



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

## SUBPROGRAMA RAMON Y CAJAL CONVOCATORIA 2009

Nombre: MANEVA, ELITZA

Referencia: RYC-2009-03975

Area: Ciencias de la Computación y Tecnología Informática

Correo electrónico: elitza.maneva@gmail.com

### Título:

Algoritmos de paso de mensajes

### Resumen de la Memoria:

La línea principal de investigación de esta propuesta es el desarrollo de nuevas técnicas para diseño y análisis de algoritmos dirigido al avance de las tecnologías de la información y sus aplicaciones en las ciencias. En concreto, las nuevas técnicas que serán investigadas son los algoritmos de paso de mensajes y métodos híbridos entre paso de mensajes y otras técnicas iterativas como búsqueda local y cadenas de Markov. Aunque algoritmos de paso de mensajes como por ejemplo belief-propagation existen desde los años 60 [Gallager62, Pearl88], éstos han recibido mucha atención durante los últimos 10 años por dos razones importantes. Primero, el diseño de códigos para corrección de errores gráficos como LDPC y Turbo, con sus algoritmos de corrección que son de paso de mensajes, han revolucionado recientemente el área de corrección de errores, llegando hasta muy cerca del límite de Shannon [RSU01, BG96]. Segundo, una técnica de paso de mensajes muy sofisticada llamada Survey Propagation fue introducida para resolver casos aleatorios muy grandes del problema de satisfacción con densidad de constricciones mucho más alta que era posible antes [MPZ02]. Éste es basado en métodos de física estadística donde la técnica de paso de mensajes fue descubierta independiente. Como consecuencia, una nueva y muy vigorosa área de investigación ha sido introducida en la frontera entre la teoría de la computación, física estadística, probabilidad, y teoría de la información. Este área tiene por objetivo construir una teoría de los métodos iterativos y extender las aplicaciones de estas técnicas. Los algoritmos considerados serán investigados en un escenario tan general como sea posible para conseguir un profundo conocimiento teórico de sus propiedades fundamentales y de su potencial. Un escenario teórico extensamente aceptado que captura simultáneamente problemas en verificación de hardware, en comunicación, y en física y biología, es el de sistemas de constricciones y de interacciones. La metodología de investigación es la de demostración matemática. Esta metodología complementará los métodos de investigación que se usan en otras áreas como comunicación, inteligencia artificial y física estadística donde estas técnicas algorítmicas están investigadas con métodos empíricos y heurísticos. Específicamente, el proyecto aprovechará sofisticados métodos formales desarrollados en la teoría de la probabilidad y la informática teórica, como por ejemplo el método de los momentos, reconstrucción en árboles, algoritmos de aproximación y apareamiento de cadenas de Markov. [Gallager62] R. Gallager, ¿Low-density parity-check codes¿, IRE Transactions on Information Theory, Vol. 8, Issue 1, pages 21-28, 1962. [Pearl88] J. Pearl, ¿Probabilistic Reasoning in Intelligent Systems¿, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. [RSU01] T. Richardson, A. Shokrollahi, R. Urbanke, ¿Design of capacity-approaching irregular low-density parity-check codes¿, IEEE Transactions on Information Theory, Vol. 47, Issue 2, pages 619-637, 2001. [BG96] Berrou, Glavieux ¿Near-Optimum Error-Correcting Coding and Decoding: Turbo-Codes¿, IEEE Transactions on Communications, Vol. 44, Issue 10, pages 1261-1272, 1996. [MPZ02] M. Mezard, G. Parisi, R. Zecchina, ¿Analytic and Algorithmic Solution of Random Satisfiability Problems¿, Science, Vol. 297, no. 5582, pages 812-815, 2002.

### Resumen del Curriculum Vitae:

1. Education: University of California at Berkeley, PhD in Computer Science, Aug 2001 - Aug 2006 Designated Emphasis in Communication, Computation and Statistics Thesis: "Belief Propagation Algorithms for Constraint Satisfaction Problems" Advisor: Alistair Sinclair California Institute of Technology (Caltech), BS in Engineering and Applied Science, Sep 1997 - Jun 2001 Thesis: "Interactive Communication on Noisy Channels", Advisor: Leonard J. Schulman  
2. Experience: Universitat Politecnica de Catalunya, Barcelona Visiting Professor, Sep 2008 - present IBM Almaden Research Center, California Postdoctoral Researcher, Oct 2006 - Jul 2008 Research Internships: IBM Almaden Research Center, California, May 2005 - Dec 2005 IBM T.J. Watson Research Center, New York, Jun 2000 - Sep 2000, IBM T.J. Watson Research Center, New York, Jun 1999 - Sep 1999 Bates Linear Accelerator Center, Massachusetts, Jun 1998 - Sep 1998 Teaching Assistantships: UC Berkeley, Fall 2004, Spring 2006 Caltech, 1999/2000, 2000/2001.  
3. Publications: P. Gopalan, P. Kolaitis, E. Maneva and C. Papadimitriou "The Connectivity of Boolean Satisfiability: Computational and Structural Dichotomies" SIAM Journal on Computing, 2009, in press. Extended abstract in Proceedings of ICALP 2006. F. Ardila and E. Maneva "Pruning Processes and a New Characterization of Convex Geometries" Discrete Mathematics, 2009, in press. E. Maneva and A. Sinclair "On the Satisfiability Threshold and Clustering of Solutions of Random 3-SAT Formulas" Theoretical Computer Science, vol. 407, pp. 359-369, 2008. E. Maneva, E. Mossel and M. Wainwright "A New Look at Survey Propagation and its Generalizations" Journal of ACM, Volume 54(4), pp. 2-41, 2007. Extended abstract in Proceedings of SODA, pp. 1089-1098, 2005. E. Maneva, T. Meltzer, J. Raymond, A. Sportiello and L. Zdeborova "A Hike in the Phases of the 1-in-3 Satisfiability" Lecture notes of the Les Houches Summer School 2006, Session LXXXV, Complex Systems, Volume 85, pp. 491-498, Elsevier 2007. E. Maneva and A. Shokrollahi "New Model for Rigorous Analysis of LT codes" IEEE International Symposium on Information Theory (ISIT), pp. 2677-2679, 2006. M. Wainwright and E. Maneva "Lossy Source Encoding via Message-passing and Decimation over Generalized Codewords of LDGM Codes" IEEE ISIT, pp. 1493-1497, 2005. A. Bogdanov, E. Maneva and S. Riesenfeld "Power-aware Base Station Positioning for Sensor Networks" 23rd Conference of the IEEE Communications Society (INFOCOM), pp. 575-585, 2004. A. Fabrikant, A. Luthra, E. Maneva, C. Papadimitriou and S. Shenker "On a Network Creation Game" 22nd ACM Symposium on Principles of Distributed Computing (PODC), pp. 347-351, 2003. E. Maneva and J. Smolin "Improved Two-party and Multi-party Purification Protocols" AMS Contemporary Mathematics Series, Quantum Computation and Quantum Information Science, vol 305, pp. 203-212, 2002. SAMPLE Collaboration: S.P. Wells et al. "Measurement of the vector analyzing power in elastic electron-proton scattering as a probe of the double virtual Compton amplitude" Phys. Rev. E, vol 63, Issue 6, id 064001, 2001. SAMPLE Collaboration: D.T. Spayde et al. "Parity Violation in Elastic Electron-Proton Scattering and the Proton's Strange Magnetic Form Factor" Phys. Rev. Lett., vol 84, pp. 1106-1109, 2000.



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

## SUBPROGRAMA RAMON Y CAJAL CONVOCATORIA 2009

**Nombre:** SAGGION , HORACIO

**Referencia:** RYC-2009-04291

**Area:** Ciencias de la Computación y Tecnología Informática

**Correo electrónico:** saggion@dcs.shef.ac.uk

**Título:**

MÉTODOS Y HERRAMIENTAS PARA EL RESUMEN AUTOMÁTICO POR ABSTRACCIÓN

**Resumen de la Memoria:**

La generación de resúmenes por ordenador es la tarea de extraer el contenido esencial de uno o más documentos y de presentar ese contenido esencial en un nuevo texto cohesivo y coherente. Este es un problema muy difícil de resolver. El paradigma corriente en generación automática de resúmenes es el de extracción de oraciones. En este paradigma primero se seleccionan un cierto número de oraciones de uno o varios documentos de acuerdo a un criterio de selección, y luego se las presenta una a continuación de la otra para producir el resumen final. Si bien, este paradigma es universalmente aceptado y corrientemente utilizado, es sabido que no permite producir resúmenes concisos, cohesivos y coherentes. Un método alternativo para la producción de resúmenes es el de extracción y generación, según el cual, informaciones (no ya oraciones) son extraídas de uno o más documentos para producir representaciones más o menos abstractas utilizando métodos de extracción de informaciones. Luego, estas informaciones o representaciones son combinadas utilizando métodos de generación automática de textos. El uso de técnicas de extracción y generación de textos para el resumen automático tiene las ventajas de producir resúmenes de buena calidad, sin embargo el costo asociado con esta solución es alto, dado que los sistemas de extracción y generación requieren la codificación del conocimiento del dominio de aplicación. Una manera de solucionar los problemas asociados a la codificación manual de conocimiento para el resumen automático es utilizar resúmenes disponibles en la Web o en colecciones de texto conjuntamente con técnicas de aprendizaje semi supervisado y no supervisado para responder de una manera empírica a las preguntas siguientes: ¿Que conceptos y relaciones son más importantes en los resúmenes de un dominio particular? ¿Como extraer estos conceptos y relaciones de los documentos? ¿Como generar un nuevo texto que contenga los conceptos y relaciones extraídos de los documentos? Estos son los problemas que pienso investigar en caso mi solicitud sea aceptada. Hoy en día vivimos en la sociedad de la información en la cual nos enfrentamos al problema de la sobre abundancia de la información, métodos para reducir textos a su contenido esencial son altamente requeridos. Esta investigación contribuirá con métodos nuevos así como efectivos en resumen automático y generación de conocimientos a partir de datos brutos (textos). Estos métodos tendrán aplicación en otras áreas del procesamiento del lenguaje natural.

**Resumen del Curriculum Vitae:**

Educación Superior Doctorado en Ciencias de la Computación. Universidad de Montreal, Montreal, Quebec, Canada. Director de Tesis: Prof. Guy Lapalme. 1997-2000. Maestría en Ciencias de la Computación. Universidade Estadual de Campinas, Campinas, Brasil. Director de Tesis: Prof. Ariadne Carvalho. 1993-1995. Licenciatura en Ciencias de la Computación. Universidad de Buenos Aires, Buenos Aires, Argentina. Director de Tesis: Prof. José Álvarez. 1981-1988. Puestos Académicos y de Investigación Investigador (Senior) (2003-2009), Universidad de Sheffield, Departamento de Computación, grupo de Procesamiento del Lenguaje Natural, Sheffield, Reino Unido. Investigador (2000-2002), Universidad de Sheffield, Departamento de Computación, grupo de Procesamiento del Lenguaje Natural, Sheffield, Reino Unido. Profesor Invitado (2008, 1 semana), Universidad de Buenos Aires, Buenos Aires, Argentina. Profesor Invitado (2004, 1 semana), Universidad de Rio Cuarto, Rio Cuarto, Córdoba, Argentina. Investigador Invitado (2001, 6 semanas), Universidad Johns Hopkins, Estados Unidos. Docente Auxiliar (1998-1999), Universidad de Montreal, Montreal, Canadá. Jefe de Trabajos Prácticos e Investigador (1993-1996), Universidad de Buenos Aires, Buenos Aires, Argentina. Jefe de Trabajos Prácticos (1992-1993), Universidad Nacional de Quilmes, Quilmes, Argentina. Docente Auxiliar (1987-1991), Universidad de Buenos Aires, Buenos Aires, Argentina. Proyectos de Investigación 2006-2009: Proyecto MUSING (Sexto Programa Marco de la Comisión Europea). Supervisión de investigadores, investigación, y desarrollo. Universidad de Sheffield. 2002-2008: SUMMA Toolkit. Sistema para el desarrollo de aplicaciones de resumen automático. Universidad de Sheffield. 2007 Evaluación SEMEVAL. Diseño e implementación de un sistema de resolución de la co referencia de nombres de personas en múltiples documentos. El sistema estuvo entre los mejores de las evaluaciones SemEval 2007. Universidad de Sheffield. 2003-2005 Evaluación TREC/QA: Diseño e implementación de un sistema de extracción de definiciones de texto. El sistema estuvo entre los mejores de las evaluaciones TREC/QA 2003 y 2004. Universidad de Sheffield. 2003-2005 Proyecto Cubreporter (EPSRC). Investigación y desarrollo en sistemas de preguntas y respuestas y resumen automático. Tareas de supervisión de investigadores. Universidad de Sheffield. 2004-2006: Evaluación de resumen automático (Document Understanding Conferences (DUC)). Diseño e implementación de sistemas de resumen multidocumento. En 2004, el sistema estuvo entre los mejores de la evaluación DUC. Universidad de Sheffield. 2003: Proyecto CLARITY (Comisión Europea). Implementación de sistemas de resumen automático en múltiples lenguas. Universidad de Sheffield. 2000-2002: GATE: General Architecture for Text Engineering (EPSRC). Implementación de varios componentes de software. Universidad de Sheffield. 2001: Proyecto SummBank (National Science Foundation). Investigación en resumen automático. Universidad Johns Hopkins. 2000-2002: Proyecto MUMIS (Comisión Europea). Diseño y desarrollo de sistemas de extracción de informaciones. Universidad de Sheffield. 2000-2002: Proyecto SOCIS (EPSRC). Diseño y desarrollo de componentes para el análisis de textos. Tareas de supervisión de investigadores. 1997-2000: SumUM (Proyecto de tesis de doctorado). Investigación en resumen automático. Universidad de Montreal.



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

**SUBPROGRAMA RAMON Y CAJAL  
CONVOCATORIA 2009**

**Nombre:** VAILARIÑO FREIRE, FERNANDO

**Referencia:** RYC-2009-05492

**Area:** Ciencias de la Computación y Tecnología Informática

**Correo electrónico:** fernando@cvc.uab.es

**Título:**

Advanced Feature Extraction and Machine Learning Approaches for Analysis and Pathology Detection in Endoscopy Imaging

**Resumen de la Memoria:**

Research on advanced features and the design of efficient classification systems constitutes the two pillars over which the investigation on computer-based methods for clinical support is founded. In this fieldwork, medical image analysis constitutes one of the main areas of investigation. In the case of intestinal endoscopy, recent works have focused on the pathology characterization by using multi-modal data from different sources, such as classical endoscopy and, lately, capsule endoscopy. We aim at tackling some of the problems which have received most relevance in the last years. On the one hand, we propose to investigate visual static (frame based) and dynamic (sequence-based) features for the automatic analysis of capsule endoscopy videos. This approach will provide a characterization of the video frames and sequences in correlation with the patterns on intestinal motility. This information will be included in a machine learning-based system, which explores the use of combination of classifiers in order to provide a probabilistic framework for the automatic characterization of severe pathologies, providing very valuable results in one of the areas in which capsule endoscopy is shown to perform better than alternative techniques. On the other hand, we propose to investigate the use of computer-based systems in classical endoscopy for two of the most relevant challenges in the fieldwork: The use of salient features for the automatic on-line detection and characterization of colon cancer, and the identification of landmarks for the automatic design of a patient-specific 2D map of the colon which allows the geographic localization of the lesions. In addition to the former, we propose the identification of different estimates for real time quality assessment in colonoscopy, which are essential for the reliability of the diagnosis, such as the overall amount of surface screened, the level of preparation and the automatic assessment of the skills of experts and novices both for evaluation and training purposes. Due to the current collaboration with projects and international groups which are leading the area of endoscopic imaging -such as Graphics Vision and Visualization, TCD, Ireland, and the Hospitals Vall d'Hebron, Barcelona, Spain and Beaumont and St. Vincent's, Dublin, Ireland- we are in a privileged position to carry out the aforementioned research lines, which will provide an impact in the improvement of the quality assessment of the endoscopic procedures and, finally, will help to save more lives.

**Resumen del Curriculum Vitae:**

Fernando Vilariño Freire obtained his B.Sc. in Physics and M.Sc. in Computer Vision and Artificial Intelligence in Univ. Santiago de Compostela, Spain, and his Ph.D. degree by Univ. Autònoma de Barcelona. His research interest is focused on feature extraction and machine learning applied to medical imaging. He has spent research stages in leading institutions in the area of computer vision and machine learning, such as the Computer Vision Centre in Barcelona, Spain, the School of Informatics in Bangor Univ. Wales, U.K., and the Computer Science Dep. in Trinity College Dublin, Ireland. He has participated in 10 competitive research projects, 3 projects of technology transfer and 3 research contracts with private institutions. The contributions of his research have been published in 7 international peer-reviewed indexed journals comprising 21.04 impact factor (8.67 associated to computer science areas exclusively). His research contributions sum up a total of 82 citations (the article with more cites having 29), and he is the co-author of 4 currently filled patents. In addition, he has published his research in multiple international conferences, he has been co organizer of different international events and he has given invited talks for the Univ. of Wales, Trinity College Dublin and Porto. Fernando Vilariño currently has a Lecturer position in Univ. Autònoma de Barcelona, he is the PI of a project for the implementation of the aforementioned lines of research in mobile technologies, and he is the coordinator of a Master module in R&D in Computer Vision and Artificial Intelligence.



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

**SUBPROGRAMA RAMON Y CAJAL  
CONVOCATORIA 2009**

**Nombre:** SALAÜN, GWEN

**Referencia:** RYC-2009-04563

**Area:** Ciencias de la Computación y Tecnología Informática

**Correo electrónico:** salaun@lcc.uma.es

**Título:**

Composition and (Self-)Adaptation of Components and Services: Detecting, Controlling and Correcting Interoperability Issues

**Resumen de la Memoria:**

Building software systems as a combination of interacting entities (components, services, etc.) is a central issue in the Software Engineering community. Reuse of existing entities is mandatory in order to avoid implementing again the same pieces of software, thus helping developers to reduce development time, respect deadlines, and have their companies save money by diminishing design costs. In this context, services must be equipped with public interfaces giving enough information about their functionality and technical characteristics in order to be reused and composed as required. Several interoperability levels are usually distinguished in interface description languages: signature level, behavioural level, quality of service level, and semantic level (sometimes referred as conceptual level). Recently, many academic approaches and industrial platforms have integrated behavioural descriptions in interface models and programming languages. Indeed, the behavioural interoperability level is essential, because even if services match from a signature point of view, their combination can lead to erroneous behaviours or deadlock situations if the designer is not aware of their execution flows, and does not take them into account while building a new system. The goal of our research program is to consider services described using signatures and protocols, and focus on interoperability issues raised during the service composition stage, in order to detect these problems and find solutions for them. The program consists of five parts: (i) compatibility and similarity of service protocols, (ii) behavioural adaptation and (iii) self-adaptation of services, (iv) re-engineering of software components, (v) choreography realizability, and peer generation. Issues overviewed in this program are of special interest to many industrial projects, and constitute an area of research where the candidate expects to develop strong relations with some private companies and local organizations which have already shown interest for the project, hence enabling the technological transfer of these research results to the industry. The candidate has been working for many years on the composition and adaptation of software entities, and his background in formal techniques and tools (PhD Thesis, and 2 years at INRIA working on formal methods and verification) has already been very helpful to his research on this topic. Since 2004, the candidate published several results on composition and adaptation in top international conferences (ICWS'04, ECOWS'04, WI'05, Coordination'07, ASE'07, SEFM'07/08, FASE'08, ICSOC'08, IFM'09, ICSE'09) and journals (IJBPIM, WIAS, IEEE TSE, JUCS). Thus, the candidate is qualified to propose deep and formal solutions to interoperability issues raised while building component and service-based systems.

**Resumen del Curriculum Vitae:**

The candidate was a PhD student at the University of Nantes (France) in 2000-2003. He defended his thesis after 2.5 years, obtaining the "outstanding" mention for his work on formal specification techniques. He spent one year on a first postdoctoral stay in Italy (University of Rome) in 2003-2004 where he worked on verification of Web services. For these contributions, he was awarded a best paper runners up prize at the International Conference on Web Services in 2004 (230 submissions, 66 accepted papers), and this paper is widely cited and used as reference by many related research works (121 citations in Google Scholar). A second paper on service compatibility checking published in 2004 is also often quoted (74 citations in Google Scholar). During a second postdoctoral stay in France (VASY Project, INRIA Rhône-Alpes) in 2004-2006, the candidate worked in the area of software and hardware verification. He collaborated with CEA (one of the top French Research Centers currently hiring more than 15,000 employees) to apply his research proposals and tools to the verification of state-of-the-art Chip architectures, including asynchronous circuits or Network-on-Chip (NoC) architectures. One of the success stories of this collaboration was the detection and correction of a routing problem in an asynchronous NoC, which was used as the backbone of a 4th Generation Wireless Telecom Baseband. The candidate also spent one month (winter 2006) at the Imperial College in London working with Pr. Jeff Kramer and Pr. Jeff Magee. The candidate collaborated with them to widen the range of verification techniques presented in their best-seller book "Concurrency: State Models & Java Programs". Since 2006 the candidate holds a "Juan de la Cierva" research grant at the Computer Science Department of the University of Málaga, where he has been working on the composition and adaptation of components and services. In these last two years, he published his results (14 papers) in top-notch international conferences and journals, including IEEE TSE (2 articles), JUCS, SEFM'07, Coordination'07, ASE'07, FASE'08, SEFM'08, ICSOC'08, ICSE'09, etc. The candidate also spent one month (summer 2008) at the University of California (Santa Barbara, USA) to collaborate with Tevfik Bultan. In addition, the candidate participated in many research projects, and set up numerous contacts in different research groups. These collaborations led to a large number of publications (about 50), and in each case he has been a principal contributing author. Regarding Committee membership, he usually serves in Organizing and Program Committees of international events related to his research areas. He will be Program Committee Chair of the 8th FOCLASA workshop in 2009. He has carried out extensive reviewing for many international workshops, conferences and journals, and has actively participated (as speaker and/or session chair) in about 20 international workshops and conferences. Furthermore, he has also been invited (14 times) to present his research contributions in the context of seminars in prestigious institutions such as several INRIA centres in France, the Microsoft Research lab in Cambridge (UK), or the University of California (Santa Barbara, USA). Finally, the candidate has taught at different Universities, supervised several Master Theses, and is now supervising 2 PhD students.



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

## SUBPROGRAMA RAMON Y CAJAL CONVOCATORIA 2009

**Nombre:** KARATZAS , DIMOSTHENIS

**Referencia:** RYC-2009-05031

**Area:** Ciencias de la Computación y Tecnología Informática

**Correo electrónico:** dimos@cvc.uab.es

**Título:**

Técnicas Inspiradas en la Percepción Humana para el Análisis del Texto en Imágenes Digitales

**Resumen de la Memoria:**

**INTRODUCCIÓN** En las últimas décadas las aplicaciones de Internet, como e-mail y la Web, han propiciado el intercambio masivo de información electrónica. En la actualidad es habitual incluir imágenes junto con texto en los documentos electrónicos. EL PROBLEMA Las imágenes digitales que contienen texto (anuncios, logotipos, titulares, etc) son entidades del documento con importancia semántica y contienen información esencial. Aunque el texto de las imágenes es importante para la comprensión del documento, los métodos automáticos de búsqueda, indexado, filtrado y análisis de información son incapaces de extraer dicho texto; y el texto-imagen es sencillamente ignorado en todos los procesos automáticos. Las imágenes digitales son claramente distintas de los documentos de papel y por tanto los métodos clásicos de análisis de imágenes de documentos, desarrollados para documentos en papel, son inaplicables a ellas. LA NECESIDAD La industria ha detectado esta cuestión y ha adoptado imágenes de texto digitales como plataforma para implementar tests CAPTCHA ("Completely Automated Public Turing Test to Tell Computers and Humans Apart") para controlar el abuso de los recursos online por parte de robots. La forma preferida de CAPTCHA presenta el texto codificado como una imagen digital ligeramente oscurecida. La lectura del texto en estas imágenes es trivial para los humanos, pero no así para los ordenadores que son incapaces de procesarla incluso con las técnicas más simples de oscurecimiento. Siguiendo esta tendencia los spammers usan cada vez más imágenes para transmitir mensajes no deseados a los usuarios, y los filtros de spam tienen dificultades para confrontar esta tendencia. La extracción del contenido de texto de las imágenes digitales es por tanto un tema candente con una plétora de aplicaciones comerciales identificadas. HIPÓTESIS INICIAL Las imágenes digitales de texto (e imágenes de documentos en general) se diferencian de la mayoría de otras imágenes (médicas, satélite, escenas reales, etc) en el sentido en que han sido preparadas por humanos para ser visualizadas e interpretadas por otros humanos. Por lo tanto, consciente o inconscientemente, el autor emplea ciertos esquemas de diseño para que el contenido de texto sobresalga y sea fácil de leer. OBJETIVO DEL PROYECTO Basándose en esta hipótesis, el candidato propone utilizar ciertas técnicas inspiradas en la percepción humana para modelar el proceso humano de la localización del texto y separación del fondo con enfoque a las imágenes digitales en color. El objetivo final del proyecto es desarrollar métodos noveles inspirados en la percepción humana para la extracción del texto de imágenes digitales en color. La idea es abordar este problema de una forma integral y examinar un amplio rango de conceptos perceptuales. Inicialmente se tratarán tres aspectos del proceso de extracción de texto y técnicas asociadas inspiradas en la percepción humana: pre-procesado basado en color perceptual; localización y extracción del texto basado en modelos de atención y saliency maps; segmentación e interpretación de sus resultados basada en leyes de agrupamiento de gestalt. El corpus de imágenes se enfocará inicialmente a tres grupos específicos de imágenes digitales: imágenes web, anuncios spam y tests CAPTCHA.

**Resumen del Curriculum Vitae:**

**BIOGRAFÍA** Karatzas es investigador en el Centro de Vision por Computador (CVC) en la Universitat Autònoma de Barcelona e investigador visitante en el PRIma lab, Univ. of Salford, UK. Se licenció en Física en la Univ. Aristóteles de Salónica, Grecia en 1998. En 2003, se doctoró en Ciencias de la Computación en la Univ. de Liverpool, UK, con "Text Segmentation in Web Images Using Colour Perception and Topological Features". Sus intereses de investigación son el análisis de imágenes de documentos (DIA) especialmente documentos históricos y en color, percepción humana, visión por computador y ciencia del color. Ha trabajado en investigación durante la última década y ha producido más de 25 publicaciones revisadas por pares en revistas, capítulos de libros y conferencia internacionales (7 indexadas en el JCR), han sido citadas más de 130 veces. Esto incluye 6 publicaciones en revistas principales (incluye Elsevier Image and Vision Computing y IEEE Sensors), 5 capítulos en libros y 14 artículos en proceedings revisados por pares en conferencias internacionales. Dr. Karatzas es director de la compañía spin-off, TruColour, especializada en soluciones de calibración del color basadas en la percepción. La compañía ha recibido financiación "venture capital" y otra financiación para desarrollar el prototipo. También ha proporcionado consultoría y soporte técnico a AIMES, UK y Perpetuum Ltd, UK. Ha solicitado con éxito subvenciones para investigación y es actualmente el investigador principal en dos proyectos supervisando a dos investigadores y dos estudiantes de doctorado. Es chair (Publications) de ICDAR 2009 y ha participado en comités de programa de ICDAR, PRIS y CVCRD, es miembro del comité de liderazgo de TC-11 (Reading Systems) de IAPR y de IEEE, SPIE y IAPR. Dr Karatzas tiene una extensa experiencia en el extranjero a través de numerosas estancias en instituciones extranjeras que suman más de 8 años. Sus estancias incluyen centros de reconocido prestigio, como el School of Electronics and Computer Science of the Univ. of Southampton, UK con rango 5\* (máximo en UK en investigación). **COMITÉS INTERNACIONALES**- Miembro del comité líder de IAPR TC 11 (Reading Systems)- Miembro fundador y del comité ejecutivo de SPIE UK chapter- Miembro de las sociedades profesionales: IEEE, SPIE, IAPR, BMVA, AERFAI **ORGANIZACIÓN DE CONFERENCIAS**- Chair (Publications) ICDAR09- Miembro de comité de programas: ICDAR09, PRIS08, CVCRD07. **CHARLAS INVITADAS**- Univ. de Salerno, Italia, 2009- Univ. Autónoma de Barcelona, 2007- Univ. de Salford, UK, 2006- Univ. de Birmingham, UK, 2005 **TRANSFERENCIA DE TECNOLOGÍA**- Creación de la compañía spin-off TruColour, en 2007.- Director técnico de TruColour Ltd. He asegurado financiación "venture capital" y otra financiación para desarrollar un prototipo.- Consultor de AIMES Centre, Liverpool, UK y Perpetuum Ltd **OTRAS ACTIVIDADES PROFESIONALES**- Investigador principal, MICINN TIN 2008-04998 y NDWA NW07/0041.- Actualmente superviso dos doctorandos, dos investigadores y dos estudiantes de ingeniería informática.- Docente en varias asignaturas en Univ. Liverpool.- Creador y responsable de un grupo de lectura en Percepción Humana y Visión por computador.- Reviewer en revistas (IJDR, SMC, IEEE Sensors etc) y conferencias (ICDAR, DAS, ICPR etc)



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

**SUBPROGRAMA RAMON Y CAJAL  
CONVOCATORIA 2009**

**Nombre:** SANCHO PITARCH, JOSE CARLOS

**Referencia:** RYC-2009-03989

**Area:** Ciencias de la Computación y Tecnología Informática

**Correo electrónico:** jcsanchop@gmail.com

**Título:**

Escalando el Rendimiento de las Aplicaciones en los Computadores con Potencia Exascale

**Resumen de la Memoria:**

El objetivo de la propuesta de investigación es permitir la escalabilidad del rendimiento de las aplicaciones en computadores capable con potencia de calculo exascale. Estos sistemas no existen todavia hoy, pero se vislumbran que puedan ser construidos en la proxima década a partir del desarrollo de investigación básica realizada hoy en día. Uno de los desafíos mas importantes en la construccion de estos sistemas reside en proveer de rapidas comunicaciones entre los diferentes elementos del sistema. Para abordar este importante problema creemos que es necesario un enfoque integrado que combine investigación en redes de interconexion con investigación en algoritmos escalables de las aplicaciones. Estas areas se complementan una a otra ya que las investigación en redes proporcionara una gran escalabilidad y rendimiento en las comunicaciones, mientras que la investigación en algoritmos escalables hace un uso eficiente de estas. La investigación en redes de interconexion se centrara en resolver los problemas de escalabilidad, limitacion de potencia y la aparicion de tecnologías disruptivas en estos sistemas como las redes opticas. Para resolver estos problemas se enfocara la investigación de una manera que unifique las diferentes redes que se encuentran en el sistema. En concreto, la investigación se centrara en el diseño de topologias y algoritmos de encaminamiento para estas redes. Por otra parte la investigación en algoritmos escalables se centrara en desarrollar ideas innovadoras en la aplicacion para que reduzcan la dependencia de esta a sus comunicaciones. Algunas de las ideas que se perseguiran en esta investigación seran el solapamiento, la localidad, y la escalabilidad debil. Estas tecnicas tendran que afrontar la complejidad cada vez mayor de las aplicaciones las cuales cada vez introducen una mayor variedad de algoritmos.

**Resumen del Curriculum Vitae:**

Obtuve mi licenciatura de Informatica en la Facultad de Informatica, Universidad Politecnica de Valencia en 1998, donde seguidamente realice mis estudios de doctorado bajo la supervision de dos profesores catedraticos, Antonio Robles y Jose Duato. Obtuve el doctorado en Computer Science en el año 2002. Mi tesis se basa en el diseño de los algoritmos de encaminamiento eficientes para redes de estaciones de trabajo. En esta tesis propuse algoritmos de encaminamiento que mejoran sustancialmente los actuales algoritmos de encaminamiento propuestos ya hace mas de 10 años. Durante mi doctorado publiqué en mas de 20 congresos y workshops presentando mi trabajo nacionalmente como internacionalmente. Tambien publiqué en una revista de investigación en computacion paralela con un alto factor de impacto, denominada IEEE Transactions on Parallel and Distributed Systems. Durante este periodo obtuve dos becas para realizar mis investigaciones, una por parte de la Universidad Politecnica de Valencia, mientras que la otra es del Ministerio de Ciencia y Tecnología, denominada Formacion de Personal Investigador (F.P.I.). Asimismo, participe en tres proyectos de investigación financiados tambien por organismos publicos. Mis contribuciones en el campo de la computacion paralela han sido reconocidos internacionalmente como se demuenstra en que he sido nombrado para presidir la session de encaminamiento en una prestigiosa conferencia internacional y servido en varias ocasiones para revisar articulos de investigación realizados por otros investigadores en revistas y conferencias de alto prestigio como International Parallel and Distributed Processing Symposium. Tras la defensa de mi tesis realice una estancia de unos seis años en el laboratorio nacional de Los Alamos en Estados Unidos de America. Empece mi estancia en una posicion tipo Postdoc Research Assistant durante tres años y ahora sostengo desde hace tres años una posicion de Computer Scientist que es la posicion típica de los investigadores que trabajan en este laboratorio. Durante este periodo he trabajado en dos tipos de proyectos, uno relacionado con la tolerancia a fallos en computadores paralelos y el otro en el solapamiento de las comunicaciones en aplicaciones científicas. Durante este periodo he publicado en 16 journals, conferencias y workshops internacionales asi como en 8 reports internos del mismo Laboratorio. Las conferencias en donde se ha publicado mi trabajo son de las de mas reputacion en computacion paralela en Estados Unidos de America, con el nivel mas bajo de aceptancia y un alto factor de impacto, mostrando asi la calidad mi trabajo, siendo estas conferencias Supercomputing Conference e International Parallel and Distributed Processing Symposium.



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

## SUBPROGRAMA RAMON Y CAJAL CONVOCATORIA 2009

**Nombre:** DOZSA , GABOR

**Referencia:** RYC-2009-04726

**Area:** Ciencias de la Computación y Tecnología Informática

**Correo electrónico:** gabordzs@yahoo.com

**Título:**

Modelos de programación novedosos para Computación de Altas Prestaciones

**Resumen de la Memoria:**

Los supercomputadores actuales están formados por decenas de miles de procesadores (ver [www.top500.org](http://www.top500.org)), y las máquinas de la siguiente generación van a tener al menos un orden de magnitud más. La complejidad de estos computadores está creciendo aún más, si cabe, con la incorporación de elementos heterogéneos. La heterogeneidad se encuentra en las más variadas formas: procesadores heterogéneos, por ejemplo la Arquitectura Cell Broadband Engine (Cell BE); sistemas heterogéneos, por ejemplo máquinas como la Bull Novascale, un sistema híbrido CPU/GPU. Actualmente, existe la creencia de que los sistemas heterogéneos son esenciales para que las aplicaciones puedan obtener ventajas de la computación de alto rendimiento (High Performance Computing, HPC). Para conseguir un alto tanto por ciento de rendimiento en estas arquitecturas es necesario explotar el paralelismo a distintos niveles, lo que requiere un gran conocimiento de la plataforma y muy posiblemente también de la aplicación. Programar eficientemente estos sistemas es extremadamente difícil. El programador tiene que conocer muchos detalles de la arquitectura, lo cual muchas veces es difícil incluso para programadores expertos. Además, actualmente cualquier PC moderno puede explotar paralelismo y la mayoría de programadores no son expertos en este tema. Ello incrementa la importancia de incrementar la productividad dada por el modelo de programación. Otro aspecto importante es cuán portable es el rendimiento obtenido (Performance Portability). Aplicaciones de alto rendimiento escritas en su mayoría en MPI no alcanzarán el rendimiento esperado, y en algún caso ni siquiera podrán ejecutarse en los computadores de la siguiente generación (>10PF). El modelo de programación de alto rendimiento, estándar de facto actual, MPI, tiene una falta considerable de soporte para productividad y portabilidad de rendimiento. Esta situación puede impedir la explotación del rendimiento de sistemas paralelos actuales y futuros. Claramente, existe la necesidad de desarrollar nuevos modelos de programación para solucionar estos problemas. Es muy importante investigar en los modelos de programación, lenguajes, librerías de soporte que puedan facilitar la portabilidad del rendimiento entre sistemas distintos (en cuanto a memoria distribuida o compartida, multi-core), teniendo en cuenta las características de sus recursos, latencia de la red de interconexión, ancho de banda, limitaciones de memoria en el chip y tamaño de la memoria caché. Deben buscarse soluciones que sean suficientemente amplias para cubrir las diversas arquitecturas y nos acerquen a los objetivos de productividad y alto rendimiento. Una propuesta interesante en esta línea son los modelos de Espacio de Direccionamiento Global Particionado (Partitioned Global Address Space, o PGAS), que combinan el concepto de memoria virtual compartida, para facilitar la productividad, con la noción de afinidad explícita, para obtener alto rendimiento. Un sistema PGAS bien estructurado y optimizado puede ser un buen candidato para mejorar la productividad y la portabilidad de rendimiento en un amplio espectro de sistemas y arquitecturas paralelas.

**Resumen del Curriculum Vitae:**

Diplomas académicos 1996-2003 Doctorado en Informática, Facultad de Informática de la Universidad de Lorand Eotvos (Budapest, Hungría), Calificación: "summa cum laude" 1991-1993 Master en informática, Universidad de Lorand Eotvos (Budapest, Hungría) 1987-1991 Ingeniería en informática, Universidad Lorand Eotvos (Budapest, Hungría) Estancias en centros de investigación 2004- IBM T. J. Watson Research Center, Yorktown Heights, EEUU, Blue Gene System Software Group. Puesto de trabajo: 2006 Ene. - Investigador miembro en plantilla 2004 Nov. - 2005 Dic. Investigador postdoctoral 1998-2004 MTA SZTAKI Computer and Automation Research Institute of the Hungarian Academy of Sciences, Budapest, Hungría. Puesto: jefe en funciones del Laboratorio de Sistemas Paralelos y Distribuidos. Proyectos de investigación y contribuciones esenciales 2007. 2008 Colaboración entre IBM y TD Bank para realizar el prototipo de una aplicación financiera de alto rendimiento basada en el modelo streaming sobre la arquitectura de Blue Gene/P. Contribución: Diseño e implementación del BGSM (Blue Gene Stream Messaging), librería de comunicación que proporciona una API streaming de baja latencia y alto ancho de banda para aplicaciones comerciales sobre BG/P. 2005- Desarrollo de la librería de soporte a paso de mensajes para superordenadores Blue Gene (IBM). Contribución en el diseño de las librerías de paso de mensajes DCMF y CCMI, diseño e implementación del componente del software para soporte de operaciones colectivas, diseño e implementación de algoritmos optimizados en latencia y ancho de banda para diversas operaciones de allreduce y broadcast, diseño e implementación de una versión multi-canal del componente software para la red DMA/Torus, y otros aspectos del stack de comunicaciones de Blue Gene... Publicaciones más relevantes: Periódicos/revistas P. Kacsuk, G. Doza, J. Kovacs, R. Lovas, N. Podhorszki, Z. Balaton, y G. Gombas P-grade: A grid programming environment, Journal of Grid Computing, 1(2):171-197, 2003. R. Wisniewski, G. Doza, y D. Drotos. Enhanced monitoring in the GRADE programming environment by using OMIS, Future Generation Computer Systems, 16(6):637-648, 2000. Capítulos de libros G. Doza Visual programming to support parallel program design, In C. Cunha, P. Kacsuk, and S. Winter, editors, Parallel Program Development for Cluster Computing, Methodology, Tools and Integrated Environments, pages 17-44. Nova Science Publishers, Inc., 2001. Conferencias de referencia y seminarios G. Doza, M. Eleftheriou, T.A. Inglett, A.J. King, T.E. Musta, J. Sexton and R.W. Wisniewski. Stream Processing Performance for Blue Gene/P Supercomputer, In Proceedings of High Performance Computational Finance, WHPCF'08, páginas 1-7, 2008. S. Kumar, G. Doza, J. Berg, B. Cernohous, D. Miller, J. Ratterman, B.E. Smith, P. Heidelberger. Architecture of the Component Collective Messaging Interface, In Proceedings of 15th Euro PVM/MPI 2008 Conference, páginas 23-32, 2008. G. Doza, P. Kacsuk, Cs. Nemeth. Grid Application Development on the basis of Web Portal Technology, Lecture Notes in Computer Science, Vol. 3732, páginas 472-480, 2006



MINISTERIO  
DE CIENCIA  
E INNOVACIÓN

**SUBPROGRAMA RAMON Y CAJAL  
CONVOCATORIA 2009**

**Nombre:** BOZZELLI, LAURA

**Referencia:** RYC-2009-05383

**Area:** Ciencias de la Computación y Tecnología Informática

**Correo electrónico:** ryc-app@clip.dia.fi.upm.es

**Título:**

Automatic verification of infinite-state reactive systems

**Resumen de la Memoria:**

Bozzelli's research activity is concentrated on decision, complexity, expressiveness, and succinctness issues related to the formal analysis of infinite-state reactive systems in the framework of automata theory, game theory, and temporal and modal logics. In particular, the focus is on algorithmic methods developed in the context of model-checking. Model-checking is one of the most successful technologies for automatic verification in design phase of software and hardware reactive systems, i.e. systems which need to interact with their environment frequently, and often do not terminate. Typical examples are concurrent and/or distributed systems. Practical examples include telephone switching networks and safety-critical systems such as nuclear power plants and air traffic control systems. In the model-checking method, the temporal behavior of a system, described by an abstract formal model, is checked against a behavioral constraint usually specified by a formula in a suitable temporal logic. Model-checking has been successfully applied to finite-state systems and the underlying theory is by now well understood. In the last ten years much interest has been addressed to the analysis of infinite-state systems. The practical motivation is to allow the automatic verification of (non-Turing-equivalent) abstract formal models which provide a more faithful representation of the real concurrent systems. These may introduce infinite states for different reasons: (1) Unbounded data structures: counters, queues, etc., (2) Unbounded control structures: recursive procedure calls and/or dynamic creation of concurrent processes (e.g., multi-threading), and (3) real-time constraints for systems which manage synchronous or asynchronous events with predictable response times. One general problem is however, the trade-off between expressivity and decidability. During her career, Bozzelli has investigated automatic verification problems for numerous infinite-state formalisms for reactive systems in the following contexts: (1) verification of recursive and concurrent programs, (2) verification of open systems, and (3) parametric real-time verification. Moreover, she has given new contributions to automata theory on infinite objects by introducing and investigating some classes of alternating automata on infinite strings. Furthermore, she has devoted to decidability and complexity issues for several extensions of standard temporal and modal logics.

**Resumen del Curriculum Vitae:**

1995-1999, Laurea Degree in Physics (summa cum laude) at Università degli Studi di Napoli Federico II, Italy. Thesis Title: Ambiente Distribuito Orientato ad Oggetti con Progetto e Sviluppo di un VideoClient di un sistema Video on Demand. Thesis Defense: March, 1999. 1999-2000, Master in Information Technology (Research area: Software Engineering) with fellowship at CEFRIEL of Polytechnic of Milan, Italy. Thesis Title: Specification and analysis of object-oriented real-time software systems. Thesis Defense: June, 2000. 2001-2004, Ph.D in Computer Science with fellowship at Università degli Studi di Napoli Federico II, Italy. Advisor: Prof. Adriano Peron. Thesis Title: Model Checking Recursive and Parallel Systems. Thesis Defense: December, 2004. 2005-2006, Post-doctoral Researcher position at Laboratoire Spécification et Vérification (LSV) - Ecole Normale Supérieure de Cachan, Paris, France. Supervisors: Dr. Patricia Bouyer and Dr. Stéphane Demri. Topic: Control and Observation of Real-Time Open Systems. 2007- Present, Research Associate at Department of the Sciences of Culture, Politics and Information - Università degli Studi dell'Insubria, Como, Italy. Topic: Automatic verification of infinite-state reactive systems.