone la utilización de técnicas basadas en riesgos para la definición de casos de prueba que validen el proceso. Se presenta un caso de estudio para ilustrar el método.

**Palabras clave:** Transacciones, pruebas de software, servicios web, análisis de riesgos

Aunque la literatura presenta diferentes enfoques para la prueba de SW, existe un vacío en la prueba de transacciones [7]. Nuestra investigación aborda la problemática de la prueba de transacciones de larga duración mediante WS-COOR y WS-BA.

En [8] se presentó un marco conceptual para la prueba de este tipo de transacciones. Dentro de ese trabajo, se estudiaron sus características desde un punto de vista de las pruebas de software. Basado en él, se definió un conjunto de propiedades [9] (*Composición, Orden, Visibilidad, Consistencia, Durabilidad, Coordinación*) donde aplicar análisis de riesgos con el objetivo de identificar escenarios de prueba. Se muestra el uso de la técnica *Fault Tree Analysis* (FTA) [18] para definir especificaciones de casos de prueba.

En este artículo extendemos el trabajo anterior mejorando el modelo para la especificación de los requisitos transaccionales así como la adaptación del enfoque basado en riesgos para WS-BA. La utilidad se ilustra con un caso de uso donde los requisitos funcionales se especifican guiados por nuestro modelo y se derivan dos casos de prueba usando los escenarios de riesgos alcanzados en el FTA.

El resto del artículo se organiza así: La **sección 2** resume el funcionamiento de WS-COOR y WS-BA. La **sección 3** presenta el modelo y notación para la especificación de los requisitos transaccionales. En la **sección 4** se ilustra nuestro enfoque con un caso de estudio. Este caso de estudio se usa para mostrar cómo los casos de prueba para WS-BA se derivan de los escenarios de riesgo en la **sección 5**. La **sección 6** compara nuestro trabajo con otras investigaciones. Algunas conclusiones y líneas de trabajo futuro se describen en la **sección 7**.

## 2. WS-Coordination y WS-Business Activity

WS-COO define un protocolo para la distribución del contexto de una transacción a todos los participantes. WS-COO especifica la interfaz del coordinador creando uno nuevo o uniéndose a una transacción ya existente. El coordinador consta de los siguientes servicios:

- Activación: creación de un nuevo contexto.
- Registro: especificación del protocolo a usar.
- Protocolo: tipo de coordinación (WS-BA para transacciones de larga duración y WS-AT para las atómicas).

Cuando un participante inicia una nueva transacción necesita un nuevo contexto que solicita a un coordinador usando el servicio de activación. Posteriormente, el participante notifica al resto reenviando el contexto de coordinación. Cada participante registra su protocolo en un coordinador usando el servicio de registro. Finalmente, el coordinador gestiona la transacción intercambiando mensajes con los participantes mediante el servicio de protocolo.

Tanto WS-AT como WS-BA están contruidos sobre WS-COOR. El propósito de WS-BA es coordinar transacciones de larga duración, compuestas por subtransacciones, utilizando compensaciones. Una vez que una subtransacción finaliza correctamente puede ser deshecha mediante la ejecución de una compensación. Imaginemos una aplicación dedicada a la reserva de un viaje compuesto por reservas de hotel, vuelo y alquiler de coche, donde cada reserva la realiza un servicio independiente. Si el vuelo se anula una vez que el hotel fue reservado, esa reserva necesita ser cancelada. Esto significa que la habitación tiene que aparecer como disponible y el cliente podría tener que pagar una penalización por la cancelación.

**3. Modelo de transacción**

Una transacción de larga duración en SW, denotada por  $wT$ , está estructurada como *saga* [13], un conjunto de subtransacciones cada cual con una acción compensatoria asociada. Si una de las subtransacciones de la secuencia aborta, se ejecutan, en orden inverso, las acciones compensatorias asociadas a todas las subtransacciones completadas. Se ve la importancia de comprobar la *recuperabilidad* de cada subtransacción para asegurar la consistencia del proceso. En esta sección se propone un modelo y notación para la especificación de los requisitos funcionales transaccionales. Este modelo será usado para la definición de casos de prueba.

Una  $wT$  se compone de actividades (subtransacciones) ejecutadas por diferentes SW (participantes) que pueden necesitar una cantidad substancial de tiempo para terminar su acción. Una  $wT$  queda definida por la quintupla  $wT = \langle S, C, \Psi_I, \Psi_E, \Psi_C \rangle$  donde  $S$  es el conjunto de subtransacciones,  $C$  el de acciones compensatorias,  $\Psi_I$  es un conjunto de *estados iniciales*,  $\Psi_E$  es un conjunto de *estados ejecutados* y  $\Psi_C$  es un conjunto de *estados compensados*.

El conjunto  $S = \{s_1, \dots, s_n\}$  define las subtransacciones donde cada  $s_i$  es una transacción atómica u otra  $wT$ . Una subtransacción es reemplazable si existe otro participante que pueda realizar la misma acción. Una subtransacción  $s_j$  es dependiente de otra  $s_i$ , denotado por  $s_j < s_i$  si puede ser ejecutada si y solo si  $s_i$  fue completada.

$S$  tiene un orden de ejecución, denotado por  $O(S)$ . Se denota  $s_i ; s_j$  si la subtransacción  $s_i$

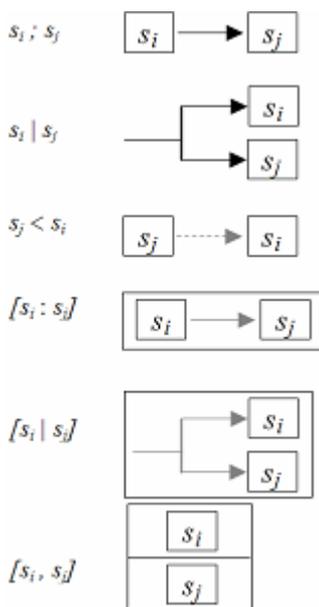


Figura 1. Relaciones entre subtransacciones.

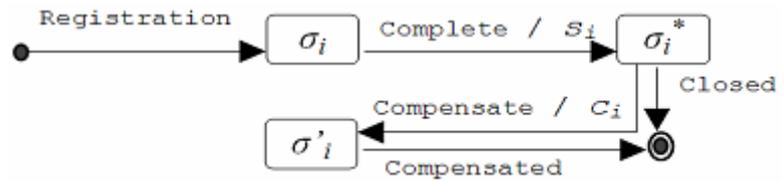


Figura 2. Estados de una subtransacción.

tiene que finalizar antes de poder ejecutar  $s_j$ . Se denota  $s_i / s_j$  si ambas subtransacciones pueden ser ejecutadas en paralelo. Se usa  $[s_i / s_j]$  para especificar subtransacciones reemplazables que pueden ser ejecutadas en paralelo y finalmente escoger cuál de ellas confirmar. Se usa  $[s_i : s_j]$  para especificar que  $s_i$  y  $s_j$  son reemplazables y  $s_j$  puede ser ejecutada si y solo si  $s_i$  falló. Si estrictamente solo una puede ser ejecutada, se denota mediante  $[s_i, s_j]$ . La figura 1 muestra la notación gráfica para estas relaciones.

Toda  $s_i$  tiene una *acción compensatoria* denotada por  $c_i$ . Una  $c_i$  deshace, semánticamente, las acciones realizadas por  $s_i$ , pero no necesariamente retorna al sistema al estado previo a su ejecución. El conjunto de todas las acciones compensatorias se denota por  $C = \{c_1, \dots, c_n\}$ . Toda  $c_i$  puede ser ejecutada si y solo si  $s_i$  ha sido completada. Cualquier  $c_i$  puede ser una acción vacía, denotado por  $\lambda$ .

**3.1. Participantes y coordinador**

Un participante  $p_i$  es el servicio responsable de ejecutar  $s_i$  y la gestión de  $c_i$ .

Una *notificación de transacción* es la comunicación entre dos participantes. Se usa la notación  $i_{wT}[m_1]j_{wT}$  para especificar que el participante  $p_i$  interviene en  $wT$  y envía el mensaje  $m_1$  a otro participante  $p_j$  también involucrado en  $wT$ . Si el participante es el coordinador se denota mediante  $K$ . Se usa  $i_{wT}[m_1]j_{wT} - l_{wT}[m_2]O_{wT} - \dots - v_{wT}[m_n]Z_{wT}$  para denotar una secuencia de notificaciones.

Un *coordinador*  $K$  es el participante encargado de manejar las subtransacciones. Ordena su ejecución, gestiona los fallos y compensaciones y recolecta los resultados de los participantes para dejar el sistema en un estado consistente después de la ejecución de la transacción.

**3.2. Estados en una subtransacción**

Un *estado inicial*  $\sigma_i$  define los requisitos necesarios para que  $p_i$  pueda ejecutar  $s_i$ . El conjunto de estados iniciales se denota  $\Psi_I = \{\sigma_1, \dots, \sigma_n\}$ .

Un *estado ejecutado*  $\sigma_i^*$  define los requisitos que  $p_i$  tiene que satisfacer una vez que ha completado  $s_i$ . El conjunto de todos los estados ejecutados se denota por  $\Psi_E = \{\sigma_1^*, \dots, \sigma_n^*\}$ .

Los requisitos especificados en  $\sigma_i^*$  pueden estar incluidos en los requisitos necesarios para la ejecución de la siguiente subtransacción  $s_{i+1}$  definidos en  $\sigma_{i+1}$ .

El estado compensado  $\sigma_i'$  define los requisitos que  $p_i$  tiene que satisfacer una vez ha ejecutado  $c_i$ . El conjunto de todos los estados compensados se denota por  $\Psi_C = \{\sigma_1', \dots, \sigma_n'\}$ .

La *acción compensatoria*  $c_i$  deshace semánticamente las acciones ejecutadas por  $s_i$ . El sistema cambia hacia  $\sigma_i'$ . Este estado no es necesariamente igual a  $\sigma_i$  dado la imposibilidad de deshacer algunas operaciones (por ej. un *email* enviado). Se usa

$$\sigma_i \xrightarrow{s_i} \sigma_i^*$$

para denotar que, empezando en  $\sigma_i$ , se alcanza  $\sigma_i^*$  después de que la subtransacción se completa y la secuencia de ejecución puede continuar. Se usa

$$\sigma_i^* \xrightarrow{c_i} \sigma_i'$$

para denotar que, empezando en  $\sigma_i^*$ , se alcanza  $\sigma_i'$  tras la ejecución de  $c_i$  y se puede continuar con la secuencia de compensaciones.

La figura 2 ilustra los estados de una subtransacción. Se puede hacer un fácil mapeo entre el modelo de requisitos con el modelo de estados de WS-BA [20]. Nótese que  $\sigma_i$  puede ser mapeado como el estado *Active* ya que es el estado anterior a la ejecución de la subtransacción (mensaje *complete*). De la misma manera,  $\sigma_i'$  está relacionado con el estado *Completed* mientras que  $\sigma_i^*$  es alcanzado a través del mensaje *compensated*.

**4. Caso de estudio**

*Servicio de Estancias (SerEs)* es una aplicación basada en SW dedicada a la gestión de las solicitudes de estudiantes para hacer estancias en otras universidades. El servicio proporciona completa funcionalidad para la asignación de las becas, registro y reserva de alojamiento en la universidad destino así como del pago.

El estudiante envía a la aplicación sus datos personales, fechas para la estancia y una lista ordenada por preferencia de las universidades en las que está interesado. La aplicación envía simultáneamente la solicitud al servicio de

$$O(S_{estancia}) = (S_{beca} \mid [S_{uo\_uni} \mid S_{obu\_uni}]); [(S_{uo\_res} < S_{uo\_uni}), (S_{obu\_res} < S_{obu\_uni})]; [(S_{uo\_inv} < S_{obu\_uni}), (S_{obu\_inv} < S_{obu\_uni})]; S_{pago}$$

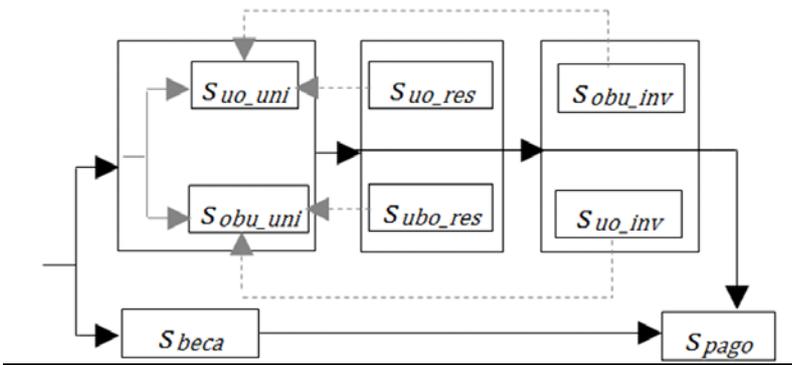


Figura 3. Relación y orden de ejecución de las subtransacciones.

becas así como a todas las universidades (*uni*) propuestas. Después de ser analizada, el servicio de becas responde con los detalles de la subvención que variará según la duración y lugar de la estancia. Obviamente la subvención puede ser denegada. Cuando todas las universidades han respondido, la aplicación selecciona la mejor acorde con las preferencias del estudiante y notifica la decisión al resto de universidades candidatas. Si la subvención es denegada, o todas las universidades candidatas rechazan la solicitud, el proceso queda cancelado y se notifica al estudiante. Una vez que la universidad acepta la propuesta y se asigna la cuantía de la beca, el estudiante es matriculado en la universidad destino. Esta acción se compone de reserva de alojamiento en una residencia universitaria (*res*) y alta en el servicio de investigación (*inv*). El último paso es cargar el pago de la residencia universitaria en la cuenta del estudiante y notificar al usuario todos los datos de la estancia.

En este ejemplo existen múltiples dependencias *inter* e *intra* transaccionales. De ahí la importancia de asegurar la consistencia tanto durante como al final del proceso, tanto si finaliza correctamente o es cancelado (y recuperado mediante acciones compensatorias). Incluso asumiendo que todos los servicios funcionan correctamente, existen escenarios de riesgo debidos a la naturaleza transaccional del proceso. A continuación se presentan algunos:

- Si un estudiante con beca la rechaza y existen otros estudiantes esperando por una, ¿se concede una nueva beca?
- Si varias universidades candidatas aceptan la propuesta, ¿se elige la preferida? ¿Cancelan el resto la propuesta?
- Si después de que todo el proceso esté confirmado el estudiante acorta la duración de la estancia, ¿se reclama la parte correspondiente de la subvención? ¿Cancela la residencia de estudiantes el alojamiento?

#### 4.1. Especificación de requisitos

En este apartado se especifican los requisitos funcionales transaccionales para *SerEs*. Se asume que el estudiante estaría interesado en hacer la estancia en la Universidad de Oviedo (UO) o en Oxford Brookes University (OBU). Esta transacción se define a continuación:

- $wT_{estancia} = S_{estancia}, C_{estancia}, \Psi_{estancia}, \Psi_{Estancia}, \Psi_{Cestancia}$
- $Participantes = \{Beca, UO_{uni}, OBU_{uni}, UO_{res}, OBU_{res}, UO_{inv}, OBU_{inv}, Pago, Coordinador\}$
- $S_{estancia} = \{S_{beca}, S_{uo\_uni}, S_{obu\_uni}, S_{uo\_res}, S_{obu\_res}, S_{uo\_inv}, S_{obu\_inv}, S_{pago}\}$
- $C_{estancia} = \{C_{beca}, C_{uo\_uni}, C_{obu\_uni}, C_{uo\_res}, C_{obu\_res}, C_{uo\_inv}, C_{obu\_inv}, C_{pago}\}$
- $\Psi_{estancia} = \{\sigma_{beca}, \sigma_{uo\_uni}, \sigma_{obu\_uni}, \sigma_{uo\_res}, \sigma_{obu\_res}, \sigma_{uo\_inv}, \sigma_{obu\_inv}, \sigma_{pago}\}$
- $\Psi_{Estancia} = \{\sigma^*_{beca}, \sigma^*_{uo\_uni}, \sigma^*_{obu\_uni}, \sigma^*_{uo\_res}, \sigma^*_{obu\_res}, \sigma^*_{uo\_inv}, \sigma^*_{obu\_inv}, \sigma^*_{pago}\}$
- $\Psi_{Cestancia} = \{\sigma'_{beca}, \sigma'_{uo\_uni}, \sigma'_{obu\_uni}, \sigma'_{uo\_res}, \sigma'_{obu\_res}, \sigma'_{uo\_inv}, \sigma'_{obu\_inv}, \sigma'_{pago}\}$

##### 4.1.1. Subtransacciones

En *SerEs* existen requisitos transaccionales que deben ser cumplidos. Por ejemplo, existe dependencia entre la residencia universitaria y el sistema de investigación: ambos deben pertenecer a la misma universidad. También existen relaciones de reemplazo ya que tanto UO como OBU pueden aceptar una propuesta. Estas relaciones se muestran en la figura 3.

$S_{beca}$ : Recibe los datos del estudiante, duración de la estancia y los detalles de la universidad destino. Si la beca es concedida, contacta con *Pago* para transferir el dinero a la cuenta del estudiante. Responde con el resultado de la solicitud.

$S_{uo\_uni}, S_{obu\_uni}$ : Reciben la propuesta del estudiante y responden si es aceptada o no.

$S_{uo\_res}, S_{obu\_res}$ : Reciben la aceptación de la universidad, fechas de reserva y los datos del estudiante. Se reserva una habitación si hay disponible. Notifica a *Pago* la cuantía. Responde con los detalles de la reserva.

$S_{uo\_inv}, S_{obu\_inv}$ : Reciben los datos del estudiante y fechas de la estancia. Registra al estudiante en el sistema y responde con el resultado.

$S_{pago}$ : Recibe los detalles de la cuenta y la cuantía ingresada o adeudada. Responde con el resultado.

##### 4.1.2. Acciones compensatorias

$c_{beca}$ : Cancela la beca otorgada e intenta concedérsela a otro estudiante.

$c_{uo\_uni}, c_{obu\_uni}$ : Cancela la aceptación del estudiante.

$c_{uo\_res}, c_{obu\_res}$ : Cancela la reserva y notifica a *Pago*.

$c_{uo\_inv}, c_{obu\_inv}$ : Da de baja al estudiante del sistema.

$c_{pago}$ : *Pago* cancela la operación.

##### 4.1.3. Estados iniciales

$\sigma_{beca}$ : El servicio debe haber recibido los datos del estudiante, duración de la estancia y la información de la universidad destino.

$\sigma_{uo\_uni}, \sigma_{obu\_uni}$ : Debe haber recibido la propuesta del estudiante.

$\sigma_{uo\_res}, \sigma_{obu\_res}$ : Debe haber recibido los datos personales del estudiante, fechas de la reserva y la aceptación de la universidad.

$\sigma_{uo\_inv}, \sigma_{obu\_inv}$ : Debe haber recibido los datos personales del estudiante, fechas de la estancia y la aceptación de la universidad.

$\sigma^*_{pago}$ : La cuantía de la beca debe haber sido depositada en la cuenta del estudiante. El pago de la residencia de estudiantes ha sido cargado en la misma cuenta.

##### 4.1.4. Estados ejecutados

$\sigma^*_{beca}$ : El servicio debe haber notificado la decisión. Si fue concedida, ha notificado *Pago* para la transferencia del dinero.

$\sigma^*_{uo\_uni}, \sigma^*_{obu\_uni}$ : Debe haber recibido la propuesta del estudiante. Se ha almacenado en el sistema y se ha respondido con la decisión tomada.

$\sigma^*_{uo\_res}, \sigma^*_{obu\_res}$ : Debe haber notificado el resultado sobre la reserva. Si había habitaciones disponibles, se registra la reserva acorde con las fechas recibidas.

$\sigma^*_{uo\_inv}, \sigma^*_{obu\_inv}$ : El estudiante debe haber sido dado de alta en el sistema. Se ha creado un perfil y una cuenta de correo usando sus datos personales. Ha notificado el resultado.

$\sigma^*_{pago}$ : La cuantía de la beca ha sido ingresada en la cuenta del estudiante. El pago de la reserva del alojamiento ha sido cargado en la misma cuenta.

##### 4.1.5. Estados compensados

$\sigma'_{beca}$ : La beca debe haber sido registrada como cancelada en el sistema de becas. Si existe algún estudiante esperando por una y el plazo está aun abierto, la aplicación ha comenzado un nuevo proceso de otorgación de beca.

$\sigma'_{uo\_uni}, \sigma'_{obu\_uni}$ : El estudiante aceptado debe haber sido borrado del sistema. Si había algún estudiante rechazado, se puede reabrir algún proceso para una nueva aceptación.

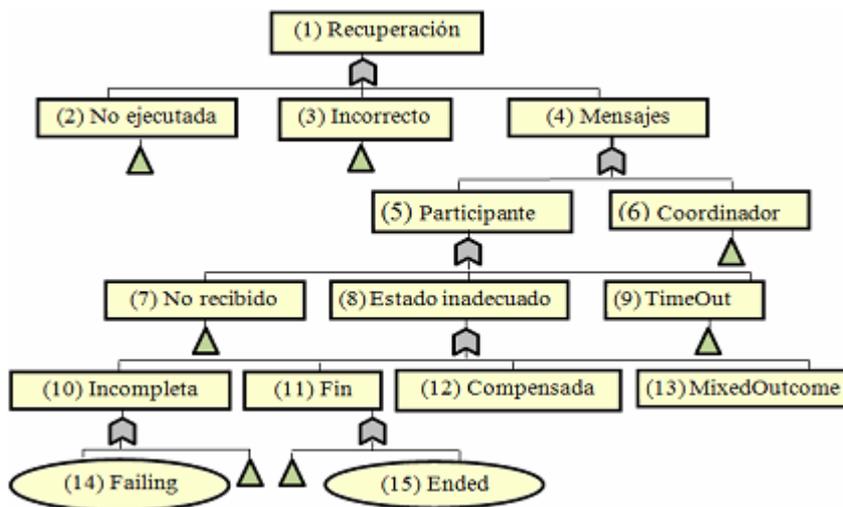


Figura 4. Árbol de fallos.

$\sigma'_{uo\_res}, \sigma'_{obu\_res}$ : La reserva de la habitación debe haber sido cancelada y aparecer como disponible en el sistema. *Pago* debe haber sido notificado.

$\sigma'_{uo\_inv}, \sigma'_{obu\_inv}$ : El usuario debe haber sido dado de baja en el sistema, quedando anulado su perfil y correo.

$\sigma'_{pago}$ : Si el cobro del alojamiento no había sido cargado aun, debe haber sido anulado. En otro caso, el dinero debe haber sido reembolsado. Si el dinero ya había sido ingresado, el servicio debe haber devuelto esa cuantía.

**5. Especificación de casos de prueba para la propiedad de recuperación**

Se define riesgo como un evento indeseado que representa una amenaza para el correcto funcionamiento de un proceso [4]. El análisis de riesgos es una técnica usada para investigar problemas y proponer medios para mitigar sus daños. Aunque originalmente fue usada en industrias como la nuclear o espacial, actualmente también es usada en el desarrollo de software donde la seguridad puede ser muy importante [3]. Nuestro enfoque usa FTA para derivar jerárquicamente escenarios de riesgo durante el ciclo de vida de una *wT*. Cada *propiedad del sistema* es la raíz de un árbol de riesgos que se elabora determinando las posibles causas hasta alcanzar escenarios individuales [9].

La **figura 4** ilustra una parte del árbol de riesgos para la propiedad *Recuperación*. El árbol comienza con el riesgo principal (fallo en el proceso de compensación) que puede ocurrir debido a diferentes causas: la acción compensatoria no se ha ejecutado (2), se ha ejecutado de forma incorrecta (3), o hubo un problema con los mensajes de compensación (4). Para ilustrar este trabajo extendemos la rama derivada del nodo (4) hasta alcanzar nodos hoja que serán usados en los siguientes apartados para definir casos de prueba.

El riesgo (4) puede ser debido a problemas con los mensajes del participante (5) o del coordinador (6). Los problemas con el participante son: que no reciba el mensaje de recuperación (7), lo reciba cuando no debería (8) o que se reciba con el *timeout* cumplido (9). Cuando un participante tiene problemas al recibir los mensajes de compensación puede ser a que lo recibió estando en un estado inadecuado y ejecutando erróneamente la acción compensatoria. Significa que el participante recibe el mensaje sin haber ejecutado su sub-transacción (10), el participante lo recibe y ya ha abandonado su participación en la transacción (11), lo recibe cuando ya ha ejecutado la compensación (12) o lo recibe cuando no es necesario ejecutar ninguna (13).

Una de las posibles situaciones identificadas en (10) es que el participante esté en el estado *failing* (14). Esta situación significa que el participante mientras estaba ejecutando su subtransacción, alcanzó el estado *failing* debido a la aparición de un fallo.

Una posible situación basada en (11) es que el mensaje de compensación se recibe tras la confirmación de que la ejecución de su subtransacción fue correcta, estando, por tanto, en el estado *ended*. El resto de nodos serían desarrollados de la misma manera.

Los nodos hojas representan situaciones de posible riesgo que deberían ser probadas. Siguiendo la derivación jerárquica de las causas se pueden definir fácilmente especificaciones de casos de prueba para aplicaciones que usan WS-BA. En los siguientes apartados se usan los nodos hojas (14) y (15) para definir casos de prueba para *SerEs*.

**5.1. Escenario failing**

Este escenario de riesgo se refiere a un participante que ha recibido el mensaje de compensación cuando está en el estado *failing* (nodo hoja 14). Significa que ocurrió un problema mientras estaba ejecutando su subtransacción, por lo que su acción pudo no haber terminado completamente. Si ejecuta la acción compensatoria ocurriría que algunas características de la subtransacción se desharían cuando realmente no habían sido realizadas.

En el ejemplo de *SerEs* se ve la necesidad de comprobar esta situación. La **figura 5** presenta el caso de prueba acorde a este riesgo para evitar que un nuevo estudiante sea aceptado cuando no es posible. Imaginemos que el servicio *UO\_uni* notifica un fallo mientras estaba estudiando una propuesta pero, indistintamente, no hay plazas disponibles para nuevos estudiantes. El estudiante original no fue aceptado pero debido a la acción compensatoria ejecutada tras la recepción del mensaje, el servicio notifica a un nuevo estudiante su aceptación. Este nuevo estudiante podría rechazar otras universidades creyendo erróneamente que tenía plaza en esta universidad.

**5.2. Escenario ended**

Este escenario se refiere a un participante que ha finalizado correctamente su transacción

Precondiciones	$wT_{estancia}$ está correctamente inicializada.
Secuencia de entrada	$UO\_universidad_{estancia} [register]K - OBU\_universidad_{estancia} [register]K - K[complete] UO\_universidad_{estancia} - UO\_universidad_{beca} [fail]K - K[compensate] UO\_universidad_{estancia}$
Salida esperada	$UO\_universidad$ ignora el mensaje de compensación por lo que no se ejecuta $C_{uo\_uni}$ Por tanto el sistema permanece en $\sigma_{uo\_uni}$ y no se cumple $\sigma^*_{uo\_uni}$ $UO\_universidad$ queda a la espera del mensaje <i>failed</i> .

Figura 5. Caso de prueba para SerEs según el escenario failing.

Precondiciones	$wT_{estancia}$ está correctamente inicializada. Se cumple el estado $\sigma^*_{obu\_inv}$
Secuencia de entrada	$OBU\_inv_{estancia}[\text{register}]K -$ $K[\text{complete}]OBU\_inv_{estancia} -$ $OBU\_inv_{estancia}[\text{completed}]K -$ $K[\text{ended}]OBU\_inv_{estancia} -$ $K[\text{compensate}]OBU\_inv_{estancia}$
Salida esperada	$OBU\_inv_{estancia}$ ignora el mensaje de compensación por lo que no se ejecuta $c_{obu\_inv}$ . Por tanto el sistema permanece en $\sigma^*_{obu\_inv}$ y no se cumple $\sigma'_{obu\_inv}$

Figura 6. Caso de prueba para SerEs según el escenario ended.

(nodo hoja 15). Por tanto recibe el mensaje de compensación cuando está en el estado ended. Podría ocurrir debido a un problema del coordinador, pero el participante debe estar preparado para ignorar este mensaje ya que si lo acepta, la subtransacción sería compensada. La figura 6 muestra un caso de prueba para SerEs siguiendo este escenario.

Este fallo podría causar que una vez la transacción haya finalizado, el servicio  $OBU\_inv$  borre de su sistema al estudiante. Dado que el participante está en el estado ended, el protocolo no permitirá notificar esta acción así que aparentemente la transacción se completó correctamente. Si  $OBU\_inv$  ejecuta la acción compensatoria, se borrará el perfil de usuario perdiendo sus correos electrónicos, documentos de trabajo, etc.

## 6. Trabajos relacionados

La mayoría de trabajos abordan el modelado de transacciones desde una perspectiva de diseño [1][6]. Chalin [10] usa casos de usos para definir los requisitos transaccionales, mientras que Banagala [2] propone usar *Problem Frames Approach*.

A diferencia de esos trabajos, nosotros proponemos un sencillo modelo para la definición de requisitos transaccionales en una transacción WS-BA, facilitando además la derivación de los casos de pruebas.

No se han encontrado trabajos específicos para la prueba de aplicaciones basadas en WS-BA. Algunos esfuerzos existentes están centrados en la verificación formal. Lanotte

[16] desarrolló un modelo para describir transacciones de larga ejecución y la verificación automática de propiedades usando *model checking*. Emmi [12] propone traducir procesos con acciones compensatorias en un árbol automata para verificar la ilusión de atomicidad. Li [17] presenta un modelo para la verificación de requisitos de atomicidad relajada usando *restricciones temporales*.

Este trabajo estudia la viabilidad de un enfoque práctico en contraste con otras investigaciones donde se usan modelos

teóricos para una verificación formal. Nuestro enfoque se basa en la identificación de escenarios de riesgo donde se deben aplicar pruebas dinámicas, siendo el objetivo la especificación de casos de prueba para el estándar WS-BA.

## 7. Conclusiones

Hemos presentado un enfoque práctico para la prueba de la recuperabilidad en transacciones según el estándar WS-BA. Este trabajo propone un sencillo modelo para la especificación de los requisitos funcionales. Los casos de prueba se definen siguiendo los escenarios alcanzados tras un análisis de riesgo usando FTA. La viabilidad de este enfoque se ilustra con un caso de estudio. Dado que las transacciones se basan en servicios independientes, el objetivo de esta línea de investigación es mejorar las pruebas en la fase de integración. Se necesita más investigación para analizar la relación y complementariedad de este enfoque con otras técnicas.

Un trabajo a corto plazo es ejecutar los casos de prueba definidos en una implementación [15] para medir su calidad. Se trabajará también en la mejora del modelo para hacerlo independiente del estándar utilizado. El objetivo es definir un modelo abstracto que incluya la definición de los requisitos así como los riesgos identificados, facilitando la generación de casos de prueba.

## Agradecimientos

Este trabajo fue parcialmente financiado por el Ministerio de Ciencia y Tecnología, España, Programa Nacional I+D+I, proyecto Test4DBS (TIN2010-20057-C03-01) y por la beca BES-2008-004355.

## Referencias

- [1] M. Alrifai, P. Dolog, W. Nejdl. Transactions Concurrency Control in Web Service Environment. *ECOWS*, 2006.
- [2] V. Banagala. Analysis of transaction problems using the problem frames approach. *IWAAPF*, 2006.
- [3] J. Bennett, G. Bohoris, E. Aspinwall, R. Hall. Risk analysis techniques and their application to software development. *EJOR*, 1996, 95, pp. 467-475.
- [4] A. Benoit, M. Patri, S. Rivard. A framework for information technology outsourcing risk management. *ACM SIGMIS Database*, 2005, 36, pp. 9-28.
- [5] OASIS. *Business Transaction Protocol: An OASIS Committee Specification*, Version 1.0, June 2002, <[http://www.oasis-open.org/committees/download.php/1184/2002-06-03.BTP\\_cttee\\_spec\\_1.0.pdf](http://www.oasis-open.org/committees/download.php/1184/2002-06-03.BTP_cttee_spec_1.0.pdf)>.
- [6] M. Butler, C. Ferreira, M. Ng. Precise modeling of transactions and its application to BPEL. *Journal of Universal Computer Science*, 11, 2005.
- [7] G. Canfora, M. Di Pentà. Service-Oriented architectures testing: a survey. *LNCS*, 2009, 5413, pp. 78-105.
- [8] R. Casado, J. Tuya. "Testing transactions in service oriented architectures", *ICWE, DC*, 2009.
- [9] R. Casado, J. Tuya, M. Younas. Testing long-lived web services transactions using a risk-based approach. *QSIC*, 2010.
- [10] P. Chalin, D. Sinning, K. Torkzadeh. Capturing business transaction requirements in use case models. *ACM SAC*, 2008.
- [11] F. Curbera, R. Khalaf, N. Mukhi, S. Tai, S. Weerawarana. The next step in web services. *Communications of the ACM*, 2003, 46, pp. 29-34.
- [12] M. Emmi, R. Majumdar. Verifying compensating transactions. *LNCS*, 2007, 4349, pp. 29-43.
- [13] H. García-Molina, K. Salem. Sagas. *ACM SIGMOD*, 1987, 16, pp. 249-259.
- [14] C. Guidi, R. Lucchi, M. Mazzara. A formal framework for web services coordination. *Electronic Notes in Theoretical Computer Science*, 2007, 180, pp. 55-70.
- [15] JBoss Community. <<http://www.jboss.org/jbosstm>>.
- [16] R. Lanotte, A. Maggiolo-Schettini, P. Milazzo, A. Troina. Design and verification of long-running transactions in a time framework. *Science of Computer Programming*, 2008, 73, pp. 76-94.
- [17] J. Li, H. Zhu, J. He. Specifying and verifying web transactions. *LNCS*, 2008, 5048, pp. 149-168.
- [18] W. Vesley, F. Goldberg, N. Roberts, D. Haasl. Fault tree handbook. *NUREG-0492, U.S. Nuclear Regulatory Commission*, 1981.
- [19] OASIS. *Web Service Atomic Transactions*, <<http://docs.oasis-open.org/ws-tx/wstx-wsat-1.2-spec-os/wstx-wsat-1.2-spec-os.html>>.
- [20] OASIS. *Web Service Business Activity*, <<http://docs.oasis-open.org/ws-tx/wstx-wsba-1.2-spec-os/wstx-wsba-1.2-spec-os.html>>.
- [21] OASIS. *Web Services Composite Application Framework*, <<http://docs.oasis-open.org/ws-caf/ws-context/v1.0/OS/wscctx.html>>.
- [22] OASIS. *Web Service Coordination*, <<http://docs.oasis-open.org/ws-tx/wstx-wscoord-1.2-spec-os/wstx-wscoord-1.2-spec-os.html>>.
- [23] OASIS. *WS-BPEL*, <<http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>>.
- [24] M. Younas, K. Chao, C. Lo, Y. Li. An efficient transaction commit protocol for composite web services. *AINA*, 2006, 1, pp. 591-596.

Las habituales referencias que desde 1999 nos ofrecen los coordinadores de las Secciones Técnicas de nuestra revista pueden consultarse en <<http://www.ati.es/novatica>>.

### Sección Técnica "Acceso y recuperación de información" (José María Gómez Hidalgo, Manuel J. Maña López)

**Tema:** Taller y colección de datos sobre correcciones ortográficas en búsquedas

Muchos estamos ya habituados a ver sugerencias de Google tales como: "Quizá quiso decir...". En realidad, esta sugerencia tan útil no pasa de ser un simple corrector ortográfico de búsquedas, con tecnología algorítmicamente más simple que, por ejemplo, la corrección ortográfica de programas como Open Office Writer o Microsoft Word. Sin embargo, el método está soportado por los datos acumulados de trillones de búsquedas y el análisis del índice de Google. Precisamente Google publicó en 2006 una serie de datos de sus índices, y que utilizan en su traductor automático, en el corrector y en otros temas. Se trata de un listado de N-gramas (secuencias de N palabras) que aparecen en páginas Web, y sus frecuencias de aparición, para distintos idiomas (incluyendo el español). Es un corpus de varios Gigabytes que se descarga directamente de su Web: <<http://googleresearch.blogspot.com/2006/08/all-our-n-gram-are-belong-to-you.html>>.

Microsoft Research ha propuesto recientemente una iniciativa relacionada, en la que proporciona un corpus extraído de la colección *TREC 2008 Million Query Track* preparado para desarrollar y evaluar algoritmos de corrección ortográfica de búsquedas. También se ha organizado una competición de algoritmos y un taller para presentar los resultados de la misma, donde la evaluación final de los algoritmos se realiza sobre búsquedas reales en Bing: <<http://web-gram.research.microsoft.com/spellerchallenge/Default.aspx>>.

**Tema:** Colletta: una herramienta para búsquedas de escritorio basada en etiquetas

En los últimos años, la explosión de información se ha extendido de Internet y los medios hasta alcanzar nuestro propio escritorio. ¿Cuántas veces hemos acabado desistiendo de encontrar un documento importante en una maraña de carpetas locales y correos? Para solucionar este tipo de problemas, raramente aliviados por los buscadores integrados en el escritorio del sistema operativo, han surgido proyectos y sistemas como Copernic <<http://www.copernic.com/>>, o la propia búsqueda de Google Desktop Search <<http://desktop.google.com/>>. Siguiendo esta línea, el proyecto Colletta es un prototipo desarrollado por *Microsoft Research Cambridge* (MSRC) para ayudar a administrar documentos y otros recursos que se utilizan en el puesto de trabajo. Este sistema permite etiquetar archivos de documentos, carpetas, imágenes, correos electrónicos, páginas web y otros objetos. A continuación, se puede acceder a ellos sin tener que navegar por el sistema de archivos, buscar en Internet, o en las carpetas de correo electrónico.

Colleta permite crear etiquetas que hacen referencia a las actividades, personas, lugares, eventos, etc. Puede asociar múltiples etiquetas a los documentos. El sistema se integra con la mayoría de aplicaciones de Microsoft Office. Para otras aplicaciones, sólo hay que arrastrar y soltar un archivo en una etiqueta. Se puede acceder a todas las etiquetas y los recursos etiquetados de la barra del escritorio Colletta o las barras de herramientas Colletta incluidas en Office. <<http://research.microsoft.com/en-us/um/cambridge/projects/research/desktop/projectcolletta/default.htm>>.

### Sección Técnica "Auditoría SITIC" (Marina Touriño Troitiño, Manuel Palao García-Suelto)

**Tema:** Cloud Computing

La divulgación de los procesos de TI bajo un esquema de "cloud computing", o de "computación en la nube" se está realizando vertiginosamente, de forma muy similar a las noticias que aparecen en los foros y chats de Internet. Ya se sabe, todo lo que se publica en Internet y si todo el mundo habla de ello (en este caso el sector de TI), es lo que prima. En definitiva, todos hablamos de "cloud computing", pero ¿sabemos realmente de qué estamos hablando? ¿Cuántos de nosotros tienen experiencia real y prologada en este tema?

Hoy en día no se sabe muy bien cuán extendida está la contratación externa de este servicio (hay entidades que ya tienen "internamente", con sus propios recursos de TI, un esquema similar de trabajo). Tampoco podemos distinguir con precisión si todos los proveedores de estos servicios hablan de lo "mismo", es decir si su oferta implica el mismo alcance, procesos y almacenamiento, y un largo etcétera.

Sin dejar de lado el tema de la seguridad y recuperación de los datos, que hasta ahora se "presuponen", pero que no suelen ser específicos en los contratos de "adhesión" de muchos ofertantes de estos servicios (contratos unilaterales). Y sin olvidarnos una cuestión "colateral", pero no menos importante, la transferencia de datos personales fuera de la UE: ¿dónde se alojan realmente los datos?

En los próximos meses, revisando las posibilidades solo en España, se ofrecen decenas de seminarios, charlas, presentaciones relacionadas con este tema. En su mayoría con una alta participación de proveedores de esta nueva tipología de servicios, de proveedores de seguridad, y de profesionales jurídicos (por cierto en casos de contrataciones online, ¿cuáles serán los tribunales de referencia en caso de reclamaciones, ¿también estarán en la nube?).

En resumen, son muchas las inquietudes con respecto a los servicios de *cloud computing*, y por esa razón la aparición de la revista "Cloud Computing", <[www.revistacloudcomputing.com](http://www.revistacloudcomputing.com)>, es una muy buena noticia. Es una buena noticia por muchas razones, entre las que se encuentran:

- Está dirigida por una persona independiente y con una larga y reconocida experiencia en TI, nacional e internacional, Fernando Piera,
- Su suscripción es gratuita,
- Tiene artículos de fondo o editoriales, y además entrevistas a todo tipo de profesionales, que esperamos en el futuro se expandan no solamente a "proveedores o consultores" de tecnologías relacionadas con esta modalidad de los procesos en TI, sino también a "clientes o usuarios" con experiencia en estos temas, para poder contrastar realidad versus ofertas.

Por lo tanto, hay que dar la bienvenida a la revista que nos puede ayudar mucho a ir desentrañando una serie de dudas que aún están "en la nube".

Fernando Piera dice en uno de sus editoriales: "*Aunque la nube proporciona muchos beneficios para las organizaciones usuarias, también expone los datos y los sistemas a una serie de riesgos. Cualquier organización, antes de entrar a trabajar en la cloud se debe asegurar de que las medidas de seguridad necesarias sean adoptadas.*"

Y la siguiente pregunta es "cómo se van a calcular los beneficios económicos" del uso externalizado de esta tecnología.

Seguiremos atentos a los artículos publicados en esta revista, de fácil acceso, y damos la enhorabuena a su creación.

### Sección Técnica "Derecho y Tecnologías" (Elena Davara Fernández de Marcos)

**Tema:** Memoria de 2010 que resume la actividad de la APDCM

El derecho fundamental a la protección de datos está garantizado y

controlado por la labor de los organismos nacionales y autonómicos creados al efecto. En el caso de Madrid y dependiendo de la tipología de los ficheros, el ciudadano puede acudir a la Agencia Española de Protección de Datos o a la Agencia de Protección de Datos de la Comunidad de Madrid (APDCM). Y precisamente en relación con la actividad de la Agencia madrileña se ha publicado recientemente la Memoria de actividad del pasado año 2010 en la misma.

Por lo que se refiere a algunos datos relevantes desprendidos del contenido de la citada Memoria, cabe destacar que el total de ficheros inscritos activos en la Comunidad de Madrid ha sido de 13.965, un 1,47% más que en 2009, pero un 11% menos que en 2008. Los ficheros registrados activos pueden pertenecer a tres tipos de administración: las Administraciones de la Comunidad de Madrid, las personas Jurídico-Públicas, o las Entidades Locales. En el año 2010, el número de ficheros registrados activos propiedad de los sujetos recién mencionados y en ese mismo orden son: 9.768, 362, 3.835. Con respecto a 2009 estas cifras representan unas variaciones del -1,15%, 6,19%, y 10,67% respectivamente. Resulta particularmente llamativo el increíble incremento de inscripciones en los últimos 5 años por parte de las personas Jurídico-Públicas y de las Entidades Locales: 152 y 236 son los tantos por ciento que han aumentado las inscripciones activas por parte de estos dos sujetos de derecho.

Respecto a las finalidades de los ficheros inscritos destacan sobre el resto la de Procedimientos Administrativos, seguida de la de fines Históricos, Estadísticos, o Científicos, y la de funciones Estadísticas Públicas. Por último y por lo que se refiere a la actitud de los ciudadanos, cabe destacar que durante el pasado año se atendieron 1.800 consultas de ciudadanos, realizadas por teléfono y por vía electrónica, solicitando información de carácter general y sobre el ejercicio de sus derechos en protección de datos, <[http://www.madrid.org/cs/Satellite?c=PAPD\\_Generico\\_FA&cid=1142646717273&language=es&pagename=PortalAPDCM%2FPAPD\\_Generico\\_FA%2FPAPD\\_fichaPublicacion](http://www.madrid.org/cs/Satellite?c=PAPD_Generico_FA&cid=1142646717273&language=es&pagename=PortalAPDCM%2FPAPD_Generico_FA%2FPAPD_fichaPublicacion)>.

#### **Tema:** Nueva regulación del "servicio universal" en España

La ley 2/2011, de 4 de marzo, de Economía Sostenible incorporó numerosas novedades en materia de protección de datos, comercio electrónico y propiedad intelectual, entre otras cuestiones relacionadas con el sector de las Tecnologías de la Información y las Comunicaciones. En este sentido, es de destacar el artículo 52 de la conocida como "Ley Sinde" donde se exigía la concreción de los términos de incorporación de la banda ancha en el servicio universal así como la necesidad de transponer los cambios introducidos en el marco del servicio universal por la Directiva 2009/136/CE. Y es en este punto donde aparece el Real Decreto 726/2011 que modifica el Reglamento sobre las Condiciones para la prestación de servicios de comunicaciones electrónicas, el servicio universal y la protección de los usuarios.

Por lo que se refiere al contenido concreto del Real Decreto, el nuevo Reglamento incluye dentro del concepto de servicio universal, o lo que es lo mismo dentro del "conjunto definido de servicios cuya prestación se garantiza para todos los usuarios finales con independencia de su localización geográfica, con una calidad determinada y a un precio asequible", la banda ancha con conexión a Internet de velocidad de, al menos, 1 Mbit por segundo. Entre las principales novedades del Real Decreto destaca la inclusión como obligación para el operador de ofrecer a todos los usuarios, si así lo desean, una media de velocidad en su conexión de al menos un Mbit por segundo. Asimismo, se incorpora en la nueva normativa la posibilidad de tener contratados dos operadores diferentes para los servicios de conexión a Internet y telefonía.

Por último, comentar que el objetivo de la nueva normativa es que el operador de telecomunicaciones cumpla con la obligación de ofrecer al ciudadano una cobertura de calidad a todos los usuarios finales, con independencia de su localización geográfica y a un precio asequi-

ble. <<http://www.boe.es/boe/dias/2011/05/24/pdfs/BOE-A-2011-9012.pdf>>.

#### **Tema:** España sigue entre los países con mayor nivel de piratería

La protección de datos de carácter personal y el comercio electrónico son cuestiones que ocupan un lugar prioritario dentro del Sector TIC; sin embargo, la propiedad intelectual y la defensa de los derechos de autor es la cuestión que, en los últimos años, ha estado (y aún está) en boca de todos por la cantidad de agentes implicados, la dificultad de regularlo y las recientes leyes de diversos Estados Miembros para reducir (e intentar erradicar de manera definitiva) la piratería de los contenidos protegidos por derechos de autor.

En el caso de España y pese a la publicación de la Ley de Economía Sostenible el pasado mes de marzo, la piratería es una realidad muy arraigada. Así lo acredita un reciente informe elaborado por senadores y miembros de la Cámara de Representantes de EEUU y que coloca a España dentro de los países con más piratería del mundo entre los que destacan Canadá, China, Rusia y Ucrania. Este quinto puesto en el ranking viene avalado por el alto nivel de intercambio de ficheros P2P en nuestro país, tal y como se afirma en el denominado "Caucus Antipiratería Internacional". El hecho de pertenecer a esta lista no acarrea sanción alguna, y en el citado informe se explica que si bien España se encuentra dando los pasos correctos en materia legislativa para combatir la piratería, aún carece de las herramientas necesarias para combatir la enorme piratería que representan las descargas P2P.

Finalmente, los responsables del informe hacen un llamamiento al gobierno español con la finalidad de que se comprometa a luchar contra todos los tipos de piratería que se encuentran en la red, para lo que lanza una propuesta que no consiste sino en identificar a los usuarios más habituales del P2P y modificar la legislación para que los titulares de derechos puedan ejercer acciones civiles contra aquellos internautas que reincidan en la conducta ilícita. <<http://www.lavanguardia.com/internet/20110602/54164682411/ee-uu-situa-a-espana-en-el-top-5-mundial-de-la-descarga-ilegal-de-archivos.html>>.

#### **Tema:** Disponibles los certificados de matrimonio y nacimiento vía electrónica

Hoy en día y desde la publicación de la Ley 11/2007 de manera especial, los ciudadanos reclaman una mayor presencia de las TIC en el conjunto de la Administración, de modo que les permita relacionarse por vía electrónica en lugar de hacer uso de los medios "tradicionales". Buena muestra de ello es que ya se puede afirmar que la Administración electrónica es una realidad en nuestro país que, poco a poco, va adquiriendo más protagonismo a través de la incorporación de las TIC a los diversos trámites y gestiones que los ciudadanos han de hacer con los diversos Ministerios, Ayuntamientos y demás entidades públicas, formando así una cultura del uso de las TIC en la Administración que, además de fomentar el cumplimiento de lo dispuesto por la Ley 11/2007, redundará en beneficio de los ciudadanos por el ahorro de tiempo y papeleo. Si bien son muchos los ejemplos que se pueden destacar, traemos a colación una medida que ha decidido incorporar el Ministerio de Justicia y que afectará a millones de españoles que, por diversos motivos, han de solicitar la expedición de su certificado de matrimonio o nacimiento. Y es que, gracias a esta modernización del Ministerio de Justicia, el Registro Civil podrá expedir, vía telemática, las certificaciones de nacimiento y de matrimonio.

Conviene tener en cuenta que no se trata de la primera iniciativa que se adopta en este sentido puesto que ya desde enero del presente año existe la posibilidad de pedir por este mecanismo los certificados de últimas voluntades, de antecedentes penales y de cobertura de seguros. Además, se puede afirmar que su aceptación ha sido más que positiva ya que en cinco meses se han expedido más de 350.000 certificaciones, con el consecuente ahorro de tiempo y papel.

Por último, destacar que los únicos requisitos para hacer uso de la modalidad electrónica de solicitud es disponer de un ordenador con conexión a Internet y un DNI electrónico, siguiendo los pasos que se indican en el sitio web. <[http://www.elpais.com/articulo/sociedad/certificados/nacimiento/matrimonio/traves/Internet/elpepusoc/20110601elpepusoc\\_13/Tes](http://www.elpais.com/articulo/sociedad/certificados/nacimiento/matrimonio/traves/Internet/elpepusoc/20110601elpepusoc_13/Tes)>.

### Sección Técnica "Entorno Digital Personal" (Diego Gachet Páez, Andrés Marín López)

**Tema:** *Elementos claves en la recuperación de información utilizando dispositivos móviles*

La recuperación de información a través de dispositivos móviles no es más que un subconjunto de la recuperación de información mediante ordenador, aunque las especiales características de los dispositivos móviles hacen que esta tarea en algunas ocasiones sea más avanzada y otras más primitiva que cuando hablamos de un ordenador clásico.

Existen dos conceptos que son clave cuando hablamos de recuperación de información mediante dispositivos móviles, uno es la Adaptación de Contenido y el otro la Atención al Contexto. En un sentido amplio el primer concepto se refiere a presentar en el dispositivo la información de una forma adecuada al usuario y el segundo corresponde a analizar la respuesta que se da al usuario a través del dispositivo. La recuperación de información móvil analiza entonces tanto el contenido como el contexto para extraer información útil y relaciones que la recuperación de información clásica no hace.

Los dispositivos móviles que están en el mercado actualmente disponen de una serie de elementos que los ordenadores tradicionales no tienen, incluyendo la posibilidad de disponer de información de localización, cámaras, dispositivos sensores de movimiento, etc. Todos estos elementos permiten el desarrollo de nuevas formas de recuperación de información que incluyen el tiempo, la localización y la semántica, cada día nos encontramos con más aplicaciones de búsqueda y procesamiento de información, específicas para dispositivos móviles, incluyendo por ejemplo realidad aumentada, geolocalización, etc.

En cuanto a la adaptación de contenido, dado que los dispositivos móviles tienen en general pantallas de reducido tamaño y una potencia de procesamiento limitada, los esfuerzos se dirigen a procesar y presentar de manera eficiente la información en este tipo de dispositivos. En este tema en concreto se están haciendo esfuerzos destacables que tienen que ver con extracción de información basada en contenido indexado, extracción de resúmenes, etc., que permiten vislumbrar en un futuro cercano la aparición de nuevas aplicaciones de búsqueda y recuperación de información para dispositivos móviles más eficientes y personalizadas.

### Sección Técnica "Informática Gráfica" (Miguel Chover Sellés, Roberto Vivó Hernando)

**Tema:** *Gráficos 3D con aceleración hardware para Internet*

La creación de contenidos gráficos para la Web con inclusión de las últimas técnicas en realismo visual será una realidad en los próximos años. Por un lado con el desarrollo de WebGL que mantiene el grupo Khronos y por otro con la nueva API Molehill que va a presentar Adobe en los próximos meses.

La principal ventaja de WebGL es que se usará en el nuevo *canvas* del lenguaje HTML 5. Además se ha adelantado en su aparición, ya que la especificación de la versión 1.0 se presentó en febrero de 2011. La ejecución de contenidos en 3D tampoco necesitará de ningún *plugin* salvo, probablemente, en el Internet Explorer. El soporte será multiplataforma, funcionará en distintos navegadores y también en dispositivos móviles con Android o iOS (incluyendo iPads, etc.).

Por lo que respecta a Molehill, no es una opción despreciable. El *plugin* de Flash está desarrollado para cualquier navegador y es usado de forma masiva en el desarrollo web. La librería se espera que incluso disponga de mayor potencia y robustez que WebGL. La intención es que dependa de DirectX en Windows, de OpenGL 1.3 en Mac OS X y Linux, y de OpenGL ES 2 en las plataformas móviles con Android. La primera versión estará disponible con la versión 11 del Flash Player.

En cualquier caso, las dos nuevas librerías son de bajo nivel y en el contexto de la creación de contenidos será necesario utilizar aplicaciones y herramientas de autor. En este sentido, han aparecido un gran número de motores de juegos que se han desarrollado para usar Molehill como: Alternativa3D, Away3d, CopperCube, Flare3D, Minko, Sophie3D o Yogurt3D. Incluso Unity 3D, uno de los motores de juegos más usado en el desarrollo web, modificará su *plugin* para soportar la librería de Adobe. Lo que nos deparará el futuro es difícil de saber, pero seguro que independientemente de la tecnología que usemos, la Web en 3D será una realidad dando lugar a la aparición de mundos virtuales que explorar, de forma social y con la última tecnología de los videojuegos.

### Sección Técnica "Ingeniería del Software" (Javier Dolado Cosín, Daniel Rodríguez García)

**Tema:** *Libros*

**Peter Smith.** *Software Build Systems: Principles and Experience.* Addison-Wesley Professional, 583 páginas, 2011. Libro que presenta de manera exhaustiva las cuestiones relativas al proceso de desarrollo de "build systems". A quien le suene raro el término "build system" seguro que encuentra más familiaridad con los términos Ant, SCons, CMake, GNU Make y algún otro y nos estamos refiriendo a las maneras de organizar procesos que automáticamente nos van a construir un sistema software ejecutable. Este libro nos acerca de manera muy detallada a estos temas a través de sus 19 capítulos: *Build System Overview, A Make-based build system, the runtime view of a program, Make, Ant, Eclipse, etc.* Recomendable para entender qué es lo que sucede cuando utilizamos "make" o recompilamos una aplicación en nuestro ordenador.

**Nuevos libros sobre CMMI del Software Engineering Institute.** El SEI ha publicado recientemente tres libros que especializan los conceptos ya conocidos de CMM a tres áreas distintas: adquisición de software, servicios y desarrollo. Los tres libros contienen ideas sobre la mejora de procesos que en algunos casos son idénticas, pero en otros casos están orientadas a los distintos objetivos de cada disciplina.

**Brian Gallager, Mike Phillips, Karen Richter, Sandra Shrum.** *CMMI for Acquisition: Guidelines for Improving the Acquisition of Products and Services*, 2nd Edition. Addison-Wesley Professional, 2011. Este libro está orientado a las empresas que trabajan con proveedores de productos o servicios.

**Mary Beth Chrissis, Mike Konrad, Sandra Shrum.** *CMMI for Development. Guidelines for Process Integration and Product Improvement*, 3rd Edition, Addison-Wesley Professional, 2011. Orientado a las empresas que se centran en el desarrollo de productos y servicios.

**Elleen Forrester, Brandon Buteau, Sandra Shrum.** *CMMI for Services. Guidelines for Superior Service*, 2nd Edition. Addison-Wesley Professional, 2011. Orientado a las empresas que se encargan de establecer, gestionar y entregar servicios.

### Sección Técnica: "Lenguajes de Programación" (Oscar Belmonte Fernández, Inmaculada Coma Tatay)

**Tema:** *Nace shema.org*

Que los buscadores sean la página de inicio seleccionada por muchos

usuarios al iniciar su navegador no es casual. En un gran número de ocasiones utilizamos un buscador para encontrar información sobre un término en Internet.

Uno de los desafíos actuales en la web es cómo añadir semántica al contenido de las páginas web para que la información que en ellas aparece pueda ser analizada de forma automática por herramientas informáticas, como por ejemplo, un buscador.

Conscientes de este problema, tres de los grandes proveedores de buscadores en Internet (Google, Yahoo y Microsoft) se han aliado para fomentar el etiquetado semántico en páginas web a través de *shema.org*. La finalidad es proporcionar a los creadores de contenido en Internet un vocabulario con el que poder añadir contenido semántico a las páginas escritas en HTML. Así, este vocabulario ayudará a los motores de búsqueda a clasificar la información que aparece en las páginas web.

#### **Tema:** Comparación entre lenguajes de programación

Aprovechando el Scala Days 2011, Google ha publicado un artículo comparando diversos aspectos de algunos lenguajes de programación. Dos de ellos son lenguajes que ya podemos considerar clásicos como C++ y Java y los otros dos los podemos entrecomillar de "modernos" como Scala y Go. Desde nuestro punto de vista la comparación hay que cogerla con pinzas, sobre todo la relativa a la sección de rendimiento, en la que, según este artículo, Java tiene un tiempo de ejecución casi seis veces mayor que C++.

De todos modos el lenguaje de programación Scala es una aproximación muy interesante al intento de unir los lenguajes de programación funcionales y orientados a objetos, <<http://code.google.com/p/multi-language-bench/>>.

#### **Tema:** Libro

Desde estas líneas, ya hemos comentado la revolución que va a suponer y que ya está suponiendo la aparición de HTML5. Para conocer las nuevas capacidades que este lenguaje de marcas va a poner a disposición de los creadores de páginas web es de gran interés el libro de **Peter Lubbers, Brian Albers y Frank Salim** titulado *Pro HTML5 Programming*.

En este libro se presentan muchas de las APIs que en parte ya están disponibles en los navegadores web actuales, como por ejemplo el API Canvas, que permite el acceso directo al hardware gráfico, con lo que es posible tener integrado OpenGL en un navegador sin necesidad de instalar ningún *plugin*. O el API de Audio y Vídeo, que va a permitir visualizar vídeo sin necesidad de tener instalado el *plugin* de Flash. También hay capítulos dedicados al API de Geolocalización que permitirá ubicar geográficamente un dispositivo aunque éste no disponga de receptor GPS; el API *WebSockets* que permitirá establecer comunicación a través de *Sockets* entre cliente y servidor; y el API de Almacenamiento que permitirá que los navegadores almacenen datos en el dispositivo cliente.

#### **Sección Técnica "Lingüística computacional"** (Xavier Gómez Guinovart, Manuel Palomar)

##### **Tema:** Curso de procesamiento del lenguaje

**Steven Bird, Ewan Klein, Edward Loper.** *Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Sebastopol, 2009, 479 páginas. ISBN 978-0-596-51649-9. Este nuevo manual de O'Reilly nos ofrece un curso práctico y actualizado de procesamiento del lenguaje natural, emanado de la docencia impartida por el área de informática de la Universidad de Pensilvania. La orientación del libro es eminentemente

aplicada, si bien también se explican de forma clara y concisa todos los conceptos teóricos necesarios para poder seguir el curso con provecho, incluso sin partir de un conocimiento previo de la materia.

Los capítulos del libro permiten adquirir unas sólidas competencias en programación lingüística en Python, en los ámbitos de la compilación y procesamiento de corpus textuales, del etiquetado morfológico automático, de la clasificación de textos, de la extracción de información textual, del análisis sintáctico automático y de la comprensión del lenguaje natural. La herramienta didáctica elegida, usada a lo largo de todo el libro en las prácticas de programación, es el paquete de recursos NLTK (*Natural Language Toolkit*), una distribución de código abierto para el procesamiento del lenguaje en Python, que incluye bibliotecas de programación para las tareas más comunes en PLN que pueden ser utilizadas por los estudiantes para escribir sus propios programas. Las tareas propuestas en las prácticas de programación son de gran valor didáctico y el desarrollo escalonado de los contenidos teóricos permite una progresión en la materia bien fundamentada. De este modo, el libro de Bird, Klein y Loper constituye un excelente manual, imprescindible en cualquier curso de lingüística computacional de orientación aplicada que adopte Python como lenguaje de programación, <<http://oreilly.com/catalog/9780596516499/>>.

#### **Sección técnica "Seguridad"** (Javier Areitio Bertolín, Javier López Muñoz)

##### **Tema:** Libros

**E. Bertino, K. Takahashi.** *Identity Management. Concepts, Technologies and Systems*. Artech House. ISBN 1608070398, 2011.

**T. Craig, M.E. Ludloff.** *Privacy and Big Data*. O'Reilly Media. ISBN 1449305008, 2011.

**M.K. Denko, L. Wougang,** *Security in Next Generation Mobile and Wireless Networks: A Collaborative Attacks Perspective*. 1st Edition. Springer. ISBN 1441961623, 2011.

**M. Dowding.** *Privacy: Defending an Illusion*. The Scarecrow Press, Inc. ISBN 0810881020, 2011.

**W.C. Easttom.** *Computer Security Fundamentals*. 2nd Edition. QUE. ISBN 0789748908, 2011.

**M. Masud, L. Khan, B. Thuraisingham.** *Data Mining Tools in Malware Detection*. Auerbach Publications. ISBN 1439854548, 2011.

**S. Nair, M. Marchetti, J. Hopkinson, S. Fogie.** *Enterprise Security: Compliance v. Competence*, 1st Edition. Springer. ISBN 0387744347, 2011.

**T. Wrightson.** *Wireless Network Security. A Beginners Guide*. McGraw-Hill Osborne Media. ISBN 0071760946, 2011.

##### **Tema:** Congresos, conferencias, symposiums

**Eurocrypt '2012.** Del 15 al 19 de abril 2012. Cambridge, UK.

**AISC '2012 (Australasian Information Security Conference).** Del 30 enero al 3 de febrero 2012. RMIT University Melbourne, Australia.

**CIBSI '2011 (Congreso Iberoamericano de Seguridad Informática).** Del 2 al 4 de noviembre 2011. Bucaramang, Santander, Colombia.

**NDSS '2012 (19th Network and Distributed System Security Symposium).** Del 5 al 8 de febrero 2012. San Diego, California, USA.

**ICNC '2012. Communications and Information Security.** Del 30 de enero al 2 de febrero 2012. Maui, Hawaii.

#### **Sección técnica "Sistemas de Tiempo Real"** (Alejandro Alonso Muñoz, Juan Antonio de la Puente Alfaro)

##### **Tema:** Libro

Acaba de aparecer el libro *Building Parallel, Embedded and Real-Time*

*Applications with Ada*. Sus autores son **John McCormick**, **Frank Singhoff** y **Jérôme Hugues**, todos ellos bien conocidos en el mundo del lenguaje Ada. El libro contiene un resumen de la parte secuencial de este lenguaje, para centrarse a continuación en los aspectos relacionados con la concurrencia y los sistemas distribuidos, con una discusión de los modelos definidos en el anexo de sistemas distribuidos de Ada y en el estándar CORBA. La planificación de sistemas de tiempo real y las técnicas de análisis temporal son el objeto de un capítulo del libro. Por último se abordan los aspectos prácticos de la programación de sistemas de tiempo real en Ada, y se describen algunas plataformas, como MaRTE, ORK+ y RTEMS, y algunas herramientas de diseño y análisis temporal.

En suma, se trata sin duda de un texto muy útil para apoyar la enseñanza de los sistemas de tiempo real, tanto en el ámbito industrial como en la universidad. El libro ha sido publicado por Cambridge University Press, con ISBN 978-0-521-19176-8.

### Sección Técnica: "Tecnología de Objetos" (Jesús García Molina, Gustavo Rossi)

**Tema:** Libro

**John Watkins**. *Agile Testing: How to succeed in an extreme testing environment*. Cambridge University Press, 2009. En entregas anteriores, nos hemos referido ya a los métodos ágiles y al uso de las pruebas durante el proceso de desarrollo. Este texto, muy completo y didáctico, aborda dicho tema desde la perspectiva de las pruebas mismas.

El libro, tal cual expresa su autor, está orientado a desarrolladores, miembros de equipos de desarrollo y pruebas, y también a investigadores interesados en métodos ágiles de desarrollo; incluye no solo una buena descripción de las ideas "teóricas" sino además algunos casos de estudio reales para enfatizar dichas ideas y mostrar su practicidad.

El texto está dividido en tres partes:

La primera, "*Review of Old Schools and Agile Approaches*" es muy breve y describe el lugar de las pruebas en el proceso de desarrollo, tanto en los enfoques más convencionales como en los modernos.

La segunda parte está planteada de manera muy original y es un conjunto de capítulos escritos por diversos autores que discuten diversos puntos de vista respecto al concepto de agilidad en base a casos de estudio reales, todos ellos descritos con el mismo formato, donde cada autor presenta su propia experiencia.

Sin embargo, cada capítulo tiene su propio interés al margen del caso de estudio ya que algunos exceden el análisis del caso y son breves ensayos teóricos como el capítulo 4 "*From Waterfall to Evolutionary Development and Test*" mientras otros abordan los casos de estudio (muy interesantes y en dominios particularmente críticos como el capítulo 8 "*Testing a Derivatives Trading System in an Uncooperative Environment*") en forma sucinta.

Algunos capítulos enfocan aspectos más específicos tales como calidad, proyectos "offshore", automatización de pruebas, etc. Todos los capítulos merecen ser leídos y representan excelentes puntos de partida para profundizar en los temas específicos.

Finalmente, la tercera parte, que representa el corazón del libro, analiza los casos de estudio desde una perspectiva más abstracta demostrando como cada proyecto puede necesitar un enfoque de pruebas ágiles particular y de que manera "personalizar" dichos procesos.

Los apéndices presentan desde datos específicos de los casos de estudio hasta plantillas utilizables en los procesos de "agile testing" discutidos en la tercera parte del libro. Si bien el texto se hace algo extenso, es de fácil lectura y muy práctico.

### Sección Técnica: "TIC y Turismo" (Andrés Aguayo Maldonado, Antonio Guevara Plaza)

**Tema:** Observatorio sobre Viajes e Internet

El Observatorio sobre Viajes e Internet <[www.minube.com/observatorio](http://www.minube.com/observatorio)> nace de la iniciativa de Minube.com, en colaboración con la Junta de Comunidades de Castilla-La Mancha y con la participación de la empresa The Cocktail Analysis como consultora, para la realización los estudios de campo, con la misión de presentar de forma semestral las tendencias en Internet de este sector. El objetivo es presentar una perspectiva más ajustada de la nueva realidad de los viajes y compartirlo con todas las instituciones, empresas y particulares tanto de España como en foros extranjeros.

Hasta la fecha, se han publicado dos estudios, titulados "*Tendencias del Nuevo Viajero*" y "*Tendencias en el uso de Internet en el móvil para los viajes*". El primero de ellos, presentado en junio de 2010, se ha centrado en el propio viajero y aborda los cambios de comportamiento que el viajero habitual ha experimentado con el crecimiento y desarrollo de Internet. El uso extensivo de la Red en todas las fases del viaje le ha llevado a convertirse en el canal más usado para elegir destino, así como el medio idóneo para resolver desde las cuestiones más prácticas de elección de lugares y ruta a seguir, hasta la resolución de problemas o la creación de contenidos personales para compartir en directo con los amigos y conocidos.

Las principales conclusiones que se desprenden de este estudio señalan que el uso de Internet se ha extendido a la totalidad de las fases del viaje (elección, preparación, viaje y post viaje) y se erige como fuente más usada y fiable, sobre todo a la hora de seleccionar destino y comunicar la experiencia del viaje, a través de las redes sociales fundamentalmente. Para el 85% de los encuestados, los amigos e Internet son la fuente de inspiración en el inicio de preparación de un viaje, aunque también tienen gran relevancia las páginas web del destino.

Durante el viaje, el uso de Internet sigue estando por encima de otras fuentes a la hora de buscar información. Un 55% de los usuarios acuden a la Red para recabar información durante su viaje; el 36% de los que se conectan a Internet durante su estancia lo hace a través del móvil y 3 de cada 10 viajeros comparten su experiencia en tiempo real, sobre todo colgando fotos o videos en las redes sociales.

El segundo de los estudios, presentado en diciembre de 2010, se centra en el uso de Internet en el móvil para los viajes y concluye que ocho de cada diez internautas en este tipo de dispositivos lo han incorporado al proceso de preparación de sus viajes, así como al propio desarrollo del mismo para buscar información, cambiar en tiempo real su plan de viaje, o incluso, compartir su experiencia con amigos, familiares y otros "compañeros de viaje virtuales". El móvil es la nueva brújula de los viajes para orientarse en la resolución de necesidades como encontrar hoteles, restaurantes, actividades de ocio, ver fotos, etc. Configurar una experiencia lo más personalizada posible es el objetivo último, así como registrar lo que nos sucede y compartirlo en redes sociales. Como elementos que frenan en cierta medida este imparable avance, los usuarios destacan la necesidad de adaptar aún más los contenidos a las características de los propios terminales y las tarifas de conexión en los desplazamientos internacionales.

Ambos estudios, junto con las encuestas empleadas y los datos en bruto están disponibles en la web de Minube.com.

Tomás F. Tornadijo Rodríguez  
*Analista informático; Encargado del dpto. de informática de Cartonajes Vir, S.A.*

<tornadijo@telecable.es>

# Criptoanálisis mediante algoritmos genéticos de una comunicación cifrada en la Guerra Civil

## 1. Introducción

Entre los diferentes métodos de cifrado utilizados durante la Guerra Civil, uno de los más empleados por ambos contendientes fue una variante del cifrado por tabla de homófonos: el criptógrafo de cinta. Este sistema consiste en una tabla que tiene un alfabeto ordenado en la fila superior, bajo el que se dispone una cinta móvil con otro alfabeto doble aleatorio que encabeza las columnas de homófonos, que son las cifras de un par de dígitos que sustituirán a las letras en el mensaje cifrado.

El emisor y el receptor del mensaje convienen una correspondencia determinada entre una letra de la primera fila y otra de la fila inferior, desplazando la cinta móvil para alinearlas. Entonces, para cifrar un mensaje se busca cada letra del mismo en la cinta móvil, substituyéndola en el texto cifrado por uno cualquiera de los homófonos de su columna, variación que hará menos susceptible esta cifra a los intentos de ruptura mediante la técnica del análisis de frecuencias.

El 29 de julio de 1936 el general Mola envió al general Franco un radiotelegrama cifrado pero con partes del texto en claro, de cuyo contenido se desprende que es una parte de operaciones de varios teatros bélicos. Este radio fue interceptado por los republicanos y una copia del mismo se puede ver en el *Taller de criptografía* [1], la extensa página Web sobre el particular del profesor Arturo Quirantes Sierra.

El encabezado del mensaje aclara que se utiliza la clave *Regidor*, que es un cifrado homofónico, sin cinta móvil [2, pág. 43], de modo que las columnas de homófonos se corresponden con un único alfabeto. Vamos a intentar romper la cifra de esta comunicación militar utilizando un algoritmo genético, ayudándolo con algún término que podamos conjeturar a partir de los fragmentos de texto en claro.

## 2. Metodología

### 2.1. Datos de partida

Supondremos que el documento que vamos a tratar de descifrar, está cifrado con una clave homofónica, de manera que cada número de dos dígitos se corresponde con una única letra en el texto en claro, en tanto que cada letra puede estar representada por varios números. Los cinco bloques en cifra se agruparán en uno solo para su descifrado.

**Resumen:** El propósito de este artículo es mostrar la aplicación de un algoritmo genético, asistido con información de contexto del mensaje, al desciframiento de un radiotelegrama cifrado mediante una tabla de substitución homofónica al comienzo de la guerra civil española.

**Palabras clave:** Algoritmo genético, cifra homofónica, criptoanálisis.

Por otra parte, se observa que el último homófono del mensaje es el 00, que no se repite en ninguna otra posición, por lo que parece razonable suponer que se trate de algún de indicativo de fin de comunicación, así que se considerará, en principio, que no codifica ninguna letra.

El mensaje se presenta tal cual se publica en la Web, no se ha investigado en los archivos militares si el documento está descifrado, ni se ha buscado por otros medios la clave del mismo.

### 2.2. Método de descifrado

Para establecer esta correspondencia entre cifras y letras del alfabeto vamos a utilizar un algoritmo genético, que es un método de ataque ampliamente utilizado en el criptoanálisis moderno [3], y que, además, ya ha sido aplicado al descifrado de documentos cifrados durante la guerra civil [2].

Sin embargo los esfuerzos desarrollados hasta el momento solo permiten atacar con éxito aquellos documentos cifrados de los que se conozca la tabla de homófonos pero no el alfabeto de cinta móvil, es decir cuando se sabe que cifras van juntas, pero no a que letra corresponden, pues de otra manera el espacio de búsqueda resulta demasiado amplio [2, pág. 17].

Para intentar resolver esta dificultad proponemos un algoritmo genético con una función objetivo que valore la diferencia entre las frecuencias encontradas de letras y bigramas y las frecuencias esperadas procedentes de un texto de referencia. Además este algoritmo tendrá un operador que permitirá vincular una palabra o texto corto a un determinado fragmento del mensaje, de forma que si esa palabra o texto corto realmente figurase en el texto en claro y en esa misma posición, entonces los homófonos correspondientes a sus letras quedarían bien asignados, resolviéndose correctamente otras porciones del mensaje y disminuyendo, en principio, la diferencia de frecuencias, mejorándose así el valor de la función objetivo.

El programa analizará la palabra clave, colocándola sucesivamente en todas las posiciones del texto y lanzará el algoritmo genético para cada una de ellas, guardando el valor de la función objetivo en una tabla, junto con texto descifrado, para obtener una estadística del desempeño.

Finalizado el análisis, comprobaremos si el texto en claro resulta inteligible, aunque sea de forma parcial, para los mejores valores de aptitud conseguidos. Si no es así, pensaremos en una nueva palabra y prepararemos el programa para realizar un nuevo análisis.

Hay que observar que este procedimiento es solo semiautomático, pues necesitamos conocer ciertos detalles pertenecientes al contexto del mensaje, por lo que resulta más apropiado para el criptoanálisis de documentos que alternen el texto en cifra con zonas en claro, como es el que nos ocupa.

## 3. Algoritmo genético

### 3.1. Introducción

Los algoritmos genéticos son unos métodos heurísticos de búsqueda y optimización, basados en la teoría de la evolución de Darwin, de acuerdo a la cual los individuos con mejor grado de adaptación al medio (*eficacia biológica* o *fitness*) sobreviven en mayor número, transmitiendo a la descendencia los rasgos adaptativos expresados por sus cromosomas.

### 3.2. Descripción de un algoritmo genético

Los algoritmos genéticos [4] imitan la acción de la selección natural a partir de una población de partida, cuyos individuos codifican en una cadena (cromosoma) una solución inicial a un problema, normalmente aleatoria. El algoritmo actúa sobre esta población sometiendo a unas operaciones similares a los mecanismos evolutivos naturales:

- Selección. Los individuos más aptos tienen mayor probabilidad de pasar a la siguiente generación.
- Cruce. Consiste en combinar las características de los cromosomas de dos individuos

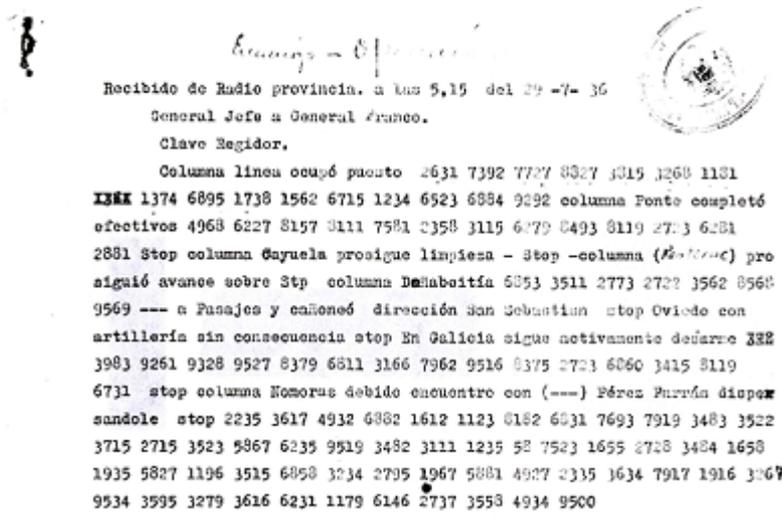


Figura 1. Radiotelegrama cifrado. Fuente: Arturo Quitantes [1].

para obtener uno nuevo.

■ **Mutación.** Con cierta probabilidad se altera aleatoriamente el cromosoma de un individuo, incrementando con ello la diversidad genética de la población.

Un algoritmo genético canónico suele seguir el siguiente esquema:

```
{
  Generar la población inicial, Pi
  Evaluar aptitud Pi
  Repetir hasta g generaciones
  {
    Seleccionar los individuos de Pi que
    pasarán a Pi+1
    Cruzar individuos de Pi y colocar en Ptemp
    Mutar individuos de Ptemp y colocar en Pi+1
    Pi = Pi+1
    Evaluar aptitud Pi
  }
}
```

Si el algoritmo está correctamente diseñado, la población convergerá hacia el mejor resultado que quedará registrado en el cromosoma del individuo más apto de la última generación.

**3.3. Codificación**

Cada individuo de la población representará una posible clave del mensaje, es decir: la relación de la tabla de homófonos con las letras del abecedario. Los cromosomas estarán formados por una cadena de texto con tantas letras del abecedario como homófonos diferentes existan en el mensaje, estando relacionada cada posición del cromosoma con

una cifra determinada. Esta representación permitirá al programa atacar indistintamente tanto la cifra homofónica convencional como la de cinta móvil, pues al final ambas no pasan de ser una correspondencia donde se relaciona cada letra con varias cifras. Por otra parte, al no utilizar todos los números del 0 al 99, sino solo los presentes en el mensaje, evitaremos generar combinaciones improductivas y ahorraremos tiempo de proceso. En la **figura 2** se muestra un ejemplo con los cromosomas de dos individuos.

**3.4. Selección**

El algoritmo emplea una selección por torneo binario determinístico, procedimiento que consiste en tomar al azar dos individuos de la población seleccionando al más apto de ellos. Además se utiliza el elitismo, para asegurar que los individuos mejor adaptados de una generación pasen siempre a la siguiente.

**3.5. Función objetivo**

Para evaluar la aptitud de un individuo, el algoritmo utiliza sus cromosomas como clave para decodificar el texto cifrado. A continuación, con el texto conseguido, calcula la diferencia entre las frecuencias encontradas de letras y bigramas y las frecuencias esperadas, procedentes de un texto de referencia (se utilizó la novela *La Primera República*, de B. Pérez Galdós). Con objeto de disminuir el tiempo de localización de los bigramas se implementó un procedimiento de búsqueda binaria en un vector ordenado.

La suma de las diferencias de letras y bigramas supone el error entre el texto descifrado y el paradigma, de manera que la idoneidad de la clave analizada, *k*, vendrá dada por la siguiente función [3, pág 43]:

$$C_k = \alpha \cdot \sum_{i \in A} |K_{(i)}^u - D_{(i)}^u| + \beta \cdot \sum_{i,j \in A} |K_{(i,j)}^b - D_{(i,j)}^b|$$

Donde *A* denota el alfabeto [A..Z], *K* denota las frecuencias de referencia, *D* las del texto analizado, los subíndices *u* y *b* las frecuencias de letras y bigramas respectivamente, en tanto que  $\alpha$  y  $\beta$  son los pesos asignados a letras y bigramas, no resultando práctico introducir las frecuencias de los trigramas [3, pág 43]. Para este problema el algoritmo mejora la convergencia utilizando como pesos los valores 1 y 2 respectivamente. Además, para conseguir una función creciente, a medida que disminuye el error, se ha aplicado:

$$C_k = 5 - C_k$$

**3.6. Población inicial**

Rellenaremos cada posición del cromosoma con letras del abecedario tomadas aleatoriamente hasta la última posición, la 50, que es el total de homófonos diferentes (excluyendo el último, como se ha dicho anteriormente) que hay en el mensaje cifrado.

**3.7. Mutación**

Con cierta probabilidad, un individuo seleccionado sufre una operación de mutación, que consiste en elegir al azar un gen, un carácter de su cromosoma, y a continuación substituirlo por una letra tomada aleatoriamente del abecedario.

**3.8. Cruce**

El algoritmo utiliza uno de los operadores más sencillos, el cruce de un punto, que consiste en cortar dos cromosomas por una posición escogida al azar, generando cada uno un segmento de cabeza y otro de cola. Entonces el segmento de cabeza de un progenitor y el de de cola de otro se unen para dar lugar a un nuevo individuo.

**3.9. Texto supuesto**

Una vía para disminuir el elevadísimo número de combinaciones con las que tiene que trabajar el programa consiste en asociar un frag-

C	A	H	G	I	E	O	D	A	N	K	Z	B	E	O	S	N	..
00	11	12	15	16	17	19	22	23	26	27	28	31	32	34	35	36	..
A	O	J	E	C	B	V	Y	S	C	E	I	N	F	E	A	S	..
00	11	12	15	16	17	19	22	23	26	27	28	31	32	34	35	36	..

Figura 2. Dos cromosomas.

mento del mensaje cifrado con un texto en claro, una palabra o una frase corta supuesta a partir de la información que nos proporcionan las zonas en claro del mensaje, un procedimiento análogo a los rastreos de contenidos estereotipados que efectuaban los criptoanalistas británicos en los partes meteorológicos cifrados de la marina alemana durante la segunda guerra mundial, con la diferencia de que, en este caso, no podemos saber donde está ese fragmento de texto, por lo que probaremos en todas las posiciones, lanzando el algoritmo genético para cada una de ellas, de acuerdo al siguiente esquema:

```

{ t = longitud del texto cifrado
  b = longitud del texto de búsqueda
  x = 0
  Repetir hasta que x = t - b + 1
  {x = x + 1
   Lanzar algoritmo genético [x, texto de
   búsqueda]
   Guardar x
   Guardar aptitud
   Guardar texto en claro
  }
}
    
```

Además el algoritmo genético cuenta con un operador adicional que modifica los cromosomas de los individuos en cada nueva generación, de forma que el mensaje en claro contenga el texto supuesto en la posición indicada.

**4. Ajuste**

**4.1. Ensayos**

Para ajustar los parámetros del algoritmo genético, se codificó un mensaje de la misma longitud que el radiotelegrama analizado, utilizando una clave con el mismo número de homófonos, y se ejecutó el programa (sin proporcionarle un texto supuesto) para un total de cinco combinaciones de pesos, contando la cantidad de letras coincidentes entre el mensaje descifrado por el programa y el texto original sin cifrar, con la intención de obtener un valor real de aptitud. Para cada combinación se ejecutó el algoritmo 50 veces,

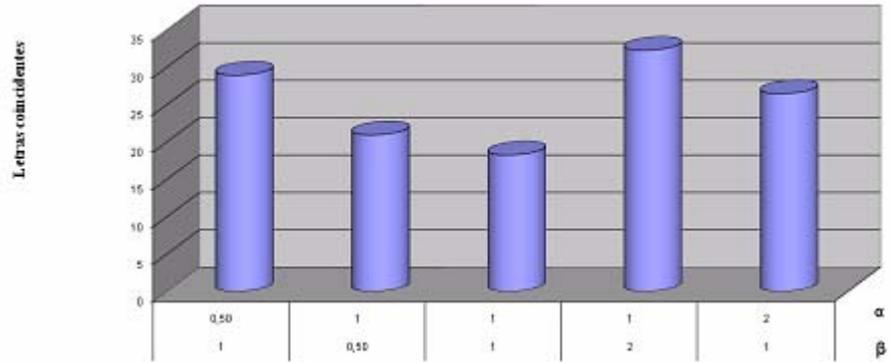


Figura 3. Aptitud real en función de los pesos α y β.

calculando el valor medio de la función de coste (ver figura 3).

Los valores de las combinaciones primera y cuarta (ver figura 3) muestran una mejora del valor de la función de coste cuando aumenta el peso asignado a los bigramas, lo que resulta coincidente con los estudios de Clark & Dawson [3, pág 44] y otras experiencias [5, pág 19]. Usaremos la mejor combinación, con los valores α = 1 y β = 2.

Asimismo se comprobó que el algoritmo se comportaba bien con un valor de mutación alto: 0,008, y con un reparto de individuos y generaciones no muy elevado, de 500 y 50, respectivamente (ver figura 4). Todos estos parámetros serán los que utilicemos en el descifrado del mensaje de la guerra civil.

**5. Resultados**

Las primeras pruebas se realizaron sin proporcionar al algoritmo un texto supuesto, pero se observó que, aunque convergía adecuadamente consiguiendo generación tras generación valores de aptitud crecientes, el texto en claro obtenido al final del proceso nunca resultaba legible.

Como en una de las zonas no cifradas del telegrama se hace mención a la columna del general Miguel Ponte y Manso de Zúñiga, que por aquellas fechas intentaba abrirse paso hacia

Madrid por la Sierra de Guadarrama, se consideró posible que el texto cifrado contuviese alguna mención a la ciudad de Madrid, objetivo último de la citada columna, información que habría resultado igualmente obvia para algún criptoanalista coetáneo al mensaje.

Sin embargo las pruebas realizadas con esos textos supuestos resultaron infructuosas, y no se pudo obtener ningún texto legible. Otros intentos utilizando algunas palabras típicas de la terminología militar, como *bombardeo*, *artillería*, *aviación*, *posición*, etc., tampoco arrojaron resultados satisfactorios.

Otro de los párrafos en claro hace referencia a la columna del capitán Pablo Díaz Duñabeitia, que por entonces operaba en Guipuzcoa, lo que hacía verosímil la presencia en el texto cifrado de alguna alusión a la capital de la provincia, San Sebastián, donde a la sazón se estaban produciendo importantes acontecimientos.

En este caso, con el valor de aptitud más alto conseguido cuando el texto se asocia a partir de la posición 158 del mensaje cifrado (ver figura 5), se obtuvo el siguiente resultado, donde se leen algunas palabras (en negrita) y donde se sugiere alguna otra más: MOMQ JASAPBTANORHANKPBREBSIALAT QQ/QARAOYONCOLSOBRETPODALR OUO/AUENAMALERIANC/UDQFPU NADEANOMERNODCALAFIBODEO/ MLKQTACO SNLOCAOSPEDIDEMV BABELSERENDICONSESCLOSAUTOS **DESANSEBASTIANDESOQALELIEK DOTENIENTELORONE FZAVESQIN**

Podríamos intentar mejorar la solución mediante un reproceso con un mayor número de generaciones para esta posición del texto de búsqueda, pero también podemos intentar romper la cifra usando las nuevas palabras que conocemos, realizando un criptoanálisis clásico. Por ejemplo, podemos completar TENIENTE CORONEL, que parece una suposición bastante obvia a partir del texto que sigue a TENIENTE, lo que nos proporciona el siguiente mensaje, compuesto únicamente con las letras que hemos ido fijando: \*O\*\*\*A\*A\*BTAN\*\*\*AN\*\*BREB\*I\*\*A\*\*/

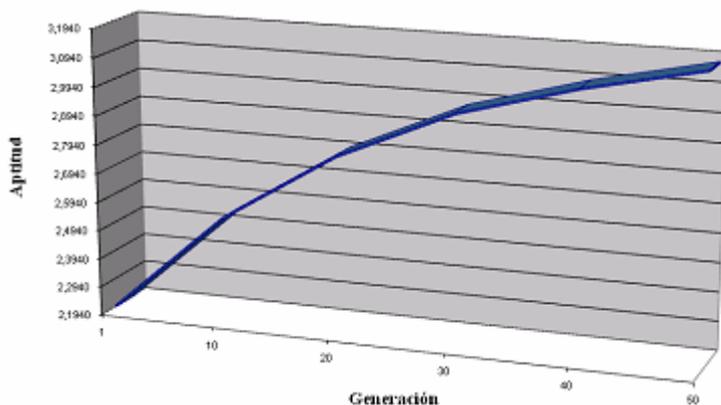


Figura 4. Aptitud en función del número de generaciones.

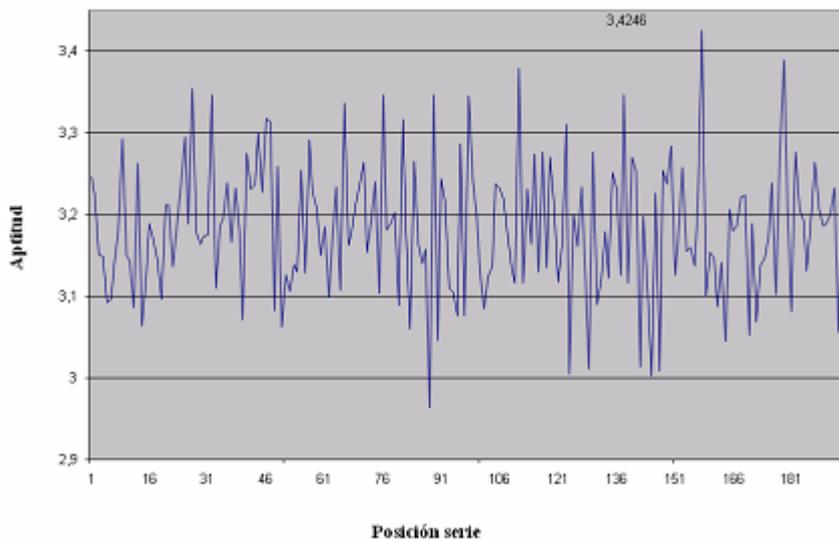


Figura 5. Aptitud en función de la posición del texto supuesto.

**\*ARA\*\*\*N\*\*\*SOBRE\*\*\*DA\*R\*\*\*/A\*ENA\*  
A\*ER\*AN\*/\*\*\*L\*\*NA\*EANO\*ERNO\*\*  
A\*A\*IB\*DEO/\*EC\*\*TA\*O\*N\*\*\*AO\*\*EDI\*  
E\*\*BABE\*SERENDI\*ON\*ES\*\*O\*A\*I\*  
OSDESANS EBASTIANDES\*\*A\*ECIE\*  
DOTENIENTECORONEL\*A\*ES\*IN**

El párrafo en negrita es claro que solo puede referirse al teniente coronel Vallespín, oficial al mando de los cuarteles de Loyola en aquella época. Completando VALESPIIN, tenemos: **\*O\*\*\*A\*A\* BTAN\*\*\*AN\*\*BREB\*I  
\*\*A\*\*\*/PARA\*\*\*N\*\*\*SOBRE\*\*\*DA\*R\*\*\*/  
A\*ENA\*A\*ER\*AN\*/\*\*\*L\*\*NA\*EANO\*ER  
NO\*\*A\*A\*IB\*DEO/\*EC\*PTA\*O\*N\*\*\*AO  
\*\*EDI\*E\*LBABE\*SERENDI\*ON\*ES\*\*ONA  
\*I\*OSDESANSEBASTIANDES\*PA\*EC  
IE\*D OTENIENTE CORONEL VALESPIIN**

Completando DESAPARECIENDO: **\*O\*\*\*  
A\*A\*BTANA\*\*ANN\*BREB\*I\*RA\*\*\*/PARA  
A\*AN\*ARSOBRE\*\*ADARRA\*A/A\*ENA  
\*ARER\*AN\*/\*\*\*L\*\*NA\*EANO\*ERNO\*\*  
ARA\*IBADEO/\*ECNPTA\*O\*NRA\*AO\*\*  
EDI\*E\*LBABERSERENDI\*ON\*ES\*RONA  
\*I\*OSDESANSEBASTIANDESAPARECI  
ENDOTENIENTECORONELVALESPIIN**

Completando AVANZARGUADARRAMA Y RIBADEO: **\*O\*\*\*A\*A\*BTANA\*\*ANN\*  
BREB\*I\*RAG\*\*\*/PARAAVANZARSOBRE  
GUADARRAMA/A\*ENA\*ARER\*AN\*/\*\*\*  
LUMNA\*EANO\*ERNO\*ZARARIBA  
DEO/\*ECNPTA\*O\*NRA\*AO\*UEDI\*E  
\*LBABERSERENDI\*ON\*ESZRONAMI  
GOSDESANSEBASTIANDESAPARECIEN  
DOTENIENTECORONELVALESPIIN**

Completamos COLUMNA, PERNOCTARA, DETECTADO, UN y vemos que hay varios homófonos confundidos en el texto cifrado: **\*O\*O\*A\*A\*BTANA\*\*ANN\*BREBUI\*RA  
GOOPARA AVANTARSOBREGUADARRA**

MA/A\*ENA\*ARER\*AN\*/\*COLUMNACEA  
NO PERNOCTARARIBA DEO/DECNPTA  
DOUNRADIOQUEDI CEDLBABERSE  
RENDIDONUESTRONAMIGOSDESAN  
SEBASTIANDESAPARECIENDOTENIEN  
TECORONELVALESPIIN

Completamos BUITRAGO: **\*O\*O\*A\*A\*  
\*BTANA\*\*ANN\*BREBUIRAGOO/PARA  
AVANTARSOBREGUADARRAMA/A\*  
ENA\*ARER\*AN\*/\*COLUMNACEANO  
PERNOCTARARIBADEO/DECNPTADO  
UNRADIOQUEDICEDLBABERSEREN  
DIDONUESTRONAMIGOSDESANSE  
BASTIANDESAPARECIENDOTENIE  
NTECORONELVALESPIIN**

Completamos AMENAZA: **\*OZO\*A\*A\*  
BTANA\*\*ANN\*BREBUIRAGOO/PARA  
AVANTARSOBREGUADARRAMA/AME  
NAZARER\*AN\*/\*COLUMNACEANO  
PERNOCTARARIBADEO/DECNPTADO  
UNRADIOQUEDICE DLBA BERSEREN  
DIDONUESTRONAMIGOSDESANSE  
BASTIANDESAPARECIENDOTENIEN  
TE CORONELVALESPIIN**

Completamos HERNANI, nuevamente se han confundido homófonos: **\*OZO\*A\*A\*  
\*BTANA\*\*ANN\*BREBUIRAGOO PARA  
AVANTARSOBREGUADARRAMA/AME  
NAZARERNANI\*COLUMNACEANOPE  
RNOCTARARIBADEO/DECNPTADO  
UNRADIOQUEDICEDLBABERSEREN  
DIDONUESTRONAMIGOSDESANSE  
BASTIANDESAPARECIENDOTENIE  
NTECORONELVALESPIIN**

Completamos LOZOYA, por ser el valle donde está Buitrago: **LOZOYA\*A\*BTANA  
\*\*ANN\*BREBUIRAGOO/PARA AVAN  
TARSOBREGUADARRAMA/AMENA  
ZARERNANI/\*COLUMNACEANOPER**

NOCTARARIBADEO/DECNPTADOUN  
RADIOQUEDICEDLBABERSERENDI  
DONUESTRONAMIGOSDESANSEBAS  
TIANDESAPARECIENDOTENIENTE  
CORONELVALESPIIN

El párrafo entre LOZOYA y BUITRAGO no se ha podido resolver, tal vez por la existencia de más errores. El texto descifrado, interpolando los errores evidentes, quedaría así: **LOZOYA\*\*\*\*\*BUITRAGO  
\*/PARA AVANZARSOBREGUADARRA  
MA/AMENAZAHERNANI/\*COLUMNA  
CEANOPERNOCTARARIBADEO/DE  
TECTADOUNRADIOQUEDICEDEHA  
BERSERENDIDONUESTROSAMIGOS  
DESANSEBASTIANDESAPARECIEN  
DOTENIENTE CORONELVALLES PIN\***

### 6. Conclusiones

Se ha conseguido romper la cifra de este mensaje y descifrarlo en su mayor parte y, aunque el método aplicado no puede considerarse una solución general para la ruptura de las cifras homofónicas, sí puede resultar de ayuda en los casos en los que se disponga de alguna información sobre el contenido o el contexto del mensaje. Como desarrollos futuros cabe imaginar un programa capaz de probar no una, sino varias palabras clave, tomándolas desde una base de datos, a costa de un importante tiempo de proceso.

### Referencias

- [1] A. Quirantes Sierra. *Taller de criptografía*, 2010, [2-12-2010]. Disponible en <<http://www.cripto.es>>.
- [2] A.S. Gascón González. *Aplicación de algoritmos genéticos en el ataque de textos cifrados durante la guerra civil española*. Proyecto Fin de Carrera, 2010. <<http://www.iit.upcomillas.es/pfc/resumenes/4c8615ad0a2b6.pdf>>.
- [3] Bethany Delman. *Genetic Algorithms in Cryptography*, 2004, <<https://ritdml.rit.edu/bitstream/handle/1850/263/thesis.pdf?sequence=1>>.
- [4] J.H. Holland. *Adaptation in Natural and Artificial Systems*. Ann Arbor: The University of Michigan Press, 1975.
- [5] Pallavi Kanagalakatte Basavaraju. *Heuristic Search Cryptanalysis of the Zodiac 340 Cipher*. Master's Projects. Paper 56, 2009. <[http://scholarworks.sjsu.edu/etd\\_projects/56](http://scholarworks.sjsu.edu/etd_projects/56)>.

### Bibliografía

- A. Clark, Ed. Dawson. *Optimisation Heuristics for the Automated Cryptanalysis of Classical Ciphers*. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 28, 1998, pp. 63-86.
- J. García Carmona. *Tratado de criptografía con aplicación especial al ejército*. Madrid: Sucesores de Rivadeneyra, 1894.
- R. Otéeme, S. Arumugam. *Applying Genetic Algorithms for Searching Key Space of Polyalphabetic Substitution Ciphers*. *The International Arab Journal of Information Technology*, vol. 5 (1), 2008.

Julio Javier Castillo, Diego Javier Serrano  
 Laboratorio de Investigación de Software MsLabs, Dpto. Ing. en  
 Sistemas de Información, Facultad Regional Córdoba - Universidad  
 Tecnológica Nacional (Argentina)

<jotacastillo@gmail.com>,  
 <diegojserrano@gmail.com>

# El problema del decodificador

Esta es una adaptación del enunciado del problema C de los planteados en la Segunda Competencia de Programación de la Facultad Regional de Córdoba (Universidad Tecnológica Nacional, Argentina) UTN-FRC celebrada el 24 de noviembre de 2010.

**Nivel del problema:** Medio

La codificación es el proceso mediante el cual ciertos símbolos se convierten en otros símbolos en un sistema de representación determinado, aplicando ciertas reglas de codificación. La decodificación puede definirse como el proceso inverso en el cual se convierten símbolos en información entendible por el receptor.

En este problema se requiere decodificar cierta información provista como entrada mostrando el mensaje decodificado en la salida.

Se sabe que el mensaje original es una frase en Español. Se conoce también que la frase no contiene los símbolos '(', ')', '{', '}', '[', ']', '!', 'y', '\_'.

Se nos informa que se aplicaron las siguientes reglas de *codificación* sucesivamente al mensaje original en este orden:

- 1) Se rota cada palabra del mensaje original. Ejemplo: "Hola" se codificó como "aloH".
- 2) Se aplica la siguiente función de transformación  $f$ , a cada palabra

$$p = a_1 a_2 \dots a_{k-1} a_k a_{k+1} \dots a_n$$

de longitud  $n$ , y  $k = g(n)$ , donde  $a_i$  un carácter de dicha palabra, donde:

$$g(n) = \begin{cases} \frac{n}{2}, & \text{si } n \text{ es par} \\ \frac{n-1}{2} + 1, & \text{si } n \text{ es impar} \end{cases}$$

$$f(a_1 a_2 \dots a_{k-1} a_k a_{k+1} \dots a_n) =$$

$$\begin{cases} a_{k-1} a_{k-2} \dots a_1 a_k a_n \dots a_{k+1}, & \text{si } n \text{ es impar} \\ \dots \\ a_k a_{k-1} \dots a_1 a_n a_{n-1} \dots a_{k+1}, & \text{si } n \text{ es par} \end{cases}$$

- 3) Si el mensaje (luego de los pasos 1 y 2) presenta palabras de dos caracteres, entonces se amplían esas palabras a cuatro caracteres duplicando cada carácter, por ejemplo "de" se ampliaría a "ddee" y "la" a "llaa".
- 4) Si en el paso 3 se generó una cantidad par positiva de palabras entonces, se toman cada una de estas palabras y se reemplazan todas las vocales 'a' por 'e', las 'e' por 'i', las 'i' por 'o', las 'o' por 'u', y las 'u' por 'a', simultáneamente.
- 5) Los caracteres 'p' se reemplazan por '\_' (guion bajo).

Se deben mostrar en la salida el mensaje original.

**Entrada:**

Cada caso de prueba se recibe en una línea que contiene una frase codificada finalizada con punto ('.'). El fin de entrada estará dado por una cadena vacía en la línea final.

**Salida:**

Por cada caso de prueba se debe imprimir una línea con el mensaje original (decodificado), respetando los espacios entre palabras en el mensaje original.

**Ejemplo:**

**Entrada:**

```
laHo donmu elucr.
taEs ssee anu eba_ru.
uuNN ssii eell 1 zev euq ala iivv.
```

**Salida:**

```
Hola mundo cruel.
Esta es una prueba.
No es la 1 vez que ala ve.
```

Julio Javier Castillo, Diego Javier Serrano,  
Marina Elizabeth Cardenas

Laboratorio de Investigación de Software MsLabs, Dpto. Ing. en  
Sistemas de Información, Facultad Regional Córdoba - Universidad  
Tecnológica Nacional (Argentina)

<jotacastillo@gmail.com>,  
<diegojserrano@gmail.com>,  
<ing.marinacardenas@gmail.com>

Para resolver este problema nos basaremos en la observación de que el Triángulo de Pascal puede ser convenientemente representado de la siguiente forma que se visualiza en la **figura 1**.

```

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
    
```

**Figura 1.** Coeficientes binomiales representados mediante una matriz triangular izquierda.

Esta forma de representar los coeficientes binomiales constituye una manera casi directa de implementación algorítmica empleando matrices.

El programa presentado a continuación ha sido modelado utilizando el paradigma orientado a objetos. Sin embargo, dada la simplicidad del problema y su resolución netamente algorítmica, la codificación de la solución se realizó solamente en una clase Main utilizando dos métodos estáticos. Por ello, la implementación de esta solución sería muy similar a la necesaria en un lenguaje de programación como C/C++ u otros lenguajes imperativos.

La clase Main contiene el método principal "main()" mediante el cual solicitaremos el ingreso de datos por la entrada estándar. Se realiza la lectura de tres números enteros: a, b, y n (donde: n <= 10, y que 0 <= a, b <= 4) correspondientes a la fórmula **(a + b)<sup>n</sup>**. De esta manera, la salida esperada del programa será el valor de **(a + b)<sup>n</sup>** seguido de los coeficientes binomiales del n-ésimo nivel del triángulo.

Dicho programa contiene un segundo método llamado "calcularTriangulo()" que implementa el cálculo de las 10 primeras filas (ya que n <= 10) de la matriz de coeficientes binomiales (ver **figura 1**). Nótese también que, tanto la fila 0 como la diagonal de la matriz, tendrán siempre asignado el valor 1 siguiendo la ecuación de la potencia del binomio que se corresponde con elevar un número a la potencia 0. El cálculo del valor del coeficiente para la i-ésima fila se realiza en base a la suma del valor de los coeficientes de la fila inmediatamente anterior (i-1), el valor de la columna actual (j) y el de la columna anterior (j-1). Esta matriz es necesaria para poder imprimir los coeficientes binomiales del n-ésimo nivel del triángulo de Pascal.

Posteriormente, una vez construida esta matriz solo resta imprimir el valor de **(a + b)<sup>n</sup>** y luego los coeficientes correspondientes de ese nivel.

Para imprimir dicho valor tenemos dos alternativas:

- La primera es simplemente computar este valor ya que los valores de las variables a, b y n son conocidos.
- La segunda alternativa es seguir la definición y calcular

$$\sum_{k=0}^n F(n, k) a^{n-k} b^k$$

# Triangulo de Pascal y la Potencia Binomial

El enunciado de este problema apareció en el número 209 de **Novática** (enero-febrero 2011, p. 76).

que es el valor de la expresión generalizada de la potencia binomial. Se proveen ambas alternativas en el código, y se comentan las líneas de código correspondientes a la segunda alternativa. Cabe destacar que entre ambas opciones es preferible la alternativa 1 ya que utiliza información conocida del problema y tiene una complejidad temporal de O(k) frente a la segunda alternativa cuya complejidad es O(n).

Finalmente, es necesario imprimir el valor de los coeficientes binomiales, para lo cual basta con recorrer la n-ésima fila de la matriz denominada "pascal" en el código fuente.

A continuación se provee el código fuente de la solución:

```

import java.util.Scanner;

public class Main
{
    static int[][] pascal = new int[10][10];
    public static void main(String[] args)
    {
        calcularTriangulo();
        int a, b, n;
        Scanner sc = new Scanner(System.in);

        while (sc.hasNextInt())
        {
            a = sc.nextInt();
            b = sc.nextInt();
            n = sc.nextInt();

            int suma = 0;

            //Alternativa 1
            suma=(int) Math.pow(a+b,n);

            //Alternativa 2:
            //for(int k = 0; k <= n; k++)
            // suma += (int) (pascal[n][k] *
            Math.pow(a,n-k) * Math.pow(b,k));

            System.out.print(suma);
            for (int k = 0; k <= n; k++)
                System.out.print(" " +
                pascal[n][k]);
            System.out.println();
        }
        return;
    }

    static void calcularTriangulo()
    {
        int i,j;
        pascal[0][0] = 1;
        for (i = 1; i < 10; i++)
        {
            pascal[i][0] = 1;
            for (j = 1; j < i; j++)
            {
                pascal[i][j] = pascal[i - 1][j -
                1] + pascal[i - 1][j];
            }
            pascal[i][j] = 1;
        }
    }
}
    
```

### Oferta de colaboración del Grupo de Lengua e Informática de ATI a los autores de Novática

El Grupo de Lengua e Informática de ATI, una de cuyas áreas de trabajo trata sobre el uso de la lengua en la Informática <<http://www.ati.es/spip.php?rubrique65>>, ofrece su colaboración a aquellos autores de **Novática** a quienes, a la hora de redactar sus artículos, les surjan dudas sobre cómo trasladar al castellano términos o expresiones procedentes de otras lenguas, y cuya traducción o adopción todavía no esté asentada en este dinámico ámbito de la terminología informática.

Para más información, por favor contactar con Carmen Ugarte <[cugarte@ati.es](mailto:cugarte@ati.es)>, coordinadora del citado Grupo de Trabajo.

### Cambios en Secciones Técnicas

Después de varios años contribuyendo a enriquecer **Novática** a través de la sección técnica "Informática y Filosofía", y en razón a que sus numerosos proyectos personales no le permiten seguir con esta dedicación, **Karim Gherab Martín** deja la labor de coordinación de la citada sección técnica.

Le sustituye **Roberto Feltrero Oreja**, profesor de la UNED, y cuya carta de presentación en **Novática** es el artículo que publicó en 2008 en nuestra revista ("*Prácticas Científicas y Sociedad del Conocimiento: el ejemplo de las comunidades FLOSS*" - **Novática** 192). Roberto compartirá a partir de ahora con **José Ángel Olivas Varela** las labores de coordinación de "Informática y Filosofía".

Karim, muchas gracias por tu dedicación y ¡mucha suerte en tus proyectos!

Roberto, bienvenido a este grupo de quienes contribuimos a hacer posible **Novática**. ¡Espero que estés muy a gusto entre nosotros!

### Programación de Novática

Por acuerdo de los Consejos Editoriales de **Novática** y **UPGRADE**, los temas y editores invitados de las monografías restantes de 2011 serán, salvo causas de fuerza mayor o imprevistos, los siguientes:

Nº 212 (julio-agosto): Número especial de verano sobre "Innovación y emprendimiento en las TIC".

Nº 213 (septiembre-octubre): "Informática verde". Editores invitados: **Juan Carlos López López** (Director del Grupo de Investigación ARCO, Universidad de Castilla-La Mancha), **Lasse Natvig** (Universidad de Ciencia y Tecnología de Noruega -NTNU-) y **Giovanna Sissa** (Universidad de Milán, Italia).

Nº 214 (noviembre-diciembre): "Gestión de riesgos". Editor invitado: **Darren Dalcher** (Middlesex University; Director del National

## Socios institucionales de ati

Según los Estatutos de ATI, pueden ser socios institucionales de nuestra asociación "*las personas jurídicas, públicas y privadas, que lo soliciten a la Junta Directiva General y sean aceptados como tales por la misma*".

Mediante esta figura asociativa, todos los profesionales y directivos informáticos de los socios institucionales pueden gozar de los beneficios de participar en las actividades de ATI, en especial congresos, jornadas, cursos, conferencias, charlas, etc. Asimismo los socios institucionales pueden acceder en condiciones especiales a servicios ofrecidos por la asociación tales como Bolsa de Trabajo, cursos a medida, *mailings*, publicidad en **Novática**, servicio ATInet, etc.

Para más información dirigirse a <[info@ati.es](mailto:info@ati.es)> o a cualquiera de las sedes de ATI. En la actualidad son socios institucionales de ATI las siguientes empresas y entidades:

AGENCIA DE INFOR. Y COMUN. COMUNIDAD DE MADRID  
AGROSEGURO, S.A.  
AIGÜES TER LLOBREGAT  
ALC ORGANIZACIÓN Y SISTEMAS,S.L.  
ALMIRALL, S.A.  
AVANTTIC, CONSULTORÍA TECNOLÓGICA, S.L.  
CENTRO DE ESTUDIOS VELAZQUEZ S.A. (C.E. Adams)  
CETICSA, CONSULTORIA Y FORMACIÓN  
CONSULTORES SAYMA, S.A.  
COSTAISA, S.A  
DEPARTAMENT D'ENSENYAMENT DE LA GENERALITAT  
ELOGOS, S. L.  
EPISER, S.L.  
ESPECIALIDADES ELÉCTRICAS, S.A. (ESPELSA)  
ESTEVE QUÍMICA, S.A.  
FUNDACIÓ BARCELONA MEDIA - UNIVERSITAT POMPEU FABRA  
FUNDACIÓ CATALANA DE L'ESPLAI  
FUNDACIÓ PRIVADA ESCOLES UNIVERSITÀRIES GIMBERNAT  
IN2  
INFORMÀTICA Y COMUNICACIONES AVANZADAS, S.L.  
INSTITUT D'ESTUDIS CATALANS  
INSTITUT MUNICIPAL D'INFORMÀTICA  
INVERGAMING GRUP  
KRITER SOFTWARE, S.L.  
MUSEU D'ART CONTEMPORANI DE BARCELONA - MACBA  
ONDATA INTERNATIONAL, S.L.  
PRACTIA CONSULTING, S.L.  
QRP MANAGEMENT METHODS INTERNATIONAL  
RCM SOFTWARE, S.L.  
SADIEL, S.A.  
SCATI LABS, S.A.  
SISTEMAS DATASIX  
SISTEMAS TÉCNICOS LOTERIAS ESTADO (STL)  
SOCIEDAD DE REDES ELECTRÓNICAS Y SERVICIOS, S.A.  
SQS, S.A  
TRAINING & ENTERPRISE RESOURCES  
T-SYSTEMS ITC Services España S.A.  
UNIVERSIDAD ANTONIO DE NEBRIJA  
UNIVERSITAT DE GIRONA  
UNIVERSITAT OBERTA DE CATALUNYA