

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).

<<http://www.ati.es/novatica/>>
<<http://www.ati.es/reicis/>>

ATI es miembro fundador de **CEPIS** (Council of European Professional Informatics Societies), representa a España en **IFIP** (International Federation for Information Processing) y es miembro de **CLEI** (Centro Latinoamericano de Estudios de Informática) y de **CECIA** (Confederation of European Computer Associations). Asimismo tiene un acuerdo de colaboración con **ACM** (Association for Computing Machinery) y colabora con diversas asociaciones informáticas españolas.

Consejo Editorial

Guillermo Alstina González, Pere Lluís Barabà, Miquel García-Menéndez (presidente del Consejo), Ernest Gijón Gil, Juan Hernández Basora, Silvia Leal Martín, David Moya Alvarez, Francesc Noguera Puig, Andrés Pérez Payeras, Víkto Pons i Colomer, Daniel Raya Demidoff, Jordi Roca i Marimon, Jorge Daniel Vigo López, Juan Carlos Vigo López

Coordinación Editorial

Llorenç Pagés Casas <lpages@ati.es>

Composición y autoedición

Impresión Olfset Derra S. L.

Traducciones

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

Administración

Tomás Brunete, María José Fernández, Enric Camarero

Secciones Técnicas - Coordinadores

Accesibilidad

Emmanuelle Gutiérrez y Restrepo (Fundación Sidar), <emmanuelle@sidar.org>

Loïc Martine Normand (Fundación Sidar), <loic@sidar.org>

Acceso y recuperación de la información

José María Gómez Hidalgo (Pragsis Technologies), <jmgomez@pragsis.com>

Enrique Puertas Sanz (Universidad Europea de Madrid), <enrique.puertas@universidadeuropea.es>

Administración Pública electrónica

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

Sabastià Justicia Pérez (Diputación de Barcelona) <sjusticia@ati.es>

Arquitecturas

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

José Flich Cardo (Universidad Politécnica de Valencia), <jflich@disca.upv.es>

Auditoría SITIC

Marina Tourinho Troitino, <marinatourinho@marinatourinho.com>

Sergio Gómez-Landero Pérez (Endesa), <sergio.gomezlandero@endesa.es>

Derecho y tecnologías

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

Enseñanza Universitaria de la Informática

Cristóbal Pareja Flores (DSIP-UCM), <cpajef@sis.ucm.es>

J. Ángel Velázquez Irujo (DLSI I, 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

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

Gobierno corporativo de las TI

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

Miguel García-Menéndez (ITI), <mgarciamenendez@ititrends.institute.org>

Informática y Filosofía

José Ángel Olivás Varela (Escuela Superior de Informática, UCLM), <joangeloliv@uclm.es>

Roberto Feltrero Oreja (UNED), <rfeltrero@gmail.com>

Informática Gráfica

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

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

Ingeniería del Software

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

Inteligencia Artificial

Vicente Botti 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), <obelform@lsi.uji.es>

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

Lingüística computacional

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

Modelado de software

Jesus Garcia Molina (DIS-UM), <jmolina@um.es>

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

Mundo estudiantil y jóvenes profesionales

Federico G. Mon Trotti (RITSJ), <fgm.fede@gmail.com>

Mikel Salazar Peña (Area de Jóvenes Profesionales, Junta de ATI Madrid), <mikelbo_uni@yahoo.es>

Seguridad

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

Miguel Sarrías Griño (ATI), <miguel@sarrias.net>

Redes y servicios telemáticos

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

Ana Pont Sanjuán (UPV), <apont@disca.upv.es>

Robótica

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

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

Seguridad

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

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

Sistemas de Tiempo Real

Alejandro Alonso Muñoz, Juan Antonio de la Puente Alfaro (DIT-UPM), <[@dit.upm.es](mailto:aalonso.jpunte)>

Software Libre

Jesús M. González Barahona (GSYC-URJC), <jgb@gsyc.es>

Fernando Tricas García (Universidad de Zaragoza), <tricas@unizar.es>

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 Vinas (Universitat de Girona), <didac.lopez@ati.es>

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

Tendencias tecnológicas

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

Juan Carlos Vigo (ATI), <juancarlosvigo@atinet.es>

TIC y Turismo

Andrés Agayo Maldonado, Antonio Guevara Plaza (Univ. de Málaga), <agayo.guevara@lcc.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 [CC BY](https://creativecommons.org/licenses/by/4.0/) o cualquier otra licencia de uso de contenidos, debiéndose en todo caso citar su procedencia y enviar a **Novática** un ejemplar de la publicación.

Coordinación Editorial, Redacción Central y Redacción ATi Madrid
Gutiérrez de Celina 24, 28017 Madrid • Tlf: 91 4029391 • novatica@ati.es

Administración y Redacción ATi Cataluña

Calle Avila 50, 3a planta, local 9, 08005 Barcelona
Tlf: 934 125 235 <secret@ati.es>

Redacción ATi Andalucía <secretand@ati.es>

Redacción ATi Galicia <secretgal@ati.es>

Suscripción y Ventas <novatica.subscripciones@atinet.es>

Publicidad Gutiérrez de Celina 24, 28017 Madrid
Tlf: 91 4029391 <novatica@ati.es>

Imprenta: Impresión Olfset Derra S.L., Lluís 41, 08005 Barcelona.

Depósito legal: B 15.154-1975 -- ISSN: 0211-2124; CODEN NOVAC

Portada: "La decisión" - Concha Arias Pérez / © ATI

Diseño: Fernando Agresta / © ATI 2003

editorial

La hora del Big Data

> 02

Periodicidad de Novática desde julio de 2016 hasta junio de 2017

noticias de ATI

Nombramiento de la nueva Directora de Novática

> 03

en resumen

Un agradecimiento muy especial para todos nuestros colaboradores

> 03

Llorenç Pagés Casas

noticias de IFIP

Asamblea General de IFIP

> 04

Ramon Puigjaner Trepal

WITFOR 2016

> 05

Ana Pont Sanjuán

Noticias del TC9: ICT and Society

> 05

Ignacio Gil Pechuán

Reunión anual del TC2 "Software: Theory and Practice"

> 06

Antonio Vallecillo Moreno

actividades de ATI

X Edición del Premio Novática: Entrega del premio al autor ganador

> 06

monografía

Big Data

Editores invitados: José María Gómez Hidalgo y Ricardo Baeza-Yates

Presentación. Big Data: Conceptos y aplicaciones

> 09

José María Gómez Hidalgo, Ricardo Baeza-Yates

Datos masivos en la Web

> 12

Ricardo Baeza-Yates

Big Data: Preprocesamiento y calidad de datos

> 17

Salvador García, Sergio Ramírez-Gallego, Julián Luengo, Francisco Herrera

Internet de las Cosas: La minería de flujos de datos masivos en tiempo real

> 24

Albert Bifet, Jesse Read

Análisis Big Data en sistemas de computación de alto rendimiento: Tecnologías, herramientas y ejemplos

> 31

Alexey Cheptsov, Bastian Koller

Big Data y sistemas de recomendación

> 39

David C. Anastasiu, Evangelia Christakopoulou, Shaden Smith, Mohit Sharma, George Karypis

Estudio sobre la escalabilidad del algoritmo de agrupamiento estructural paralelo para redes en Big Data

> 46

Weizhong Zhao, Gang Chen, Venkata Swamy-Martha, Xiaowei Xu

Introducción a la analítica de texto con Spark

> 53

José María Gómez Hidalgo

Cómo mejorar el conocimiento de tu audiencia: Experiencias de la CCMA en un entorno Big Data

> 60

Xavier Ferrándiz Bofill, Alberto Alejo Marcos

Privacidad en la analítica masiva de datos

> 65

José María del Álamo Ramiro, Esmeralda Saracibar Serradilla, Emilio Aced Féliz

secciones técnicas

Tendencias Tecnológicas

¿Nos está haciendo felices la tecnología?

> 70

Dorian Peters

Referencias autorizadas

> 72

sociedad de la información

Programar es crear

El problema del robot de exploración de Marte

> 78

(Competencia UTN-FRC 2014, problema 5, enunciado)

Julio Javier Castillo, Diego Javier Serrano, Marina Elizabeth Cárdenas

Discos duros

> 79

(Competencia UTN-FRC 2015, problema A, solución)

Julio Javier Castillo, Diego Javier Serrano, Marina Elizabeth Cárdenas

asuntos interiores

Coordinación editorial / Programación de Novática / Socios Institucionales

> 80

Monografía del próximo número: "Seguridad digital"

La hora del *Big Data*

Informática puede venir de **información**, y la información se nutre de **datos**. Los datos en sí mismos suelen ser poco significativos dado que su valor se manifiesta cuando los relacionamos, porque es entonces cuando se transforman en información.

Esta reflexión, posiblemente elemental, configura en parte la historia de nuestra informática, que consiste, entre otros aspectos, en ofrecer información elaborada. Aunque para completar el objetivo que todo proceso informático requiere faltará un tercer elemento que es la **tecnología**, absolutamente necesaria para poder acceder a la información de forma eficiente.

En esta etapa actual de nuestra profesión impera la necesidad de obtener información eficiente y en este cometido toma papel predominante la cultura "*Big Data*". Por ello, el posicionamiento de los profesionales de la Informática que necesitan estar al corriente de las novedades del sector y de sus posibilidades, se debe orientar a entender la utilidad del "*Big Data*" en la empresa, y a traducir en términos de rentabilidad las posibilidades de su aplicación como concepto, así como también conocer el uso de la tecnología que acompaña al mismo. Y esto requiere formación y recursos.

¿Más formación y más recursos todavía? dirán algunos informáticos seniors, hartos del cambio tecnológico continuo. Y así es. En nuestra profesión, lo cotidiano es cambiar por lo general, nuestras competencias en sistemas operativos, lenguajes de programación y conocimientos de estructuras de información, todo ello a muy corto plazo, y sin tregua hacia un recorrido que no tiene por qué tener un final.

Las estructuras de datos siempre han sido motor de cambio en nuestra profesión. De los sistemas de ficheros a los modelos de bases de datos, de las primeras estructuras, jerárquicas o Codasyl, a las aceptadas como más eficaces, y basadas en el modelo relacional. Ésto solo fue un inicio que evolucionó en términos de eficacia, ya que aparecieron los modelos documentales y los orientados a objetos. Todo ello en paralelo con la evolución del hardware y los lenguajes de programación.

El "*Big Data*" es el protagonista del momento actual en gestión de los datos, porque dispone de muchas herramientas para tratar la información de forma masiva, que es lo que define este nuevo concepto. Y por ello se va a imponer probablemente como una de las herramientas más útiles para tomar decisiones y prever futuras operaciones de largo alcance relacionadas con la información obtenida.

Pero existe otra faceta que suscita reflexión sobre esta nueva tecnología, que es la gestión de unos datos, donde va a ser difícil averiguar si los resultados obtenidos se dan a conocer por el gran público, y si atentan a la seguridad de las personas, empresas y de la sociedad en general. Y también si los datos manejados son obtenidos con el consentimiento de su uso por parte de los ciudadanos implicados.

Si las reglas de juego son laxas, el poseedor de la información elaborada a través de procesos de tratamiento de información de forma masiva, y que además también domina parte de determinadas redes sociales, nos puede llegar a conocer y manejar sin que seamos conscientes de ello, con más intensidad de lo que ya lo debe estar haciendo actualmente.

No hay que ser demasiado observador para ver paseando por nuestras calles ciudadanos que realizan, por lo general, una actividad casi frenética: llenar de forma continua y a través de sus *smartphones* ficheros de datos e imágenes, algunos a bases de datos privadas, y otros a determinados bancos de datos. Y lo que se está publicando en estos bancos de datos son referencias a quienes somos, que nos apetece, cuáles son nuestras intenciones personales a corto y medio plazo, y por qué estaríamos dispuestos a jugarlos el tipo. Todo un manjar para los servicios de inteligencia y las empresas que comercializan productos dirigidos al gran público.

Por lo demás, centrándonos más ya en la evolución de los productos y conceptos para tratar la información, y aprendiendo del curso de la Informática y de nuestra experiencia como profesionales, el "*Big Data*" no tiene por qué ser el final de ninguna historia, ya que sabemos que lo único estable que tiene la gestión de datos es su evolución continuada.

Podemos imaginar ya un futuro de evolución del "*Big Data*" con la aparición de nuevos conceptos, del tipo: "*Very Large Big Data*" o "*Super Big Data*" que pueden estar a la vuelta de la esquina.

La Junta Directiva General de ATI

Periodicidad de Novática desde julio de 2016 hasta junio de 2017

Con la publicación de este número, el cual significa la despedida de Llorenç Pagés Casas, quien ha venido dirigiendo la revista desde julio de 2006, se inicia una fase de remodelación de la revista, acorde con los nuevos tiempos y con los retos presupuestarios que debe afrontar ATI.

Para facilitar esta transición, ya iniciada desde el nombramiento de Encarna Quesada Ruiz como nueva

directora de la revista, y poder realizar mejor todas las adecuaciones necesarias, hemos decidido cambiar provisionalmente la periodicidad de publicación de la revista, pasando a ser ésta cuatrimestral entre julio de 2016 y junio de 2017. Por lo tanto, en este periodo además del número actual se publicarán otros dos números más: Novática 238 (noviembre 2016-febrero 2017) y Novática 239 (marzo 2017-junio 2017).

Es de preveer que a partir de julio de 2017 podamos reanudar la periodicidad trimestral que nos ha sido habitual desde 2012, de modo que Novática 240 y Novática 241 se puedan publicar también a lo largo de 2017.

Nombramiento de la nueva Directora de Novática

Nos complace informaros de que en la última reunión de Junta Directiva General, y tras el pertinente concurso, celebrado en fechas recientes, se acordó nombrar a **Encarna Quesada Ruiz** como nueva Directora de *Novática*, en sustitución de **Llorenç Pagés Casas**, después de que éste comunicase a esta JDG, hace unos meses, su intención de ir cerrando un ciclo que ya se prolongaba por espacio de diez años.

Encarna Quesada es especialista en procesos de Transformación Digital. La investigación, la divulgación y el asesoramiento son los tres ejes principales de su labor, en relación a los avances e innovación en tecnolo-

gía que afectan tanto a las empresas, como a la sociedad en su conjunto.

Encarna fue Directora de W3C España, un consorcio internacional que trabaja en el desarrollo de protocolos y pautas que permiten asegurar el crecimiento futuro de la Web y hacer posible la comunicación humana, el comercio y las oportunidades para compartir conocimiento. Asimismo, ha trabajado con una serie de organismos y empresas nacionales e internacionales en proyectos de transformación digital, lo que le ha permitido adquirir una visión estratégica en innovación tecnológica, en diferentes sectores.

De forma constante, Encarna investiga y analiza la evolución de la tecnología, asesorando y tomando parte activa en proyectos empresariales para la transformación tecnológica de procesos, negocios y personas.

Finalmente, no queremos dejar pasar la ocasión sin agradecer a Llorenç su labor al frente de *Novática* durante todos estos años. Vayan para él nuestros mejores deseos para cuantos retos, personales y profesionales, aborde de aquí en adelante.

La Junta Directiva General de ATI

En resumen Un agradecimiento muy especial para todos nuestros colaboradores

Llorenç Pagés Casas
Coordinación Editorial de *Novática*

Cuando en el año 2006, **Rafael Fernández Calvo** me dio a conocer que me consideraba un candidato idóneo para sucederle al frente de *Novática*, he de confesar que me dio un vuelco el corazón al tiempo que se abrían ante mí un montón de ilusiones y buenas expectativas. Se me vino a la mente en aquel momento el modo en el que, a finales de los años 1970 y sin haber cumplido todavía los 18 años, esperaba con devoción la llegada vía postal de la revista a casa para devorarla con fruición tratando de descubrir las claves de esa “mitad ciencia-mitad arte” que empezaba ya a ser mi profesión.

Una vez que acepté ese para mí increíble reto y pude ganar el pertinente concurso, he tenido la suerte y el privilegio de haber podido dirigir la revista en un periodo tan largo como son algo más de 10 años. En los cuales hemos podido publicar, contando con éste, 56 números de la revista, más de la mitad de los cuales en conjunción con **UP-GRADE**, una revista promovida por CEPIS donde publicábamos básicamente una versión inglesa del monográfico de *Novática*.

Han sido para mí años de intenso trabajo y mucha dedicación procurando dar lo mejor de mí mismo para mantener ese nivel y prestigio en la revista que le habían ido otorgando con el paso del tiempo todos y cada uno de mis antecesores. Ahora, en el momento de la despedida, puedo decir que, aun a pesar de que quedaron muchas cosas en el tintero, y otras muy mejorables en diversos aspectos, me voy con la satisfacción del deber cumplido ante ese reto que solo a base de entusiasmo e ilusión he conseguido que no me viniera excesivamente grande.

Pero me despido sobre todo con la enorme satisfacción de haber conocido y tratado

con personas maravillosas que me han facilitado muchísimo el trabajo y sin los cuales todo ese entusiasmo e ilusión no hubieran servido para nada. Así, he podido coincidir tanto con excelentes profesionales como con muchas personas que voluntariamente han cedido su tiempo en beneficio de que *Novática* publicase unos contenidos mejores y más adecuados. Ha sido para mí una experiencia impagable en este sentido.

Empezando por el principio, solo puedo tener palabras de agradecimiento para **Fernando Piera** y **Rafael Fernández Calvo**, quienes me dieron “acogida” en el proyecto en su calidad, por entonces, de vocal de *Novática* en la JDG y antecesor mío, respectivamente. Posteriormente, Rafael ha continuado colaborando con *Novática* hasta fechas bien recientes en su calidad tanto de mentor, colaborador y en muchos casos inspirador mío, haciendo que ese gratísimo recuerdo vaya a permanecer eternamente en mi memoria.

Por otra parte, tuve también la suerte de trabajar con **Jorge Llácer** (q.e.p.d) una excelente persona de quien uno piensa que la vida no lo trató como él merecía. Jorge fue maquetador de la revista durante más de 20 años y cuidó perfectamente de todos los detalles como gran profesional que era.

Otra magnífica profesional es **Concha Arias**, portadista de la revista con la misma antigüedad que la mía, y que ha tenido la virtud de irme sorprendiendo día tras día, portada tras portada, con su ingenio, imaginación y buen hacer. Estoy seguro de que, en virtud de su trabajo, nuestros lectores aprecian hoy especialmente la calidad de nuestras portadas.

He de agradecer también el apoyo encontrado en las secretarías, a lo largo de todos estos años, por parte de **Enric, Feli, Pepa** y **Tomàs**, así como en Derra, S.L., con especial mención para **Josep** y **Lola**, siendo ésta

la empresa que cuidó inicialmente de la impresión de la revista, para posteriormente hacerse cargo de la maquetación, en ausencia de Jorge.

Y finalmente, los voluntarios. Esas personas que han conseguido sacar tiempo de sus quehaceres habituales para favorecer la calidad de nuestros contenidos. Especialmente mencionable es el grupo de coordinadores de sección que ha cumplido con la misión tanto de cuidar de las revisiones de artículos no solicitados como de redactar esas impagables “referencias autorizadas” que aparecen regularmente en cada número. Así como el grupo de traductores voluntarios que, con el paso del tiempo, se ha ido nutriendo de extraordinarios colaboradores, y sin cuyo trabajo no hubiéramos podido publicar los varios centenares de artículos que nos llegaron originalmente en inglés.

A todas esas personas maravillosas, muchas gracias por vuestra colaboración.

Ahora, la revista queda en unas excelentes manos. Por una parte, mi sucesora, **Encarna Quesada Ruiz**, una gran profesional, experta en comunicación, y por otra, **Miguel García Menéndez**, en su calidad de vocal de *Novática* en la JDG. Ambos han sido magníficos colaboradores de la revista durante años, y por lo tanto están preparados para los grandes retos que afronta *Novática* en el futuro. Como ya saben ellos, siempre podrán contar con mi modesta ayuda, allá donde sea necesario.

Un fuerte abrazo a todos,