

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 **CECUA** (*Confederación 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

Guillem Alstina González, Pere Lluis Barà, 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 Marimón, Jorge Daniel Vigo López, Juan Carlos Vigo López

Coordinación Editorial

Llorenç Pagés Casas <pages@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 Martínez Normand (Fundación Sidar), <loic@sidar.org>

Acceso y recuperación de la información

José María Gómez Hidalgo (Praxis Technologies), <jmgomez@praxis.com>

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

Administración Pública electrónica

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

Sebastià 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@marinatourino.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), <cpareja@sip.ucm.es>

J. Angel Velázquez Irujo (DL SI I, URJC), <angel.velazquez@urj.es>

Entorno digital personal

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

Diego Gachet Paez (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), <jbaiget@ati.es>

Gobierno corporativo de las TI

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

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

Informática y Filosofía

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

Roberto Feltre Oreja (UNED), <rfeltre@gmail.com>

Informática Gráfica

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

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

Ingeniería del Software

Luis Fernández Sanz, Daniel Rodríguez García (Universidad de Alcalá),

<luis.fernandezsan@uah.es>

Inteligencia Artificial

Vicente Boti Navarro, Vicente Julián Inglada (DSIC-UPV), <vbotti.vinglada@disic.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), <ougarte@ati.es>

Lenguajes Informáticos

Oscar Belmonte Fernández (Univ. Jaime I de Castellón), <obelfern@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 Róssa (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 (Área de Jóvenes Profesionales, Junta de ATI Madrid), <mikelbo_uni@yahoo.es>

Seguridad

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

Miguel Sarrías Griño (ATI), <miquel@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), <joscortesa@gmail.com>

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

Seguridad

Javier Arellito 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),

<aalonso.jpunte@dit.upm.es>

Software Libre

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

Fernando Tricas García (Universidad de Zaragoza), <ftricas@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) <aalg@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@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» o «p» indicadas 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
Gutiérrez de Cetina 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 125235 <secret@ati.es>

Redacción ATI Andalucía <secretand@ati.es>

Redacción ATI Galicia <secretgal@ati.es>

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

Publicidad Gutiérrez de Cetina 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 NOVACV

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 Trepast

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 Félez

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