

**Novática**, founded in 1975, is the oldest periodical publication amongst those specialized in Information and Communication Technology (ICT) existing today in Spain. It is published by **ATI (Asociación de Técnicos de Informática)** which also publishes **REICIS (Revista Española de Innovación, Calidad e Ingeniería del Software)**.

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

ATI is a founding member of **CEPIS (Council of European Professional Informatics Societies)**, the Spain's representative in **IFIP (International Federation for Information Processing)**, and a member of **CLEI (Centro Latinoamericano de Estudios en Informática)** and **CECUA (Confederation of European Computer User Associations)**. It has a collaboration agreement with **ACM (Association for Computing Machinery)** as well as with diverse Spanish organisations in the ICT field.

#### Editorial Board

Guillem Alsina González, Rafael Fernández Calvo (presidente del Consejo), Jaime Fernández Martínez, Luis Fernández Sanz, José Antonio Gutiérrez de Mesa, Silvia Leal Martín, Didac López Vilas, Francesc Noguera Puig, Joan Antoni Pastor Collado, Viktu Pons i Colomer, Moisés Robles Gener, Cristina Vigil Díaz, Juan Carlos Vigo López

#### Chief Editor

Llorenç Pagés Casas <pages@ati.es>

#### Layout

Jorge Llacer Gil de Ranales

#### Translations

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

#### Administration

Tomás Brunete, María José Fernández, Enric Camarero

#### Section Editors

##### Artificial Intelligence

Vicente Botti Navarro, Julián Inglada (DSIC-UPV), <{vbotti,viglada}@dsic.upv.es>

##### Computational Linguistics

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

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

##### Computer Architecture

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

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

##### Computer Graphics

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

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

##### Computer Languages

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

Inmaculada Coma Talsy (Univ. de Valencia), <Inmaculada.Coma@uv.es>

##### e-Government

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

Sebastià Justícia Pérez (Diputació de Barcelona) <sjusticia@ati.es>

##### Free Software

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

Israel Herráiz Tabernero (Universidad Politécnica de Madrid), <isra@herraiiz.org>

##### Human-Computer Interaction

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

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

##### ICT and Tourism

Andrés Aguayo Maldonado, Antonio Guevara Plaza (Universidad de Málaga),

<{aguayo, guevara}@cc.uma.es>

##### Informatics and Philosophy

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

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

##### Informatics Profession

Rafael Fernández Calvo (ATI), <rfcalvo@ati.es>, Miquel Sarriés Grijó (ATI), <miquel@ati.es>

##### Information Access and Retrieval

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

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

##### Information Systems Auditing

Marina Tourinho Troitino, <marinatourino@marinatourino.com>

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

##### IT Governance

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

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

##### Knowledge Management

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

##### Language and Informatics

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

##### Law and Technology

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

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

##### Networking and Telematic Services

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

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

##### Personal Digital Environment

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

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

##### Software Modeling

Jesús García Molina (DIS-UM), <jmolina@um.es>

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

##### Students' World

Federico G. Mon Trotti (RITSI), <gnu.fede@gmail.com>

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

##### Real Time Systems

Alejandro Alonso Muñoz, Juan Antonio de la Puente Alfaro (DIT-UPM),

<{aalonso,puente}@dit.upm.es>

##### Robotics

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

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

##### Security

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

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

##### Software Engineering

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

##### Technologies and Business

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

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

##### Technologies for Education

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

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

##### Teaching of Computer Science

Cristóbal Pareja Flores (DSIP-UCM), <cpareja@sip.ucm.es>

J. Angel Velázquez Ilurbide (DLSI I, URJC), <angel.velazquez@urjc.es>

##### Technological Trends

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

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

##### Web Standards

Encarna Quesada Ruiz (Virali), <encarna.quesada@gmail.com>

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

#### Copyright © ATI 2014

The opinions expressed by the authors are their exclusive responsibility

#### Editorial Office, Advertising and Madrid Office

Plaza de España 6, 2ª planta, 28008 Madrid

Tlf. 91 4029391; fax 91 3093685 <novatica@ati.es>

#### Layout and Comunidad Valenciana Office

Av. del Reino de Valencia 23, 46005 Valencia

Tlf. 963740173 <novatica\_val@ati.es>

#### Accounting, Subscriptions and Catalonia Office

Calle Avila 48-50, 3a planta, local 9, 08005 Barcelona

Tlf. 934125235; fax 934127713 <secregen@ati.es>

#### Andalucía Office

<secreand@ati.es>

#### Galicia Office

<secregal@ati.es>

#### Subscriptions and Sales

<novatica.subscriptions@atinet.es>

#### Advertising

Plaza de España 6, 2ª planta, 28008 Madrid

Tlf. 91 4029391; fax 91 3093685 <novatica@ati.es>

**Legal deposit:** B-15-154-1975 - ISSN: 0211-2124. CODEN NOVAEC

**Cover Page:** Mineral, Vegetable, Animal - Concha Arias Pérez / © ATI

**Layout Design:** Fernando Agresta / © ATI 2003

#### editorial

##### ATI: Boosting the Future

> 02

##### From the Chief Editor 'Pen

##### Process Mining: Taking Advantage of Information Overload

> 02

Llorenç Pagés Casas

#### monograph

##### Process Mining

Guest Editors: Antonio Valle-Salas and Anne Rozinat

##### Presentation. Introduction to Process Mining

> 04

Antonio Valle-Salas, Anne Rozinat

##### Process Mining: The Objectification of Gut Instinct - Making Business

##### Processes More Transparent Through Data Analysis

> 06

Anne Rozinat, Wil van der Aalst

##### Process Mining: X-Ray Your Business Processes

> 10

Wil van der Aalst

##### The Process Discovery Journey

> 18

Josep Carmona

##### Using Process Mining in ITSM

> 22

Antonio Valle-Salas

##### Process Mining-Driven Optimization of a Consumer Loan Approvals Process

> 30

Arjel Bautista, Lalit Wangikar, S.M. Kumail Akbar

##### Detection of Temporal Changes in Business Processes

##### Using Clustering Techniques

> 39

Daniela Luengo, Marcos Sepúlveda

Antonio Valle-Salas<sup>1</sup>, Anne Rozinat<sup>2</sup>

Managing Partner of G2; Co-founder of Fluxicon

<avalle@gedos.es>, <anne@fluxicon.com>

# Presentation. Introduction to Process Mining

During the last few decades information technology has touched every part of our lives. From cellular phones to the most advanced medical information processing systems, vending machines and PLCs in production lines, computerized components are everywhere. All these components generate vast amounts of information that are growing exponentially. Relatively few years ago the challenge was finding digitized information, whereas now the problem is being able to process and give meaning to all the information we generate.

In recent years we have seen how the information analysis industry has proposed various approaches to this problem. Some of them have been covered in one way or another in previous editions of *Novática*: starting with *VLDB* (Very Large Databases) in volume 91 (1991) and *Datawarehouse* approaches attempting to discover patterns in these data stores with *Data Mining* in volume 138 (1999), then followed *Knowledge Management* in volume 155 (2002). We realized how complex the problem was in the monograph on *The Internet of Things* in volume 209 (2011) and how we could exploit this information in volume 211 on *Business Intelligence* (2011). Finally, the industry is also moving in a direction not yet covered in *Novática* but certain to be addressed in the near future: *Big Data*.

In the present volume of *Novática* we address a particularly interesting topic within this broad range of techniques for data analysis: *Process Mining*. This is a variant of data mining in which we focus on analyzing the information generated by the processes that have been computerized and whose executions have been traced. As **Anne Rozinat** and **Wil van der Aalst** explain in the opening article, we will see that the first traces are found in the late nineteenth century, although in terms of modern science we refer to the seminal work of Myhill / Nerod (1958), or Viterbi algorithms (1978).

In the late 90s there were already some specific research teams in universities around the world, especially the University of Colorado and *Technische Universiteit Eindhoven*. These teams developed their research defining algorithms and methods that allow the treatment of process execution traces for

### Guest Editors

**Antonio Valle-Salas** is Managing Partner of G2 and a specialist consultant in ITSM (Information Technology Service Management) and IT Governance. He graduated as a Technical Engineer in Management Informatics from UPC (*Universitat Politècnica de Catalunya*) and holds a number of methodology certifications such as ITIL Service Manager from EXIN (Examination Institute for Information Science), Certified Information Systems Auditor (CISA) from ISACA, and COBIT Based IT Governance Foundations from IT Governance Network, plus more technical certifications in the HP Openview family of management tools. He is a regular collaborator with itSMF (IT Service Management Forum) Spain and its Catalan chapter, and combines consulting and project implementation activities with frequent collaborations in educational activities in a university setting (such as UPC or the *Universitat Pompeu Fabra*) and in the world of publishing in which he has collaborated on such publications as *IT Governance: a Pocket Guide*, *Metrics in IT Service Organizations*, *Gestión de Servicios TI. Una introducción a ITIL*, and the translations into Spanish of the books *ITIL V2 Service Support* and *ITIL V2 Service Delivery*.

**Anne Rozinat** has more than ten years of experience with process mining technology and obtained her PhD cum laude in the process mining group of Prof. Wil van der Aalst at the Eindhoven University of Technology in the Netherlands. Currently, she is a co-founder of Fluxicon and blogs at <<http://www.fluxicon.com/blog/>>.

discovery, analysis and representation of the underlying processes.

However no tools that implemented these algorithms with appropriate degrees of usability had yet reached the market. By the end of 2003 the *processmining.org* specialized community (a working group of the TU/e) was created, and in early 2004 the first version of ProM was developed, a generic and open source framework for process mining that has become the primary tool for researchers and analysts, now at version 6.3 and including more than 500 plugins that implement state of the art in this field.

In 2009 a Task Force of the IEEE focused on process mining was created that now has members from over 20 countries including software vendors (such as Software AG, HP, IBM or Fluxicon, among many others), consulting firms and analysts (Process Sphere, Gartner and Deloitte, among others) and a wide range of educational and research institutions (TU/e, Universitat Politècnica de Catalunya or Universität zu Berlin, to name but a few). One of the key objectives of this task force is to spread the concepts, techniques and benefits of process mining. In 2011 they published the *Process Mining Manifesto*, a document signed by more than 50 professionals translated into 12 languages. We cannot include here the full text of the manifesto, but the reader will find the reference in the links section of this monograph.

For this present edition of *Novática* we have been privileged to have a group of authors that give us different perspectives on the matter. We begin with an introductory article in which **Anne Rozinat** and **Wil van der Aalst** set the context for process mining concepts and state, in a very enlightening process mining message, that it allows us to get an objective vision of our processes.

In the second paper **Wil van der Aalst** guides us through the different uses we can make of process mining: to create of a model of the process, to check the compliance of the model or to improve an existing model. Here another key message is presented: the use of process mining as X-rays that allow us to see "inside" the process, based on the analysis of real data from the execution of all cases (as opposed to the statistical sampling we would do in an audit, for example).

In the next article you will find the **Josep Carmona's** vision of the task of discovering a process from its traces. Here Josep makes an entertaining approach to how we could use the mining process to decrypt the message of an alien explaining his visit to planet Earth, while showing us the anatomy of the discovery process.

The introductory papers will be followed by a set of articles focusing on case studies. First **Antonio Valle-Salas'** article presents an application of process mining in a specific industry, focusing on the processes in an IT

Department and showing the different uses we can make of these techniques in the world of IT Service Management (ITSM)

Then **Arjel D. Bautista**, **Lalit M. Wangikar** and **Syed Kumail Akbar** present the work done to optimize the loan approval process of a Dutch bank, a remarkable work that was awarded the BPI Challenge 2012 prize.

Finally **Daniela Luengo** and **Marcos Sepúlveda** give us a research perspective on one of the challenges stated in the manifesto: dealing with the concept drift. This term is used to refer to the situation in which the process is changing while it is being used. Detecting these changes and including these features in the analysis is essential when working on rapidly changing environments because, otherwise, it could lead to erroneous conclusions in analysis.

These authors have contributed with their articles to give a clearer vision of what process mining is, what it is useful for and what its future is. Process mining is a relatively new science but is already reaching the level of maturity required to become standard practice in companies and organizations, as reflected in the articles' practical uses. However, there are still many challenges ahead and a long way to go: Will we be able to overcome the problems introduced by the concept drift? Can we use process mining not only for knowing the past of a process but also to predict its future? Will we implement these techniques in the management systems of business processes in order to provide them with predictive systems or support operators?

We are sure we will see great advances in this area in the near future.

## Useful References of "Process Mining"

In addition to the materials referenced by the authors in their articles, we offer the following ones for those who wish to dig deeper into the topics covered by the monograph:

■ **W.M.P. van der Aalst**. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer Verlag, 2011. ISBN 978-3-642-19344-6.

■ **IEEE Task Force on Process Mining**. *Process Mining Manifesto* (en 12 idiomas). <[http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:process\\_mining\\_manifesto](http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:process_mining_manifesto)>.

■ **Fluxicon TU/eProcess Mining Group**. *Introduction to Process Mining: turning (big) data into value* (video). <[http://www.youtube.com/watch?v=7oat7MatU\\_U](http://www.youtube.com/watch?v=7oat7MatU_U)>.

■ **Fluxicon**. *Process Mining News*. <<http://fluxicon.com/s/newsarchive>>.

■ **TU/e Workgroup**. <<http://www.processmining.org>>.

■ **Fluxicon**. *Process Mining Blog*. <<http://fluxicon.com/blog/>>.

■ **IEEE Task Force on Process Mining**. <<http://www.win.tue.nl/ieeetfpm/doku.php?id=start>>.

■ **LinkedIn**. *Process Mining* (community) <<http://www.linkedin.com/groups/Process-Mining-1915049>>.

■ **TU/e**. *Health Analytics Using Process Mining*. <<http://www.healthcare-analytics-process-mining.org>>.