

nov@tica



CEPIS UPGRADE

Revista de la Asociación
de Técnicos de Informática

Nº 197, enero-febrero 2009, año XXXV

Web universal, ubicua e inteligente

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

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

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

Consejo Editorial

Joaquín Ballester Montserrat, Rafael Fernández Calvo, Luis Fernández Sanz, Javier López Muñoz, Alberto Lirio, Gabriel Martí Fuentes, Josep Moias i Bertan, José Onofre Montes Andrés, Olga Pallás Codina, Fernando Píera Gómez (Presidente del Consejo), Ramon Puigjaner Trepal, Miquel Sarries Grifó, Adolfo Vázquez Rodríguez, Asunción Yturbe Herranz

Coordinación Editorial

Llucen Pagés Casas <pages@ati.es>

Composición y autoedición

Jorge Llácer Gil de Ramales

Traducciones

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

Administración

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

Secciones Técnicas - Coordinadores

Acceso y recuperación de la información

José María Gómez Hidalgo (Uptenet), <jmgomez@yaho.es>

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

Administración Pública electrónica

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

Arquitecturas

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

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

Auditoría SITIC

Marina Touriño Troitino, <marinatourino@marinatourino.com>

Manuel Palao García-Suñeto (ASIA), <manuel@palao.com>

Derecho y tecnologías

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

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

Escuela Universitaria de la Informática

Cristóbal Pareja Flores (DSIP-UCM), <cpareja@sp.ucom.es>

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

Entorno digital personal

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

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

Estándares Web

Encarna Quesada Ruiz (Pez de Babel), <equesda@pezdebabel.com>

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

Gestión del Conocimiento

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

Informática y Filosofía

José Ángel Olivas Varea (Universidad Superior de Informática, UCLM), <joseangel.olivas@uclm.es>

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

Informática Gráfica

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

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

Inteligencia Artificial

Vicente Botti Navarro, Vicente Julián Inglada (DSIC-UPV)

<vbotti.vinglada@dsic.upv.es>

Información Persona-Computador

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

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

Lengua e Informática

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

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

Lingüística computacional

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

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

Mundo estudiantil y jóvenes profesionales

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

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

Profesiones Informáticas

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

Miquel Sarries Grifó (Ayto. de Barcelona), <msarries@ati.es>

Redes y servicios telemáticos

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

Germán Santos Boada (UPC), <german@ac.upc.es>

Seguridad

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

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

Sistemas de Tiempo Real

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

Software Libre

Jesús M. González Barahona, Israel Herráiz Tabernero (GSYC-URJC), <jmhb.herraz@gsyc.es>

Tecnología de Objetos

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

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

Tecnologías para la Educación

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

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

Tecnologías y Empresa

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

Francisco Javier Cantais Sánchez (Indra Sistemas), <fjancanis@gmail.com>

Tendencias tecnológicas

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

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

TIC y Turismo

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

Publicidad

Padilla 66, 3º dcha., 28006 Madrid

Tfn. 914029391; fax. 913093685 <novatica@ati.es>

Composición, Edición y Redacción ATI Valencia

Av. del Reino de Valencia 23, 46005 Valencia

Tfn. fax 963330392 <secretari@ati.es>

Administración y Redacción ATI Cataluña

Via Laietana 46, ppal. 1ª, 08003 Barcelona

Tfn. 934129235; fax. 934127713 <secretgen@ati.es>

Redacción ATI Aragón

Lagasca 9, 3-B, 50006 Zaragoza

Tfn. fax 976235181 <secretara@ati.es>

Redacción ATI Andalucía <secretaria@ati.es>

Redacción ATI Galicia <secretgal@ati.es>

Suscripción y Ventas <<http://www.ati.es/novatica/interes.html>>, ATI Cataluña, ATI Madrid

Publicidad

Padilla 66, 3º dcha., 28006 Madrid

Tfn. 914029391; fax. 913093685 <novatica@ati.es>

Imprenta: Derra S.A., Juan de Austria 66, 08005 Barcelona.

Diseño layout: B 15, 154-1975 -- ISSN: 0211-2124; CODEN NOVAEC

Perifoneo: Quejido de la mente - Concha Aras Pérez / © ATI

Diseño: Fernando Agresta / © ATI 2003

editorial

ATI y la profesión informática

> 02

en resumen

Tres objetivos, tres deseos

> 02

declaración de ATI

Ingeniería Informática y Profesión Informática

> 03

noticias IFIP

TC13: Human-Computer Interaction

> 04

Julio Abascal González

monografía

Web universal, ubicua e inteligente

(En colaboración con UPGRADE)

Editores invitados: *Klaus Birkenbihl, Encarna Quesada Ruiz y Pablo Priesca Balbín*

Presentación. Presente y futuro de la World Wide Web

> 05

Klaus Birkenbihl, Encarna Quesada Ruiz, Pablo Priesca Balbín

La Web: ¡qué construcción!

> 08

Klaus Birkenbihl

Uso de la Web en móviles: algunos detalles y tendencias de 2008

> 13

Charles McCathieNeville

La utilización de tecnologías web y teléfonos móviles para el desarrollo social: W3C

> 19

Stephane Boyera

Convirtiendo la semántica en sintaxis

> 22

Bert Bos

Anotación de contenidos en la Web del futuro

> 28

Raúl García Castro, Andrés García Silva

Web Semántica para espacios de conocimiento evolutivos entre iguales

> 33

Ossi Nykänen

La Web del Ayuntamiento de Zaragoza como servicio

> 40

de Atención al Ciudadano

María Jesús Fernández Ruiz, Víctor Morlán Plo

secciones técnicas

Estándares Web

Integración en DENEb de componentes para la conectividad dinámica de los procesos Web. Aplicación a escenarios de gestión de emergencias basados en Sensor Web

> 47

Javier Fabra Caro, Pedro Alvarez Pérez-Aradros, José Ángel Bañares Bañares,

Joaquín Ezpeleta Mateo

Gestión del Conocimiento

Escenarios del conocimiento en la organización

> 53

Juan Baiget Solé

Ingeniería del Software

Una metodología basada en ISO/IEC 15939 para la elaboración de planes de medición de calidad de datos

> 56

Eugenio Verbo García, Ismael Caballero Muñoz-Reja, Ricardo Pérez del Castillo,

Coral Calero Muñoz, Mario Piattini Velthuis

Tecnologías para la Educación

Metadatos de atención contextualizada en entornos de aprendizaje

> 63

Martin Wolpers

Referencias autorizadas

> 67

sociedad de la información

Presente y futuro de ATI

ATInet 2.0

> 73

Entrevista a Gabriel Martí Fuentes

asuntos interiores

Coordinación Editorial / Programación de Novática / Socios Institucionales

> 77

ATI y la profesión informática

Cuando se habla de "profesión" comúnmente se utiliza el concepto que el diccionario de la RAE expresa como "*Empleo, facultad u oficio que alguien ejerce y por el que percibe una retribución*" (Dicc. RAE 22ª edición). Tradicionalmente se opone a la actividad del aficionado "*que cultiva o practica, sin ser profesional, un arte, oficio, ciencia, deporte, etc.*", y se asume que no hay mayor compromiso con el buen hacer, la seguridad y el servicio al cliente o usuario que el que nuestros ingresos dependan de ello.

Qué duda cabe que, como cualquier otra actividad remunerada y ejercida por personas u organizaciones, la profesión informática debe respetar los compromisos establecidos entre proveedor y cliente, amparados por disposiciones legales. En algunos casos, la sociedad puede considerar insuficiente la responsabilidad "a posteriori" en determinadas actividades por lo que adopta regulaciones para aportar ciertas garantías a priori sobre el ejercicio profesional para minimizar el número de reclamaciones sobrevenidas o el riesgo de que algunos efectos no puedan rectificarse o compensarse.

En cualquier caso, éstas deberían centrarse en los elementos de competencia (Dicc. RAE: *Pericia, aptitud, idoneidad para hacer algo o intervenir en un asunto determinado*) que aporten esa garantía o seguridad que reclama la Sociedad. En el ámbito europeo, se ha adoptado la idea del Marco Europeo de Cualificaciones <http://ec.europa.eu/education/policies/educ/eqf/eqf08_en.pdf> aprobado por el Parlamento Europeo el 23 de abril de 2008, donde la competencia se describe en términos de responsabilidad y autonomía sin prejuzgar el método para conseguirla.

Sin embargo, en cuanto a su desarrollo, la profesión informática supone un marco conceptual diferente al de otras disciplinas técnicas. La informática abarca todos los ámbitos de tratamiento automatizado de la información y por ello:

- Es claramente transversal y ubicua toda vez que el procesamiento de información se ha convertido en una herramienta de base en la gran mayoría de sectores y actividades e, incluso, como actividad estratégica para los estados y gobiernos (por ejemplo, la estrategia de Ciberinfraestructura de EE.UU. impulsada por el Presidente Obama) con un efecto multiplicador de negocio y competitividad aunque también con un coste significativo para garantizar sus efectos y su calidad.

- Por otra parte, es una disciplina todavía en rápida evolución y expansión y que va incorporando especializaciones reconocidas internacionalmente como los perfiles formativos de los ACM Currícula <<http://www.acm.org/education/curricula-recommendations>>.

- Por ello, es habitual que las soluciones requieran equipos multidisciplinares y un número creciente de perfiles híbridos de varias disciplinas "tradicionales" para las necesidades que continuamente aparecen.

- En esta disciplina es, si cabe, más crítica la filosofía europea de Life-Long Learning puesto que el reciclaje y actualización es imprescindible para mantener una mínima calidad en el trabajo realizado.

El profesionalismo informático implica la disposición de competencias demostradas y adecuadas en la materia, es decir de conocimientos, habilidades y capacidades que per-

mitan el ejercicio de la profesión informática en las debidas y requeridas condiciones. Lógicamente la enseñanza reglada en sus distintos niveles (formación profesional, ingeniería técnica, ingeniería, grado, master, etc.) aporta una cualificación prominente sin descartar otros medios de formación permanente y reciclaje con garantías de calidad.

En ATI creemos que quienes sientan la vocación por esta profesión deben buscar la formación específica para sus aspiraciones, rechazando tanto la tendencia al autodidactismo no soportado por una formación de base como la concepción no profesional de su trabajo. Y, precisamente para asentar y madurar la profesión, todos los implicados (Administraciones Públicas, asociaciones, empresas, profesionales, etc.) deben fomentar una realidad alejada del amateurismo.

Sin embargo, además de la cualificación, el profesionalismo informático también implica integridad, responsabilidad, "accountability" o rendición de cuentas, y un leal respeto a valores éticos e implicación social frente a retos como la e-inclusión y la alfabetización digital.

Por último, pero no menos importante, se deben señalar algunos de los desafíos que la profesión informática tiene ante sí: a) la mejora de su imagen social y laboral, b) la diversidad de talento, en especial la atracción de las mujeres hacia los estudios reglados donde siguen siendo minoría, c) la defensa de la calidad de los trabajadores cualificados en el competitivo mercado de la tecnología frente a las presiones centradas únicamente en el precio y en valorar tecnología como gasto y no como inversión y d) hacer llegar a la sociedad el valor de la Informática.

en resumen Tres objetivos, tres deseos

Llorenç Pagés Casas

Coordinación Editorial de *Novática*

La eclosión de la Web ha venido acompañada de una loable actitud de sus precursores e ideólogos (agrupados principalmente en torno al W3C) de procurar que todo el mundo pueda participar de sus beneficios.

Así, los tres objetivos presentes en el título de la monografía de este número, que han coordinado **Klaus Birkenbihl** (W3C), **Encarna Quesada Ruiz** (Pez de Babel) y **Pablo Priesca Balbín** (Fundación CTIC), como son la universalidad, ubicuidad e inteligencia de la Web, son en realidad los tres objetivos principales que plantea el W3C.

Son éstos objetivos no solamente alcanzables en el futuro sino que en cierta forma estamos

ya viendo incorporados paulatinamente a nuestro presente. Y que, en nuestra opinión, implican otras metas de nivel más global como serían la contribución relevante de las tecnologías Web a la cultura, la participación y el progreso de la Humanidad.

Y ahí es donde, ya fuera de nuestro ámbito, los informáticos que participamos en estas tecnologías necesitamos del apoyo de nuestros gobernantes y de quienes tienen capacidad de liderazgo y de inversión financiera.

En concreto, necesitamos apoyo a iniciativas en el ámbito educativo y formativo para conseguir que la Web sea ese extraordinario instrumento de difusión de la cultura que promete ser. Necesitamos que la Web sea instrumento de participación, no solamente en discusiones reducidas y en pequeñas decisiones, sino también de los grandes debates

y decisiones de nuestro tiempo. Y, por último, necesitamos favorecer iniciativas y proyectos reales que procuren el logro del bienestar de las personas, en especial el de aquellas que viven en condiciones más difíciles.

Será ésta la manera en la que verdaderamente veamos materializado el "sueño" de los apóstoles de la Web. Y el nuestro propio que sería poder editar en el futuro otra monografía bajo el título "*La Web, instrumento clave para la cultura, la participación y el progreso*".



Ingeniería Informática y Profesión Informática

En los últimos meses se está produciendo un intenso **debate sobre el papel de la Ingeniería Informática en España**, acompañado de numerosos pronunciamientos y movilizaciones, estas últimas protagonizadas principalmente por los estudiantes de dichas Ingenierías.

ATI

La Asociación de Técnicos de Informática (ATI) ha seguido con mucho interés ese debate, y ha participado en él interviniendo en eventos de todo tipo y dando su opinión a través de su sitio web <<http://www.ati.es/>> y de su revista **Novática**, una participación obligada por tratarse de la mayor y más antigua de las asociaciones de profesionales informáticos de nuestro país, de los que desde hace años es representante oficial y exclusivo a nivel internacional (en Europa a través de CEPIS, <<http://www.cepis.org/>>, y en el mundo a través de IFIP, <<http://www.ifip.org/>>).

ATI, como asociación de ámbito estatal de la que forman parte cerca de 4.000 profesionales informáticos – casi la mitad de ellos titulados en Ingeniería Informática Superior y Media o Licenciados en Informática, y los demás titulados en otras materias y de otros niveles como Formación Profesional, así como también no titulados y estudiantes de Informática en sus diversas vertientes –, tiene el propósito de seguir contribuyendo a dicho debate de forma positiva y constructiva y por ello *desea hacer públicas las siguientes reflexiones, abiertas a las aportaciones de todos los profesionales informáticos, independientemente de que sean titulados en Ingeniería Informática o no, o de que sean miembros o no de nuestra asociación.*

PREMISAS

Partiendo del objetivo irrenunciable de que la profesión informática en su conjunto progrese y contribuya al desarrollo armónico de la Sociedad de la Información, consideramos que las siguientes **premisas** son objetivamente válidas:

1. La Ingeniería Informática debe tener el mismo grado de reconocimiento académico, profesional y social que cualquier otra Ingeniería.
2. La Ingeniería Informática debe seguir siendo considerada plenamente como una rama de Ingeniería específica y diferenciada, no absorbible por otras Ingenierías, con sus propios planes de estudio adaptados al continuo avance tecnológico y a los requerimientos sociales y del mercado.
3. Las titulaciones de Ingeniería Informática deben ser reconocidas como el nivel formativo más completo en el ámbito de las Tecnologías de la Información.
4. Como consecuencia de lo anterior, los Ingenieros Informáticos deben contar con un conjunto razonable de atribuciones profesionales compatible con las Directivas europeas sobre la materia.

Creemos que **también son válidas estas otras premisas**, que complementan a las anteriores sin contraponerse a ellas:

- a. La profesión informática es multidisciplinar y abarca también a

multitud de profesionales que no son Ingenieros Informáticos o Licenciados en Informática (en concreto, titulados de otras carreras con especialidades informáticas, graduados de la rama de Informática de Formación Profesional, poseedores de certificados otorgados por diversas empresas y organizaciones, e incluso personas sin titulación alguna).

- b. En España y en todo el mundo, el modelo dominante se caracteriza por la diversidad de titulaciones y procedencias de los profesionales informáticos, diversidad que existe no sólo por razones históricas sino también por necesidades imperiosas del mercado laboral y de reciclaje continuo de conocimientos, y que enriquece a nuestra profesión.

- c. Los intereses de los profesionales informáticos que no son Ingenieros Informáticos deben ser también objeto de reconocimiento y tutela, tanto por empresas como por Administraciones Públicas.

- d. En el ámbito profesional y laboral ha existido siempre una convivencia armoniosa entre los profesionales informáticos que cuentan con títulos de Ingeniería Informática y quienes no los poseen, convivencia que debe ser preservada frente a visiones exclusivistas y discriminatorias.

- e. En correspondencia con la diversidad de la profesión informática, los profesionales informáticos españoles están representados por asociaciones profesionales (ATI, ALI y AI2 fundamentalmente) y Colegios de Ingenieros de Informática.

CONCLUSIONES

Como consecuencia de lo anterior ATI quiere expresar lo siguiente:

- *Nuestra asociación tiene la decidida voluntad de participar, aportando sus ideas y propuestas, en todas las instancias en las que se debaten los asuntos que se refieren al presente y el futuro de la profesión informática, junto a los diversos actores implicados (Colegios de Ingenieros de Informática y asociaciones profesionales informáticas, Conferencia de Decanos y Directores – CODDI –, representantes de alumnos – RITSI –, y Administraciones Públicas, así como asociaciones empresariales y sindicales del sector).*

- *Nuestra asociación defenderá por todos los medios lícitos a su alcance los intereses del conjunto de los profesionales informáticos, tratando de armonizar los legítimos derechos y aspiraciones de los diversos colectivos que la componen. En este sentido:*

- defenderá que se otorguen a los titulados en Ingeniería Informática Superior y Media las atribuciones que sean compatibles con la legislación europea sobre la materia, lo que implica que no puedan ser omnicomprendivas ni exclusivistas;
- defenderá también que el acceso a las funciones y niveles de carácter informático, en empresas y Administraciones Públicas, siga abierto, como hasta ahora, a los profesionales informáticos que cuenten la cualificación adecuada y con el nivel de titulación requerido para el puesto, independientemente de que esta titulación sea o no en Ingeniería Informática.

30 de marzo de 2009

TC13: Human-Computer Interaction

Julio Abascal González

*Laboratorio de Interacción Persona-Computador para Necesidades Especiales,
Universidad del País Vasco-Euskal Herriko Unibertsitatea; Representante de ATI en el TC13*

<julio.abascal@ehu.es>

1. Historia y objetivos del TC13 HCI

El profesor británico Brian Shackel, que colaboraba en IFIP desde 1970, reconstruyó en 1981 el grupo de trabajo WG 6.3 (*Man-Computer Communication*). En 1982 este WG se transformó en el *Task Group on Human-Computer Interaction*, que organizó el primer congreso de la serie INTERACT.

En 1989 pasó a ser el Comité Técnico TC13 por decisión de la Asamblea General de IFIP.

Entre sus objetivos fundacionales están:

- Fomentar la investigación empírica (uso de metodologías válidas y fiables).
- Promover el uso de los conocimientos y métodos de las ciencias humanas en el análisis, diseño y evaluación de sistemas informáticos.
- Promover una mejor comprensión de la relación entre los métodos formales de diseño y la usabilidad y aceptación del sistema.
- La elaboración de directrices, modelos y métodos para ayudar a los diseñadores a ofrecer mejores sistemas informáticos orientados a la persona.
- Cooperar con otros grupos, dentro y fuera de la IFIP, a fin de promover la orientación al usuario y la "humanización" en el diseño de sistemas.

Su ámbito de trabajo y áreas de estudio incluyen:

- Los problemas que tienen las personas con los computadores.
- El impacto de los computadores sobre las personas en los contextos individual, organizativo, social y de aprendizaje.
- Los factores determinantes de la utilidad, usabilidad, satisfacción y aceptabilidad.
- El modelado del usuario como una ayuda a un mejor diseño del sistema.
- La armonización de la computadora a las características y necesidades de los usuarios.

2. Estructura del TC13 HCI

El TC13 se reúne una vez al año (dos en los años impares cuando se celebra INTERACT). Cuenta actualmente con 43 miembros: 35 representantes de las sociedades nacionales de informática de 34 países¹, los presidentes de los 7 grupos de trabajo y un experto invitado.

Actualmente está presidido por la profesora danesa Annelise Mark Pejtersen asistida por

dos vicepresidentes, un tesorero y un secretario.

Para el trabajo especializado el TC13 cuenta con seis *Working Groups*:

- WG 13.1: Education in HCI and HCI Currículo.
- WG 13.2: Methodologies for User-Centred Systems Design.
- WG 13.3: Human-Computer Interaction and Disability.
- WG 2.7(13.4): User Interface Engineering.
- WG 13.5: Human Error, Safety, and System Development.
- WG 13.6: Human-Work Interaction Design group.
- WG 13.7: HCI and Visualisation.

3. Actividades del TC13

El TC 13 organiza cada dos años el prestigioso congreso INTERACT². Este año se celebrará entre el 24 y el 28 de agosto en Upsala (Suecia)³.

Por otro lado, es frecuente que cada sociedad nacional que acoge una de las reuniones del TC13 organice un seminario para intercambio de ideas entre sus miembros y los expertos locales. Los dos últimos seminarios se celebraron en Salamanca (2007) "*Human-Computer Interaction Research and Development Challenges*"⁴ y Atenas (2009) "*Human-Computer Interaction: theory and practice of design of usable and accessible technologies*".

En los últimos años el TC 13 ha publicado varios libros, la mayoría de los cuales son resultados de congresos y seminarios organizados por sus miembros.

4. Representación española en el TC13

La Federación Española de Sociedades de Informática (FESI), que era por entonces corresponsal de IFIP en España, me nombró en mayo de 1980 representante en el TC13. Desde esta fecha el TC13 ha celebrado 28 reuniones plenarios de las que he asistido personalmente a 26.

Durante estos años he participado activamente en las actividades del TC13, organizando reuniones (San Sebastián 1999 y Salamanca 2007), participando en el comité de programa y de organización de la mayoría de los INTERACT celebrados hasta ahora y copresidiendo el congreso INTERACT 2007 celebrado en Río de Janeiro, etc. En 1993

propuse la creación del Grupo de trabajo WG 13.3 *HCI and Disability*, que presidí hasta 1999⁵. Soy miembro del WG 13.1 *Education in HCI and HCI Curriculum*.

5. La comunidad española de HCI

En 1990 el profesor Jesús Lorés impulsó la creación de la Asociación de Interacción Persona-Ordenador (AIPO) para propiciar la colaboración de todos aquellos que trabajan en este ámbito en el área iberoamericana. Esta sociedad, muy activa especialmente en España, mantiene una web⁶ y viene organizando el congreso anual Interacción desde 2000. AIPO está tratando con ATI su colaboración en las actividades de Interacción Persona-Ordenador en IFIP. Por mi parte he tratado de actuar como puente entre ATI, AIPO e IFIP TC 13 para promocionar y dar a conocer la actividad española en HCI.

6. Conclusión y llamada a la participación

A pesar de que en España hay una intensa actividad en Interacción Persona-Ordenador (IPC), tal como lo demuestra la extensa participación en los congresos Interacción, ésta no está suficientemente reflejada en los foros internacionales. Así pues, uno de nuestros objetivos es tratar de aumentar la presencia de la IPC española en el exterior.

Con respecto al TC13, actualmente no hay otros miembros españoles que colaboren en sus grupos de trabajo⁷, por lo que animo a todas las personas que pudieran estar interesadas en participar a ponerse en contacto conmigo. El nivel de actividad en estos grupos de trabajo es diverso y están, en general, abiertos a la participación de quienes estén interesados. La pertenencia oficial a un WG suele requerir un periodo de implicación activa en el trabajo del mismo.

¹ USA cuenta con dos representantes: IEEE y ACM.

² De tipo A en el 2007 Australian Ranking of ICT Conferences CORE <<http://www.core.edu.au/>>.

³ <<http://www.interact2009.org/>>.

⁴ Una selección de los artículos presentados se publicó posteriormente en el Journal of Universal Computer Science Vol. 14 Issue 16 <http://www.jucs.org/jucs_14_16>.

⁵ Por lo que me fue concedido el *IFIP Silver Core Award* en 1998.

⁶ <<http://www.aipo.es/>>.

⁷ El profesor Lorés fue miembro del WG13.1 hasta su fallecimiento.

Klaus Birkenbihl¹, Encarna Quesada Ruiz², Pablo Priesca Balbín³

¹ICT Media GMBH, Coordinador de Operaciones de las oficinas del W3C; ²Directora General de Pez de Babel; ³Director General de la Fundación CTIC

<Klaus.Birkenbihl@ict-media.de>, <equesada@pezdebabel.com>, <info@fundacionctic.org>

En los últimos años son muchos los cambios que estamos presenciando no sólo en Internet, sino también en una de las herramientas que, desde su origen, más ha revolucionado el mundo. Este es el caso de la World Wide Web.

Estos cambios, que afectan a nuestra forma de comunicarnos, interactuar y hacer negocios, acompañados por una auténtica revolución en lo que a dispositivos de acceso a la Web se refiere, han generado profundas transformaciones en el conjunto de la sociedad.

Hay que admitir que la Web se ha convertido en una herramienta imprescindible en nuestro día a día y, cada vez más, demandamos servicios innovadores, sobre la base de tecnologías web, que sean rápidos, robustos, atractivos e interoperables; sin olvidar, por supuesto, que sean accesibles desde cualquier lugar, independientemente del dispositivo de acceso utilizado. Yes aquí donde los estándares tienen mucho que decir y que ofrecer, con el objetivo de garantizar una web universal, ubicua e inteligente.

Entrando en el tema del acceso a Internet desde dispositivos móviles, nos encontramos con el surgimiento de una gran variedad de nuevos dispositivos, con características muy diversas, a lo que hay que sumar un número creciente de usuarios, también con diferentes características, que cada vez más demanda una Web accesible, desde cualquier terminal o punto de acceso, lo que genera la necesidad de crear aplicaciones que favorezcan la independencia de dispositivo y tengan en cuenta la heterogeneidad de las preferencias y limitaciones de los distintos usuarios y de los entornos desde los que interactúan con la Web.

Acceder a la Web desde cualquier lugar se está convirtiendo en una realidad, pero esta realidad se enfrenta a ciertas limitaciones cuando sobre la base de su desarrollo no se encuentra la idea de universalidad. La Web, creada por Tim Berners-Lee, fue diseñada para ser universal. Y es a través de este concepto de universalidad como se hacen posibles la comunicación, el intercambio de información y el desarrollo de aplicaciones dinámicas para todo el mundo, desde cualquier lugar, en cualquier momento y utilizando cualquier dispositivo, evitando así problemas de interoperabilidad y usabilidad.

Presentación Presente y futuro de la World Wide Web

Editores invitados

Klaus Birkenbihl es el fundador y Director General de ICT Media GMBH. Actualmente, es coordinador de operaciones a nivel internacional de todas las oficinas del W3C (*World Wide Web Consortium*). Klaus es licenciado en Matemáticas por la Universidad de Bonn (Alemania), donde comenzó su carrera profesional en el GMD National Research Lab for Information Technology. Posteriormente, trabajó en el Instituto de Medios de Comunicación de Fraunhofer-Gesellschaft. Klaus fue profesor de Informática en la Universidad de Colonia y en la Universidad de Ciencias Aplicadas Bonn-Rhein-Sieg. Comenzó a trabajar en temas de redes informáticas en 1983 cuando era Director de GMD's Computing Center en Bonn. Se unió a la Red Académica y de Investigación Europea (*European Academic Research Network* - EARN, 1983) como miembro fundador. Fue también Director Adjunto en EARN Alemania (1987-1990), y miembro del Comité Directivo de EASnet. Klaus fue también administrador y coordinador de la división alemana de la Internet Society, donde continúa ahora en el comité de supervisión. En 1997, se lanzó la Oficina del W3C en Alemania y estuvo en el Comité Asesor del W3C desde 2002-2005. Klaus dejó Fraunhofer-Gesellschaft en el año 2005, comenzando así su trabajo dentro del equipo de W3C, así como su propia empresa.

Encarna Quesada Ruiz es la fundadora y Directora General de Pez de Babel, empresa que nace con un objetivo principal: proporcionar soluciones integrales a las necesidades de globalización e internacionalización de todo tipo de empresas. Previamente, fue Directora de la Oficina Española del W3C (*The World Wide Web Consortium*), donde centró su trabajo en la divulgación de las tecnologías del W3C dentro de la comunidad hispano-hablante. La función principal de la oficina ha sido promover la difusión, la adopción y la utilización de las Recomendaciones del Consorcio W3C, con el objetivo de guiar la Web hacia su máximo potencial, así como fomentar la adscripción al W3C de organizaciones clave para la creación de futuras recomendaciones. Licenciada en Sociología y especializada en Sociedad de la Información, Encarna ha trabajado en las áreas de internacionalización, globalización y localización durante varios años en empresas multinacionales de gran peso dentro del mundo de la tecnología. Es coordinadora de la sección técnica "Estándares Web" de *Novática* desde su creación hace aproximadamente dos años.

Pablo Priesca Balbín es Director General de la Fundación CTIC (Centro Tecnológico de la Información y de la Comunicación). Es licenciado en Psicología por la Universidad de Oviedo, y experto en sociedad de la información y desarrollo territorial. Lleva más de 14 años trabajando en proyectos informáticos. Dentro de la Fundación CTIC su objetivo principal es promover y estimular actividades, relacionadas con el desarrollo de la Informática en todos los campos de la vida económica y social, que conduzcan al impulso y fortalecimiento de la Sociedad de la Información.

Son ya muchas las organizaciones, empresas y desarrolladores que utilizan estándares a la hora de diseñar sitios web, pero queda todavía mucho camino por recorrer. Este es el caso, por ejemplo, de la accesibilidad Web. Es cierto que la accesibilidad ha experimentado en los últimos años un avance importante. Prueba de ello es que, cada vez más, tanto la Administración Pública como empresas y desarrolladores Web se preocupan por proporcionar a los usuarios servicios "accesibles", que sigan los estándares elaborados por el W3C (*The World Wide Web Consortium*). Esta preocupación por la accesibilidad nace del derecho de todos los ciudadanos a acceder a la información electrónica a través de la Web; es decir, el derecho a poder disfrutar de una web universal

y para todos.

Es complicado hablar con precisión sobre la futura evolución de la Web ya que, como hemos visto, en la Web se produce una compleja interacción entre tecnología y sociedad. Lo que sí se puede decir, y esto es algo que los autores de esta monografía mencionan, es que la Web seguirá tendiendo hacia la constitución de una plataforma cada vez más universal e independiente, tanto del dispositivo de acceso como de las características del usuario, y con mecanismos de estructuración de la información que permitirán vincular los datos entre sí, a través de formatos estándares, formando de manera conceptual una gran base de datos.

Esta monografía, que lleva por título "Web universal, ubicua e inteligente", tiene como objetivo mostrar los avances producidos hasta este momento y los posibles desarrollos futuros en la web, a través de los ojos de varios expertos en las áreas de accesibilidad, movilidad y web semántica. Sus análisis y opiniones nos permitirán conocer más acerca del estado actual y del futuro de aspectos claves en el crecimiento de la Web.

La monografía comienza con un artículo de **Klaus Birkenbihl**, uno de los editores invitados, que bajo el título "*La Web: iqué construcción!*" proporciona una introducción clara, precisa y detallada sobre la Web, sus estándares y sus tecnologías subyacentes, poniendo especial hincapié en aquellas desarrolladas dentro del W3C.

Si algo venimos demandando como usuarios de la Web desde hace años es la accesibilidad a la información independientemente del tipo de hardware, software, infraestructura de red, idioma, cultura, localización geográfica y capacidades de los usuarios, lo que desde el W3C, consorcio internacional de la Web, se viene promoviendo como uno de sus pilares principales a la hora de desarrollar tecnologías.

En esta línea de accesibilidad desde cualquier lugar y dispositivo, publicamos las opiniones

de expertos como **Charles McCathieNevile**, Director de Estándares de Opera Software, quien nos habla de las últimas tendencias en lo que a acceso a la web desde dispositivos móviles se refiere. Por su parte, **Stephane Boyera**, responsable del área de Web Móvil en países en desarrollo dentro del W3C, nos presenta una perspectiva diferente de la misma temática, abordando en su artículo el uso de las TIC (Tecnologías de la Información y la Comunicación) desde teléfonos móviles como una posible solución para reducir la brecha digital en países en desarrollo.

Bert Bos, Coordinador de las Actividades de Estilo y Matemáticas del W3C, dedica un artículo a hablar de la Web como un sistema de comunicación con características que invitan al procesamiento automático, de forma que usando estas técnicas el usuario obtenga una versión mejorada, que combine recursos y muestre información adaptada a sus necesidades. La potencialidad de la Web queda reforzada mediante la expresión de la información en lenguajes formales que un ordenador pueda manipular.

En conexión con la temática de la web inteligente, **Raúl García Castroy** y **Andrés García Silva** del Departamento de Inteligencia Artificial de la Facultad de Informática de la Universidad Politécnica de Madrid nos presentan el estado

y las tendencias actuales de los enfoques y herramientas de anotación de recursos Web en el contexto de las áreas de la Web Semántica, la Web 2.0 y lo que se ha denominado la Web 3.0, entendiendo esta última como la fusión de la Web Semántica y la Web 2.0.

Desde otra universidad, la *Tampere University of Technology* (Finlandia), **Ossi Nykänen** trata sobre la Web Semántica como una tecnología fundamental para aplicaciones de explotación de datos a través de semántica comprensible para las máquinas.

Finalmente, desde un punto de vista más pragmático, **María Jesús Fernández Ruiz** (responsable de la web municipal del Ayuntamiento de Zaragoza) y **Victor Morlán Plo** (Ingeniero Técnico en Informática del Ayuntamiento de Zaragoza) nos hablan de lo que en estos días se está convirtiendo para las administraciones públicas en un auténtico reto, la accesibilidad a la información y el acercamiento de la administración pública al ciudadano a través de servicios telemáticos accesibles e inteligentes.

Rememorando el título del artículo introductorio de **Klaus Birkenbihl**, la Web es exactamente eso, una "construcción gigantesca" donde los conceptos de universalidad, ubicuidad e inteligencia son, o al menos deberían ser, la base y los pilares de su futuro crecimiento.

Referencias útiles sobre "Web universal, ubicua e inteligente"

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

Libros/Publicaciones

- **Tim Berners-Lee.** "Weaving the web. The original design and ultimate destiny of the World Wide Web, by its inventor". Harper San Francisco (1999). ISBN: 0062515861.
- **Shawn Lawton Henry.** "Just Ask: Integrating Accessibility Throughout Design". Lulu.com (2007). ISBN: 1430319526.
- **J. Davies, D. Fensel, F. van Harmelen, John Wiley.** "Towards the Semantic Web: Ontology-Driven Knowledge Management". Chichester, UK: John Wiley & Sons, Ltd. (2002).
- **S. Powers.** "Practical RDF". O'Reilly Media (2003). ISBN: 0-596-00263-7.
- **D. Fensel, J. A. Hendler, H. Lieberman, W. Wahlster.** "Spinning the Semantic Web". MIT press (2003). ISBN: 0-262-06232-1.
- **G. Antoniou, F. van Harmelen.** "Semantic Web Primer" (2004). ISBN: 026201210.

- **Tomi Ahonen.** "Mobile as 7th of the Mass Media: Cellphone, cameraphone, iPhone, smartphone". futuretext (2008). ISBN: 0955606950.
- **Fundación Orange.** "eEspaña 2008: Informe anual sobre el desarrollo de la sociedad de la información en España". Fundación Orange (2008). Depósito Legal: M-32980-2008.
- **Fundación Orange.** "Estudio comparativo 2008 de los servicios públicos online en las comunidades autónomas españolas". Fundación Orange (2008).
- **European Commission.** Joint Research Centre. "The future of eGovernment: An exploration of ICT-driven models of eGovernment for the EU in 2020". European Commission (2007). ISSN: 1018-5593.

Enlaces Web

- Consorcio World Wide Web <<http://www.w3.org/>>.
- *Pila tecnológica del W3C* <<http://www.w3.org/2004/10/RecsFigure.png>>.
- *Sobre las tecnologías del W3C* <<http://www.w3.org/Consortium/technology>>.
- *Publicaciones e informes técnicos del W3C* <<http://www.w3.org/TR/>>.
- *Sobre las actividades del W3C* <<http://www.w3.org/Consortium/activities>>.

- Iniciativa de accesibilidad Web del W3C <<http://www.w3.org/WAI/>>.
- Pautas de accesibilidad para el contenido en la Web del W3C (WCAG) 2.0 <<http://www.w3.org/TR/WCAG20/>>.
- Componentes esenciales de la accesibilidad <<http://www.w3.org/WAI/intro/components.php>>.
- Iniciativa de Web móvil del W3C <<http://www.w3.org/Mobile/>>.
- Webmóvil para el desarrollo social (MW4D) <<http://www.w3.org/2008/MW4D/>>.
- Proyecto MobiWeb2.0: «Mobile Web 2.0» <<http://www.w3.org/2008/MobiWeb20/>>.
- El rincón del desarrollador - *Developer's corner* (herramientas y referencias del W3C para desarrollos de web móvil) <<http://www.w3.org/Mobile/Dev>>.
- Especificaciones técnicas desarrolladas por la Iniciativa de Web móvil del W3C <<http://www.w3.org/Mobile/Specifications>>.
- Planet Mobile Web (blogs relacionados con la web en móviles) <<http://www.w3.org/Mobile/planet>>.
- Wiki del grupo de Web móvil en países en desarrollo del W3C <<http://www.w3.org/2008/MW4D/wiki/Stories>>.
- Aplicaciones web ubicuas del W3C <<http://www.w3.org/2007/uwa/>>.

- Actividad de Web Semántica del W3C <<http://www.w3.org/2001/sw/>>.
- Archivo de publicaciones de Web Semántica del W3C <<http://www.w3.org/2001/sw/publications.html>>.
- Wiki del grupo de W3C de Web Semántica <<http://www.w3.org/2006/07/SWD/wiki/FrontPage>>.
- eGovernment en el W3C <<http://www.w3.org/2007/eGov/>>.
- TIC para la Administración y los servicios públicos (Sociedad de la Información Europea) <http://ec.europa.eu/information_society/activities/egovernment/index_en.htm>.

Artículos/Entrevistas

- **Tim Berners-Lee.** "What is the future of the internet". Entrevista de Radio 4 de la BBC (julio 2008). <http://news.bbc.co.uk/today/hi/today/newsid_7496000/7496976.stm>
- **Tim Berners-Lee.** "Sir Tim Berners-Lee Talks About the Semantic Web". Talis Group (febrero 2008) <http://talis-podcasts.s3.amazonaws.com/twt20080207_TimBL.html>.
- **L. Feigenbaum, I. Herman, T. Hongsermeier, E. Neumann, S. Stephens.** "The Semantic Web in Action". Scientific American, 297(6), pp. 90-97, (diciembre 2007). <<http://thefigtrees.net/lee/sw/sciam/semantic-web-in-action>>

- **Tim Berners-Lee.** "Q&A with Tim Berners-Lee". BusinessWeek (abril 2007). <http://www.businessweek.com/technology/content/apr2007/tc20070409_961951.htm>.
- **Tim Berners-Lee.** "Tim Berners-Lee on the Semantic Web". Video Technology Review (marzo 2007). <<http://www.technologyreview.com/Infotech/18451/?a=f>>.
- **Ivan Herman.** "Le Web Sémantique: un interview avec Ivan Herman". Xavier Borderie en Journal du Net, développeurs (junio 2006). <<http://www.journaldu.net.com/developpeur/itws/060608-itw-w3c-herman.shtml>>.

Congresos/Eventos

- Conferencia de la World Wide Web – WWW 2009 (abril 2009). Madrid, España. <<http://www.www2009.org/>>.
- Congreso mundial de móviles GSMA (febrero). Barcelona, España. <<http://www.mobileworldcongress.com/>>.
- 2º Congreso anual de IIR sobre Internet móvil (marzo 2009). Viena, Austria. <<http://www.iir-events.com/IIR-conf/Telecoms/EventView.aspx?EventID=2029>>.
- Perspectiva de África sobre el papel de las tecnologías móviles en el incremento del desarrollo económico y social (abril 2009). Maputo, Mozambique. <http://www.w3.org/2008/10/MW4D_WS/>.

- VII Taller MobEA de widgets móviles (abril 2009). Madrid, España. <<http://www.research.att.com/~rjana/mobe2009.htm>>.
- W4A2009 – Conferencia de Accesibilidad Web (abril 2009). Madrid, España. <<http://www.w4a.info/>>.
- Foro europeo de accesibilidad (marzo 2009). Frankfurt, Alemania. <<http://eafra.eu/>>.
- III Foro europeo de accesibilidad (marzo 2009). Paris, Francia. <http://inova.snv.jussieu.fr/evenements/colloques/mobel/Description/index_en.php>.
- VI Conferencia europea de Web Semántica (mayo-junio 2009). Heraklion, Grecia. <<http://www.eswc2009.org/>>
- VIII Conferencia Internacional de Web Semántica (octubre 2009). Washington, DC. <<http://iswc2009.semanticweb.org/>>

W3C

- *W3C y su pila de tecnologías.* <<http://www.w3.org/Consortium/technology>>.
- *Informes técnicos y publicaciones.* <<http://www.w3.org/TR/>>.
- *Actividades.* <<http://www.w3.org/Consortium/activities>>.

Program Committee Chair

* Idit Keidar, Technion

Program Committee

- * Ittai Abraham, Microsoft Research SCV
- * Yehuda Afek, Tel-Aviv University
- * Marcos K. Aguilera, Microsoft Research SCV
- * James Aspnes, Yale
- * Christian Cachin, IBM Zurich Research Laboratory
- * Gregory V. Chockler, IBM Haifa Research Laboratory
- * Carole Delporte-Gallet, University of Paris Diderot
- * Pascal Felber, University of Neuchatel
- * Seth Gilbert, École Polytechnique Fédérale de Lausanne
- * Danny Hendler, Ben-Gurion University
- * Ricardo Jiménez-Peris, Universidad Politécnica de Madrid
- * Zvi Lotker, Ben-Gurion University
- * Thomas Moscibroda, Microsoft Research
- * David Peleg, Weizmann Institute
- * Eric Ruppert, York University
- * Elad M. Schiller, Chalmers U. of Technology and Gothenburg U.
- * Mark R. Tuttle, Intel
- * Robbert van Renesse, Cornell University
- * Jay J. Wylie, HP Labs
- * Lidong Zhou, Microsoft Research Asia

Organizing Committee Chairs

- * Vicent Cholvi Juan, Universitat Jaume I
- * Antonio Fernández Anta, Universidad Rey Juan Carlos

Keynote Speakers

- * Lorenzo Alvisi, University of Texas at Austin
- * Nir Shavit, Tel-Aviv University
- * Willy Zwaenepoel, École Polytechnique Fédérale de Lausanne

September 23-25, 2009

DISC 2009

23rd INTERNATIONAL SYMPOSIUM ON
DISTRIBUTED COMPUTING
ELCHE/ELX ■ SPAIN

<http://disc2009.gsync.es>

Submission deadline: May 2, 2009



Klaus Birkenbihl

Director General de ICT Media GMBH,
Coordinador de Operaciones de las oficinas
del W3C

<Klaus.Birkenbihl@ict-Media.de>

La Web: ¿qué construcción!

Traducción: Encarna Quesada Ruiz (editora invitada de la monografía)

1. Introducción

Hace 17 años que la Web apareció por primera vez y ha cambiado mucho desde entonces. Continúa cambiando su tecnología, los servicios que proporciona y la forma en que son utilizados. Hay tantas opciones que un solo punto de vista sobre lo que ofrece es muy complicado. Por este motivo, los editores invitados a esta monografía decidimos escoger algunos desarrollos que consideramos relevantes para ofrecer una perspectiva sobre lo que está por venir.

- Web desde cualquier lugar – la Web Móvil.
- Web para todo el mundo – la Web accesible.
- Web de información estructurada – la Web Semántica.
- Web de participación y de comunicación – colecciones, blogs y foros.
- Web como servicio – Aplicaciones Web Enriquecidas (*Rich Web Applications*).
- Web para los negocios – Servicios Web.
- Web para multimedia – video y voz en la Web.

Todo esto debería apoyarse en un conjunto común de tecnologías y estándares. Las funcionalidades de los distintos aspectos deberían interrelacionarse entre ellas.

Otros temas son importantes para la mayoría de estos aspectos. Son horizontales por naturaleza y algunos están sujetos a los diferentes puntos de vista culturales.

- Privacidad en la Web.
- Seguridad y confianza en la Web.
- Tratamiento de la propiedad intelectual.
- Clasificación de contenido y valoración.
- Accesibilidad de la Web.

En este artículo no vamos a cubrir las diferentes perspectivas en temas como la privacidad, propiedad intelectual y clasificación de contenidos. Éstos están todavía sujetos a discusiones políticas y sociales. Las tecnologías web deberían acercarse a todo aquello que pareciera beneficioso en un contexto social y político.

La mayoría del trabajo en estándares y tecnologías que se describe en este artículo es o fue realizado en el W3C (*World Wide Web Consortium*), consorcio creado en 1994 por el inventor de la Web Tim Berners-Lee.

2. Los cimientos

Tres tecnologías utilizadas de forma combinada han hecho posible el éxito de la Web:

- Una sintaxis común para documentos

Resumen: este artículo ofrece un resumen de los estándares y tecnologías web, haciendo hincapié en aquellos desarrollados por el W3C. Algunas de las investigaciones y desarrollos que se están llevando a cabo alrededor de ellos son presentados en esta edición especial sobre tecnologías web. Este artículo introductorio, muestra el contexto de utilización y el propósito de las diferentes tecnologías, ilustrando la relación y dependencia entre ellas.

Palabras clave: accesibilidad, aplicaciones web, comunicación por voz, privacidad, seguridad, servicios web, web móvil, web semántica, W3C.

Autor

Klaus Birkenbihl es el fundador y Director General de ICT Media GMBH. Actualmente, es coordinador de operaciones a nivel internacional de todas las oficinas del W3C (*para más detalles sobre su biografía, ver el artículo de presentación*).

que puedan contener hipervínculos (HTML).

- Una sintaxis para la localización de documentos (o fragmentos de documentos) y la opción de utilizar esa sintaxis dentro de HTML para construir hipervínculos (URL).
- Un protocolo para enviar peticiones y respuestas entre clientes (normalmente navegadores) y servidores (HTTP).

Estas tecnologías permiten ofrecer una web global de documentos que (desde la perspectiva de un usuario) proporciona una interfaz gráfica y una forma muy sencilla de ir de un documento a otro relacionado con él, utilizando enlaces proporcionados en el primer documento.

Esta web de documentos, pronto se convirtió en una víctima de su propio éxito. Había una gran cantidad de documentos y el deseo de integrar la información desde bases de datos. Ofertas de tiendas, horarios de vuelos... etc., necesitaban extensiones del concepto. A corto plazo, fue el *scripting* (del lado del servidor y del cliente) y CGI los que permitían páginas web dinámicas. Pero hacían falta también soluciones más estructuradas y sistemáticas, por lo que tenía que hacerse una generalización y extensión de las tres tecnologías originales. En términos de estructuración de documentos y datos, se introdujeron dos conceptos: XML Infoset y Grafos RDF.

XML Infoset es una generalización del concepto HTML (aunque siendo más estricto en la sintaxis). Al igual que HTML, XML estructura un documento, u otras entidades, en un árbol con nodos denominados (etiquetados). HTML utiliza un conjunto predefinido de nombres de nodos y reglas para restringir su uso en un documento. XML Infoset no tiene estas restricciones. La introducción de restricciones específicas puede ser utilizada para

crear nuevas estructuras de datos o lenguajes de etiquetado. De hecho, miles de estos lenguajes XML existen actualmente con diferentes propósitos. Unas pocas tecnologías de ayuda surgieron a partir de XML, tales como:

- Espacios de nombres XML (*XML Namespaces*) como forma para separar un espacio de etiquetas del otro.
- Esquemas XML (*XML Schemas*) para permitir la definición de restricciones para especificar un nuevo lenguaje XML.
- XQuery/Xpath como mecanismo para navegar o mandar una petición a documentos XML.
- XML Base para administrar URL relativos.
- XPointer para permitir referencias a documentos XML y fragmentos.
- DOM, una interfaz abstracta, para procesar XML.

XML se basa en el concepto de definir entidades complejas, componiéndolas a través de otras más simples. Los Grafos RDF (*RDF Graphs*) tratan de definir entidades, describiendo sus relaciones con otras. Un grafo RDF está compuesto por *triples* de la forma *sujeto predicado objeto*. Donde *sujeto* (un recurso) denota algo que se quiere definir, *predicado* define la relación (propiedad), y *objeto* (un recurso) es algo por lo que la relación se mantiene. Los *objetos* (e incluso los *predicados*) en un triple pueden aparecer como *sujetos* en otros, por lo que podemos ver enormes redes de relaciones. Las propiedades que tienen su origen en las matemáticas o en la lógica (como "es un subconjunto de", "implica"...) basan las peticiones complejas y el razonamiento en esas redes. Del mismo modo que XML, RDF viene con un conjunto de tecnologías asociadas:

- RDFS es un lenguaje de esquema que

define un vocabulario que da significado a algunos recursos (como *Class*, *Literal*, *Datatype...*) y propiedades (como *subClassOf*, *subPropertyOf*, *domain*, *range...*) Esto abre un camino de estandarización para expresar conjuntos o propiedades lógicas relacionadas. OWL complementa y extiende estas definiciones.

- SPARQL es un lenguaje de búsqueda para RDF.

A pesar de que XML y RDF siguen diferentes conceptos (jerárquicos y relacionales) a la hora de estructurar información, no están completamente separados el uno del otro. RDF/XML es un lenguaje XML especial para representar (serializar) Grafos RDF (que no están entre los más comprensibles para los seres humanos). Los tipos de datos básicos utilizados en RDF provienen de *XML Schema*.

Junto a las tecnologías introducidas aquí hay un conjunto denominado "Principios de Arquitectura Web" que constituyen reglas sobre cómo utilizar estas tecnologías para construir nuevas.

3. Principios básicos

Queremos que la Web sea universal. Es una petición muy noble con bastantes implicaciones.

En conclusión significa que todo el mundo debería poder utilizar la Web con una amplia gama de dispositivos en cualquier idioma que conozca. Es obvio que esto tiene sus limitaciones (por ejemplo, el autor de este artículo probablemente no sea capaz de comprender un artículo sobre microbiología en chino). Esto no es algo sorprendente. Lo sorprendente es cuanta universalidad se puede obtener siguiendo unas pocas pautas. Las áreas para las cuales estas pautas son desarrolladas son: accesibilidad, independencia de dispositivo e internacionalización.

El W3C desarrolla algunas de estas pautas. Al final, hay algunas reglas generales que son refinadas y complementadas:

- Utilizar estándares siempre que existan.
- Añadir redundancia en la forma en la que se presenta la información.
- Estructurar la información de forma correcta para que sea reconocible.
- Separar contenido de presentación.

Hay un solapamiento en los requerimientos técnicos para alcanzar esto. Una página que separa contenido de presentación puede ser adaptada fácilmente a diferentes dispositivos pero también a necesidades de accesibilidad. Una imagen de un edificio en un documento acompañada de texto puede ayudar a un invidente a entenderla pero también puede ayudar a alguien que por ejemplo nunca ha visto el edificio antes a identificarlo o a un

motor de búsqueda a encontrarlo. El texto llega a más gente cuando se añade información complementaria al proporcionarla en diferentes idiomas. La utilización de conjuntos de caracteres estandarizados ayuda a la traducción automática proporcionando un único modo de representación para símbolos y caracteres.

Pero las implicaciones van más allá de los temas mencionados aquí. Por ejemplo, un sitio web que es accesible es más fácil de encontrar. (O como Karsten M. Self, dijo en *Aracnofobia*: "*Google es, en todos los sentidos, un usuario ciego. Un usuario ciego millonario, con millones de amigos, que escucha cada una de sus palabras. Sospecho que Google tendrá un impacto mayor que ADA a la hora de crear sitios web accesibles*"). En general: la accesibilidad, la independencia de dispositivo y la internacionalización (como un efecto secundario bien recibido) también incrementan significativamente la usabilidad y el valor de un sitio.

4. Los pilares

Teniendo todas estas tecnologías y el conocimiento, ¿qué podríamos esperar de la Web?

Cada tecnología está ahí para hacer nuestra vida más fácil. Debería reducir el esfuerzo que tenemos que emplear para alcanzar nuestros objetivos. Así como la tecnología de la información en general, la Web ofrece un amplio espectro de áreas de aplicación.

Además de la tecnología de la información, la web ofrece estas aplicaciones enlazadas en una escala global. Las conexiones de banda ancha y las casi ilimitadas capacidades de almacenamiento y potencia, junto con tecnologías interoperables abren el campo para la multimedia y bases de datos de gran tamaño. Los modelos de negocio triunfan y fracasan, las denominadas "killer applications" emergen. Las redes sociales, la privacidad, la posibilidad de compartir conocimiento e inteligencia, la gestión de la propiedad intelectual, la participación en procesos de toma de decisiones, el marketing de ideas y productos, la importancia de la *long tail...*, son todos ellos aspectos que se vuelven más relevantes con la aparición de la Web.

Existe una convergencia entre la tecnología web y la tecnología de teléfonos móviles. Los teléfonos móviles digitales surgieron entre los usuarios durante la primera mitad de los años 90 al mismo tiempo que la Web. Desde entonces el crecimiento de este mercado sobrepasa incluso al crecimiento de Internet. Los teléfonos móviles juegan un papel crucial en la economía de muchos países en vías de desarrollo. En la actualidad, casi todos los teléfonos móviles pueden acceder a la Web. Por lo que ambas tecnologías convergen.

Por otro lado, hay ciertas cosas que aún son

algo confusas. Navegar puede ser en muchas ocasiones una auténtica pesadilla. El *phishing* y el abuso de las vulnerabilidades de los navegadores amenazan a los usuarios. Los desarrollos con respecto a la traducción automática son aún primitivos para ayudar a acceder a información en idiomas que el usuario no conoce. Hay mucho que mejorar e investigar en tecnologías, y hay muchas decisiones que tomar en tecnología, sociedad y política.

En las siguientes subsecciones mostraremos una clasificación de las tecnologías disponibles o que están siendo desarrolladas.

4.1. Aplicaciones Web

Para desarrollar una aplicación web lo único que se necesita es HTML. Por otro lado, hay muchas aplicaciones, por ejemplo, para el trabajo en la oficina o para el procesamiento de imágenes. También se puede incluir vídeo, audio, gráficos sofisticados, software para la representación de matemáticas o química. Los *plugins* y los intérpretes de lenguajes de programación soportan casi cualquier clase de procesamiento que funcione en un entorno de escritorio para ser utilizado en una página web. Tenemos bastante tecnologías preparadas para realizar esto. Algunas están estandarizadas, y otras son propietarias. Algunas son "abiertas" (por ejemplo, facilitando código abierto), otras son "cerradas".

Algunas tecnologías y estándares importantes en este contexto son desarrollados por el W3C (de forma gratuita y abiertos a todo el mundo). Entre estos estándares, HTML es el más utilizado. Es también el estándar que proporciona formas de incrustar objetos que siguen otros estándares para hacer gráficos, vídeos, imágenes...

Para lo bueno o para lo malo, hay dos HTML que en la actualidad están bajo estandarización por el W3C, conocidos como HTML 5 y XHTML 2, pero este no es el lugar para entrar en más detalles sobre este tema. Aun corriendo el riesgo de simplificar demasiado, estas dos tendencias se caracterizan de la siguiente forma:

- El principal objetivo de XHTML 2 es utilizar de forma más eficaz las ventajas de XML en HTML. Al mismo tiempo uno quisiera eliminar construcciones (elementos) cuya funcionalidad sería mejor implementar de una forma más limpia y general. Otro objetivo de diseño es permitir la integración de otras tecnologías o requisitos (documentos compuestos, Web semántica, accesibilidad).
- HTML 5 es más un avance de HTML (HTML 4 o XHTML 1). No se apoya necesariamente en XML aunque soporta su utilización, pero es muy importante su compatibilidad con versiones anteriores de HTML.

Por supuesto, hay solapamientos en los objetivos y principios de ambos HTML, pero las diferencias en los objetivos de diseño son importantes.

Para poder separar la presentación del contenido de documentos HTML o XML, se desarrolló una tecnología llamada CSS. Este lenguaje proporciona una forma flexible de definir para cada elemento la forma en la que se va a presentar en el documento. CSS es una tecnología muy potente que permite a los diseñadores Web adaptar sus páginas a una gran variedad de dispositivos móviles o tamaños de ventanas. Un enfoque más general es XSL. XSL permite no sólo darle formato a documentos ya existentes, sino que permite también transformar su contenido. De esta forma los elementos de un documento pueden ser reorganizados, compactados o extendidos.

Para permitir al lector de una página web enviar información a modo de respuesta (por ejemplo, enviar una orden de trabajo o inscribirse en un evento) los documentos HTML pueden contener formularios. Los formularios están disponibles como parte del estándar HTML. Mucho más potente (e integrado de forma más adecuada en el mundo de XML) es XFORMS que puede ser utilizado con cualquier lenguaje XML.

Un lenguaje XML bastante potente para los gráficos es SVG. Este lenguaje puede manejar gráficos de dos dimensiones vectoriales (además de imágenes basadas en *pixels*) que pueden ser escalables sin perder calidad. SVG permite también animaciones vectoriales. Muchos de los navegadores actuales son compatibles con SVG o subconjuntos de éste.

Como mencionamos anteriormente, (X)HTML permite incluir objetos que no son (X)HTML. A éstos se hace referencia o bien a través de un enlace o incluyéndolos directamente, frecuentemente como documentos XML con otro espacio de nombres (por ejemplo, SVG o el lenguaje de etiquetado de matemáticas, MathML). Esto debería funcionar de forma correcta a no ser que haya interacciones del usuario o sincronización de animaciones. Hay conjuntos de estándares de documentos compuestos que se están desarrollando (WICD, CDF...), los cuales permitirán mejorar la integración de partes de un documento que utilice diferentes tecnologías.

La Web (especialmente cuando se combina con lenguajes de programación potentes e interfaces) puede actuar como una plataforma para aplicaciones sofisticadas. Podemos ver un montón de *applets*, *widgetsy* aplicaciones que se basan en tecnologías web. Juegos, mapas, procesamiento de imagen o paquetes de oficina completos... pueden basarse en navegadores utilizando tecnologías como Formatos de Documento Compuesto, XML,

DOM (y eventos DOM), XFORMS, XHTML, SVG y algunos más.

Para poder crear presentaciones con multimedia desde diferentes fuentes de forma sincronizada hay que utilizar SMIL. SMIL fue publicado en 1998 y se trata de un lenguaje XML que permite organizar en tiempo y espacio elementos multimedia.

4.2. La Web Semántica

Encontrar cosas en la Web es todavía uno de los mayores retos. Hay muchos motores de búsqueda, de texto, de catálogos por temas o motores especializados como buscadores de precios más bajos, etc. No obstante, encontrar por ejemplo información sobre todos los eventos musicales en Berlín que tendrán lugar en abril el próximo año puede ser un reto si no se encuentra una base de datos especializada.

El motivo de esto es que los documentos web normalmente no dicen "esto trata de un evento que tendrá lugar en abril" de forma que los motores de búsqueda puedan entenderlo. Una persona puede, de forma sencilla, llegar a estas conclusiones cuando leen la información en la página pero no un motor de búsqueda. Encuentra la palabra "abril" pero no tiene ni idea de que esto pueda ser el mes en el cual un evento tenga lugar, así como tampoco sabe que esa página trata sobre un evento, ni tampoco qué puede ser la palabra "evento", y lo mismo es aplicable a la palabra "música".

Para poder solucionar estos problemas tenemos que "enseñar" a los ordenadores. Esto se lleva a cabo a través de la definición de relaciones entre las cosas de las que hablamos. Tenemos que (encontrar caminos para) decirle a los ordenadores que la música es un arte escénico. Que un arte escénico es un arte. Que cualquier cosa del tipo música tiene una duración. Que la música puede ser una representación en directo. Que una representación en directo es un evento. Que cada evento tiene una hora de inicio, etc. Segundo, necesitamos una forma de que los ordenadores sepan que ciertas partes de los datos son por ejemplo la hora de inicio. Y tenemos que ser capaces de especificar la hora de inicio que corresponde a cada evento.

Anteriormente mencionamos que RDF es un lenguaje que nos permite hablar de relaciones, pero para que los ordenadores lo puedan utilizar necesitamos también una forma única de nombrar a las *propiedades* y *recursos*. En este caso, el concepto de URL es muy útil. URL constituye una forma exclusiva de identificar documentos web. Pero no hay una razón técnica para constreñir el mecanismo de las URL a las páginas web. Al eliminar esta restricción lo que obtenemos es URI, el cual es utilizado por RDF para identificar de forma exclusiva *recursos* y *propiedades*.

Las *propiedades* son *recursos* también, por lo

que podemos hacer declaraciones sobre *propiedades*. Por ejemplo, podríamos definir que una *propiedad* "isSubClassOf" nos permite concluir que si "car isSubClassOf vehicles" y "myCar isMemberOf car" entonces "myCar isMemberOf vehicles". Esto nos permite llegar a nuevas declaraciones desde las ya existentes, aplicando la lógica. Hay diferentes vocabularios RDF (y reglas) que definen el significado de esas propiedades. Comenzando por el RDF más simple, se puede hablar (actualmente) de tres niveles en el estandarizado OWL: OWL Lite, OWL DL y OWL Full. Cada nivel ofrece diferentes opciones para especificar relaciones. Hay algoritmos que ayudan a encontrar estas conclusiones (pero en general no es seguro que esos algoritmos que procesan algunos OWL Full finalicen nunca).

¿Qué significa para la Web tener estas tecnologías? Podemos imaginar una Web "mejorada" donde por ejemplo, cada fecha de inicio de un evento se identifique a sí misma como fecha de inicio, haga referencia al evento al que pertenece, el cual se identifica a sí mismo como una representación y de igual manera se identifica como música... Hay un estándar llamado RDFa que permite incluir declaraciones RDF en un documento XHTML.

Aparte de las búsquedas, ha surgido en la Web otra utilización de los datos enlazados: los *mash-up*. A pesar de que actualmente están basados en API propietarios, la idea es utilizar los datos desde un recurso y unirlos con información de otro recurso. La utilización de tecnologías de web semántica para enlazar los datos haría mucho más fácil la construcción y mantenimiento de los *mash-ups*.

Otro ejemplo de utilización de datos enlazados es el proyecto FOAF. Este proporciona un vocabulario RDF para describir personas. Parte de la descripción puede estar compuesta por datos sobre a quién conoce esa persona - lo cual puede enlazar con otra descripción FOAF o contenerla. Algunas redes sociales abiertas están basadas en FOAF.

Aparte de RDF existe otra tecnología para enlazar datos en la Web: los microformatos. Los microformatos usan el atributo "class" de (X)HTML para declarar que un elemento HTML contiene una serie de datos. Con el objetivo de ser interoperables, los microformatos tienen que ser administrados de forma centralizada. Las especificaciones de microformatos están disponibles en el sitio web *microformats.org*.

4.3. Servicios web

Los servicios web (*Web Services*), lejos de la parte de navegación, están más relacionados con la interoperabilidad de aplicaciones software que se ejecutan en ordenadores interconectados por una red.

Un conjunto de estándares ayudan a alcanzar esta interoperabilidad y a conectar aplicaciones para llevar a cabo operaciones complejas. Juegan un papel central en los negocios de red (e-business). La oportunidad de transferir datos desde una etapa del negocio a las siguientes utilizando un formato de datos XML es algo muy atractivo y tiene la potencialidad de automatizar partes del proceso de negocio. Hemos visto un montón de conceptos y de palabras (EAI, ESB, MOM, SOA...) en esta área que están basados en servicios web. Hay que cumplir algunos requisitos para conseguir que los servicios web funcionen:

- Es necesario un protocolo para intercambiar mensajes entre aplicaciones.
- Es necesario ser capaces de direccionar servicios y mensajes.
- Es necesario describir servicios de forma que las máquinas los puedan comprender.
- Es necesario aplicar directivas para controlar la utilización de los servicios.

Hay unos cuantos estándares para servicios web basados en estos servicios:

- SOAP especifica el formato de los mensajes que se pueden utilizar para comunicarse con un servicio.
- WS-Addressing describe como direccionar servicios web y mensajes.
- WSDL es un lenguaje para describir un servicio por el tipo de mensaje que recibirá o enviará, y su funcionalidad. Hay también un estándar para anotaciones semánticas de WSDL (SAWSDL), el cual permite utilizar ontologías para representar la semántica de los servicios web.
- Un mecanismo para especificar directivas para la utilización de un servicio, que puede ser especificado y adjuntado a un servicio a través de la utilización de las especificaciones de Directivas de Servicios Web (*Web Services Policy specifications*).

4.4. Web móvil

El uso móvil de la Web se caracteriza por dispositivos con capacidades limitadas (por ejemplo teléfonos móviles) en términos de conectividad, banda ancha y poder de procesamiento del cliente. Los estándares de Web Móvil hacen posible la navegación y obtener una buena experiencia mientras se utilizan dispositivos móviles. Los estándares de Web Móvil se centran en varios aspectos:

- Reducir los estándares ya existentes para resolver las restricciones de dispositivos móviles.
- Desarrollar pautas especialmente para autores (pero también para fabricantes de herramientas de autor) sobre cómo hacer un sitio web que se vea en móviles.
- Encontrar formas para que los servidores sepan más sobre las capacidades de los clientes.

XHTML Básico, SVG Tiny, SVG Básico,

SMIL móvil, XFORMS Básico y CSS móvil son estándares que incluyen esos módulos de las tecnologías respectivas que los dispositivos móviles deberían proporcionar. Para poder acoplar el contenido enviado por un servidor a las capacidades de un dispositivo móvil (o de un navegador utilizado en el dispositivo) el servidor debe ser capaz de conocer las características del dispositivo. El API Simple para DDR describe una interfaz que los servidores pueden utilizar para buscar repositorios de descripción de dispositivos que contienen esta información.

Un segundo aspecto de la movilidad es que el conocimiento de la ubicación puede ser una ventaja para el usuario. Algunos dispositivos móviles tienen capacidad para ubicar su posición. Para poder utilizar esta información ofreciendo servicios al usuario se han comenzado trabajos sobre una especificación API de geolocalización.

4.5. Comunicación por voz

Nos podemos preguntar por los requisitos mínimos para acceder a la Web "¿funcionará un teléfono normal?", "¿sólo algunas veces?". Alguna información es distribuida o comunicada a través del teléfono, por lo que es razonable preguntarnos cómo las tecnologías web soportarán esto. Todo aquel que anteriormente ha utilizado un servicio de telefonía basado en el ordenador, sabe que, hablando coloquialmente, hay maneras de mejorar la experiencia de usuario. Para poder construir el camino hacia mejores comunicaciones por voz (semi-) automatizadas, se ha desarrollado un conjunto de estándares para cubrir las siguientes áreas:

- VoiceXML define infraestructuras para diálogos. Es un lenguaje de etiquetado para recoger información del usuario. El flujo del diálogo es controlado por el etiquetado VoiceXML dependiendo de las respuestas del usuario. VoiceXML es un lenguaje XML con características de *scripting*.
- Dos estándares para el habla soportan la utilización de VoiceXML. SSML es un lenguaje XML para ayudar en la producción de habla sintética. SRGS permite la especificación de gramática para el reconocimiento del habla.

Por otro lado, CCXML proporciona soporte al control de llamadas telefónicas y puede ser utilizado de forma conjunta con sistemas de diálogo como VoiceXML. EMMA es un lenguaje XML que sirve para describir la entrada de datos desde diferentes canales de forma simultánea (voz, escritura, teclado...).

4.6. Privacidad y seguridad

La Web ha cambiado nuestras vidas en muchos sentidos. Uno de los cambios más ambivalentes es nuestra privacidad. Mientras usamos la Web estamos mostrando un mon-

tón de datos personales, de intereses y preferencias. Cada pregunta que hacemos, cada sitio web que visitamos, puede ser registrado. Es difícil para el usuario adivinar cómo se va a utilizar esta información. Los gobiernos, para poder proteger a sus ciudadanos han creado leyes de privacidad pero el ámbito de la legislación normalmente suele terminar en las fronteras nacionales, lo que limita sus efectos en Internet. Este problema no puede ser resuelto únicamente de forma técnica. La gente tiene que manejar sus datos privados con cuidado, y han de encontrarse nuevos modelos de privacidad socialmente aceptados y técnicamente implementables.

P3P es un estándar para especificar una descripción procesable por máquinas sobre directivas de privacidad. Permite, por ejemplo, a las empresas publicar sus directivas de privacidad de forma que un agente de usuario (por ejemplo, como parte de un navegador web) pueda entenderlas. Ya que no hay forma de hacer respetar la directiva, P3P es útil en los casos en los que el usuario puede confiar en un sitio que muestra una directiva. Esto es lo que normalmente se asume cuando se hacen negocios. Algunos navegadores (por ejemplo, Internet Explorer 6+) utilizan información P3P proporcionada por los servidores para permitir a los usuarios tener el control del uso de *cookies*.

Confiamos en algunos, pero ¿podemos estar seguros de que el sitio web que pretende ser la página principal de un banco, es realmente un banco?, ¿podemos transferir datos confidenciales de forma segura en la Web?, ¿puede un banco asegurar que el dinero transferido desde la cuenta de un cliente fue realmente enviado por él?

Estos problemas no están especificados en la Web. Las matemáticas (especialmente la teoría del número) nos han proporcionado un instrumento de encriptación que resuelve muchos de los problemas en esta área. Los protocolos de Internet con TLS/SSL aplican estos instrumentos a los protocolos de Internet. Evidentemente hay todavía cosas que tienen que resolverse para aplicar esta tecnología a nivel de documento o dato. XML Encryption, XML Signature, y XML Key Management son tres tecnologías relacionadas que permiten aplicar conocidos métodos criptográficos en fragmentos de documentos XML, por ejemplo para firmar digitalmente una parte del cuerpo de un mensaje SOAP.

Se está desarrollando trabajo sobre cómo evaluar o informar a los usuarios sobre el contexto de seguridad de un entorno que están utilizando. Este trabajo gira alrededor de un mejor uso de los certificados, encriptación, sitios web dinámicos e informaciones proporcionadas por el servidor. Esto debería permitir a los usuarios comprender mejor el contexto de seguridad de un sitio web a la hora de tomar

decisiones basadas en la confianza.

A pesar de que hay bastante progreso en esta área, existen todavía bastantes problemas por resolver en relación con la privacidad y seguridad.

5. El tejado: Sitios web y aplicaciones revolucionarias

Todas estas tecnologías y estándares son más un medio que un fin en sí mismos y la lista todavía está lejos de ser completada.

El crecimiento conjunto y continuo de la banda ancha y de la potencia de los ordenadores ha hecho de la Web un lugar donde la gente utiliza su creatividad para hacer negocios, llevar a cabo discusiones, distribuir ideas, o establecer relaciones.

Durante los últimos 17 años la Web ha transformado Internet pasando de ser un extraño juguete a una infraestructura muy potente formada por millones de individuos, gente de negocios, periodistas, artistas, científicos... Google, Yahoo, Flickr, Ebay, Youtube, Myspace, Wikipedia, Mozilla, son sólo algunos de los proyectos más populares dentro de este espacio.

Pero hay todavía un montón de preguntas abiertas que están esperando ideas, soluciones técnicas, o acuerdos sociales para obtener respuesta. De forma paralela al trabajo realizado sobre millones de sitios web, bitácoras, wikis, tiendas y aplicaciones de moda, hay un trabajo en marcha para obtener mejores tecnologías y estándares. El W3C proporciona una infraestructura donde proveedores, autores y usuarios cooperan para mejorar el entorno de la Web.

Glosario

ADA: Americans with Disabilities Act.
API: Application Programming Interface.
CCXML: (Voice Browser) Call Control XML.
CDF: Compound Document Formats.
CGI: Common Gateway Interface.
CSS: Cascading Style Sheets.
DDR: Device Description Repositories.
DOM: Document Object Model.
EAI: Enterprise Application Integration.
EMMA: Extensible MultiModal Annotation markup language.
ESB: Enterprise Service Bus.
FOAF: Friend of a Friend.
HTML: Hypertext Markup Language.
HTTP: Hypertext Transfer Protocol.
MathML: Mathematical Markup Language.
MOM: Message Oriented Middleware.
OWL: Web Ontology Language.
P3P: Platform for Privacy Preferences.
RDF: Resource Description Framework.
RDF/XML: RDF in XML Syntax.
RDFS: RDF Schema.
RDFa: RDF in attributes.
SAWSDL: Semantic Annotations for WSDL.
SMIL: Synchronized Multimedia Integration Language.
SOA: Service-Oriented Architecture.
SOAP: (originalmente) Simple Object Access Protocol (la versión larga es confusa y está en desuso).
SPARQL: Protocol and RDF Query Language.
SRGS: Speech Recognition Grammar Specification.
SSML: Speech Synthesis Markup Language.
SVG: Scalable Vector Graphics.
TLS/SSL: Transport Layer Security / Secure Sockets Layer.
URI: Uniform Resource Identifier.
URL: Uniform Resource Locator.
VoiceXML: Voice Extensible Markup Language.
W3C: World Wide Web Consortium.
WICD: Web Integration Compound Document.
WSDL: Web Service Description Language.
XFORMS: (Un nuevo formato XML para poder definir interfaces de usuario).
XHTML: Extensible Hypertext Markup Language.
XML: Extensible Markup Language.



Pez de Babel
servicios lingüísticos

Soluciones integrales a necesidades de

globalización

internacionalización

www.pezdebabel.com

Info@pezdebabel.com T.985 308 911 F.985 308 228

Charles McCathieNevile
 Director de Estándares de Opera Software

<chaals@opera.com>

Uso de la Web en móviles: algunos detalles y tendencias de 2008

1. Estado de la Web Móvil

1.1. Introducción

Durante toda esta década se ha estado hablando de la *Web móvil* como "el próximo gran hito". A finales de los 90 se tenía la frecuente sensación de que la única forma de utilizar un móvil era por medio de pasarelas diseñadas especialmente, y accediendo a sitios web hechos para este fin. Entre otras cosas, esto llevó a enormes inversiones en tecnología WAP (*Wireless Application Protocol*) [1] por parte de un enorme consorcio de compañías [2], y en paralelo al desarrollo de cHTML (*Compact Hypertext Markup Language*) [3] y del servicio iMode lanzado en primer lugar por el operador japonés NTT DoCoMo.

Algunos sistemas pioneros de transcodificación dieron acceso a sitios webs generalistas convirtiéndolos a WML, pero la mayor parte de la inversión se dirigió a desarrollar lo que se llamaron "jardines vallados" (*walled gardens*), portales basados en WML (*Wireless Markup Language*) y sitios generalmente proporcionados por operadores que controlaban los enlaces que se podían seguir y en algunos casos navegadores restringidos para llevar a los usuarios a su propio portal (ver **figura 1**).

Durante la última década las mejoras en terminales telefónicos y en ancho de banda han supuesto una evolución desde aquellos primeros días del sencillo WAP y de los contenidos y navegadores iMode. A lo largo de este tiempo el lenguaje WML ha evolucionado hasta el punto de que su versión actual,

Traducción: José Alfonso Accino (Grupo de Trabajo Lengua e Informática de ATI)

Resumen: este texto hace una breve revisión de las tendencias en el uso de la Web en dispositivos móviles, estudiando específicamente datos referentes al navegador Opera Mini. Intenta encontrar resultados imprevistos y diferencias geográficas, e identifica algunas tendencias y temas para un estudio posterior del uso de la Web móvil.

Palabras clave: Opera Mini, sistemas de transcodificación, sitios para móviles, Web, Web móvil.

Autor

Charles McCathieNevile es Director de Estándares en Opera Software, una compañía noruega cuya visión es proporcionar la mejor experiencia en Internet para el usuario de cualquier dispositivo. Charles está a cargo de los estándares y especificaciones de Opera. Sus intereses se centran en la accesibilidad web para todos y en facilitar todo lo posible la creación para la Web, así como en las herramientas de autoría funcionales para gente corriente. Es vicepresidente de la "Fundación SIDAR – Acceso Universal" una sociedad que se propone conseguir que la Sociedad de la Información, en toda Iberoamérica, sea accesible e inclusiva.

XHTML-MP 1.2 [4] es la misma que el XHTML 1.1 Basic [5] del W3C, y en la práctica la mayoría de los navegadores de los teléfonos móviles, más que un manejo adecuado de XHTML (*eXtensible HyperText Markup Language*) de cualquier tipo, aceptan ya algo que se aproxima al contenido de "sopa de etiquetas" (*tag-soup*) que se espera de la Web "standard" [6].

Opera fue probablemente la primera en poner un navegador estándar para "Web total" en un teléfono móvil, hace 6 años, utilizando el navegador para móvil que habían desarrollado para los dispositivos de tipo Palmtop que se empezaron a comercializar en 1998. Desde entonces, los navegadores "Web total" para

móviles se han convertido en navegadores modernos capacitados para Ajax (*Asynchronous JavaScript And XML*) como el Opera Mobile para s60 (2005), seguidos por el navegador de los Nokia, los iPhone de Apple y los navegadores Opera para dispositivos Windows Mobile de gama alta tales como los Touch Diamond de HTC, Omnia de Samsung y Xperia de SonyEricsson.

Aunque se han vendido ya cientos de millones de dispositivos, tanto con navegadores WAP como "Web total", los usuarios han sido generalmente más lentos en hacer uso de las posibilidades que ofrecen sus teléfonos para la navegación web. Los últimos tres años han visto cambiar esta situación, con navegadores para



Figura 1. Web de la BBC en Opera Mini (vista general y detalle) y página WML de Google en 2006.

móviles que empiezan a mostrarse como serios contendientes en el mercado web en general [7].

Este estudio analiza los datos de Opera Mini, uno de los navegadores para móviles más populares hoy en día, para determinar si hay diferencias significativas entre lo que hace la gente cuando navega desde un móvil y desde un ordenador de mesa. Aunque es de esperar alguna diferencia entre el uso de Opera Mini y el de otros dispositivos móviles, los datos disponibles constituyen un interesante objeto de estudio. Queda para otra investigación saber si se dan tendencias comunes para los distintos dispositivos.

1.2. Metodología

Desde abril de 2008 Opera Software ha publicado un informe mensual titulado "Informe del Estado de la Web Móvil" (*State of the Mobile Web Report*) [8]. En realidad, este informe se deriva del uso de Opera Mini, uno de los navegadores web para móviles de mayor crecimiento (ver **figura 2**) con cerca de 20 millones de usuarios únicos al mes a finales de 2008 (comparados con los cincuenta mil de enero de 2006, el mes de su lanzamiento).

Opera Mini funciona mediante un *proxy* de transcodificación, lo que permite a un cliente muy ligero (alrededor de 100Kb, como un Midlet² Java) entregar páginas web incluso en dispositivos con una potencia relativamente baja. Aunque no tiene un soporte completo para Ajax consigue dar una experiencia de usuario significativa para una gran proporción de páginas web. El *proxy* recoge algunos datos estadísticos básicos que se utilizan para generar el informe (hay información más detallada disponible sobre la Política de Privacidad de Opera Mini³).

Naturalmente, esto no es una muestra totalmente representativa. Opera Mini está diseñada

especialmente para teléfonos con una característica determinada, lo que es tanto como decir dispositivos de capacidad media o baja, más que para los terminales de gama alta como los Touch Diamond de HTC, Xperia de SonyEricsson y Omnia de Samsung, todos los cuales utilizan Opera Mobile (el moderno navegador de gama alta con capacidad Ajax), o teléfonos avanzados tales como el Nokia N95 con su navegador S60 de Nokia, o los iPhone de Apple que ejecutan versiones de Safari para móviles.

Opera Mini ha estado disponible para pruebas en algunos países desde finales de 2005, distribuyéndose ya en general desde principios de 2006 y el autor ha podido consultar datos de uso más detallados proporcionados por Opera Software. Aunque no es posible divulgar todos esos datos en relación con este artículo, parece que en gran medida repiten los ya proporcionados por el Informe del Estado de la Web Móvil.

1.3. Tendencias de alto nivel

Mirando todos los informes conjuntamente, hay algunas tendencias evidentes. Lo más obvio es que los países con mayor presencia son relativamente estables. Desde el primer Informe del Estado de la Web Móvil, Rusia ha sido siempre el primer país en cuanto al número de usuarios finales. Indonesia ha sido siempre el país número dos en cuanto al número de usuarios únicos. Es más, en ocho meses sólo once países han estado entre los diez primeros por número de usuarios, con Egipto desplazando temporalmente a Alemania a finales de 2008. Sin embargo, ha habido movimientos, cambiándose las posiciones entre los puestos 3 al 11.

De manera similar al número de usuarios, las páginas vistas y el volumen de datos consumido han experimentado un crecimiento bá-

sicamente estable. Un resultado interesante es que ha crecido el número de páginas vistas por usuario, lo que significa que, al parecer, los usuarios hacen más uso de Opera Mini conforme pasa el tiempo. Con un crecimiento en usuarios de alrededor del 300% anual, esto significa que deberíamos esperar ver cierta madurez en los patrones de uso.

2. Tipos populares de sitios

2.1. Redes sociales

La búsqueda de datos y las redes sociales son los dos usos más populares de Opera Mini, con gran diferencia. Vkontakte.ru, una red social en ruso, es generalmente el sitio más popular, con otros sitios de redes sociales tales como Facebook, Friendster, Yahoo, Orkut y My.Opera siguiéndole de cerca en la lista (con algunas variaciones según el país; parece que muchos países tienen sólo uno o unos pocos sitios favoritos, y éstos recibieron un mayor tratamiento en el informe de diciembre [10]).

2.2. Motores de búsqueda

Los motores de búsqueda como Google, Yahoo, Yandex y Baidu también están consistentemente en los primeros lugares. Generalmente son de la versión "Web total" más que de sitios específicos para móviles, aunque algunos de los sitios adaptan automáticamente su respuesta dependiendo del cliente. Sin embargo, los resultados de búsqueda son para la Web en su conjunto más que para la Web móvil.

Cuando el buscador Yahoo! era el fijado por defecto en Opera Mini, proporcionaba resultados para la "Web total" y para la "Web móvil"; por ejemplo, sitios que identificaba como específicamente adaptados para dispositivos móviles. Desafortunadamente, no hay datos disponibles sobre qué resultados eran más populares pero sería un tema de estudio

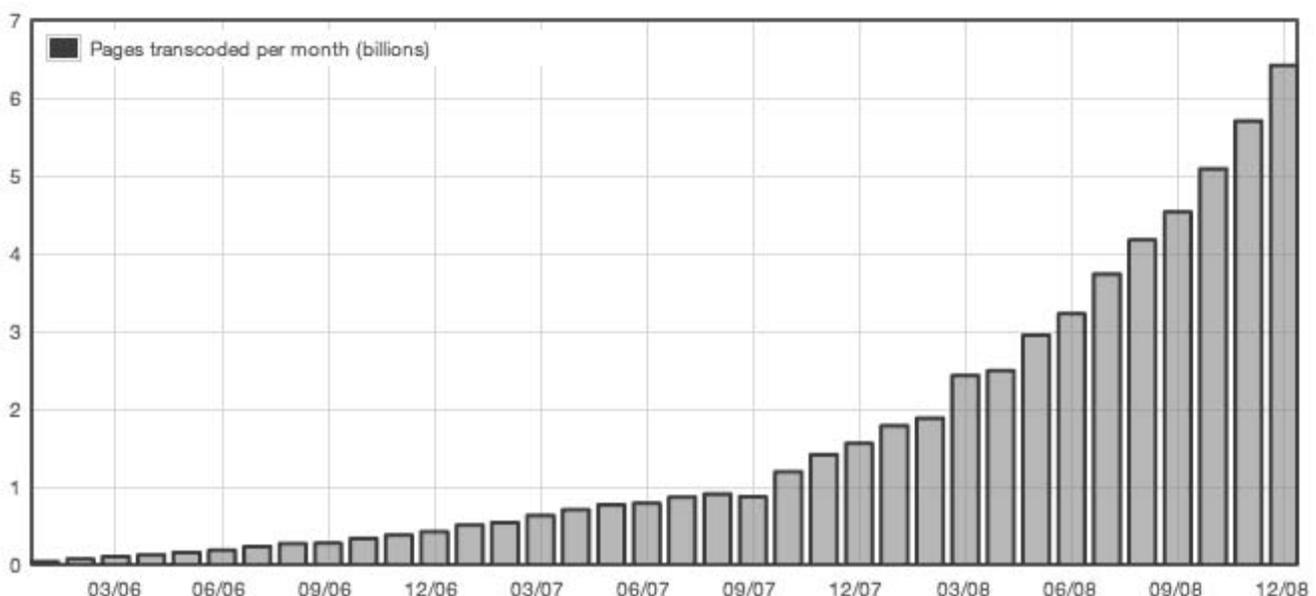


Figura 2. Páginas vistas al mes en Opera Mini [9].

muy interesante.

2.3. ¿Sitios específicos para móviles o no?

Un resultado que puede sorprender es el uso de sitios orientados a móviles comparado con sitios diseñados simplemente para la web en general. Esto es bastante más difícil de estudiar ya que hay varias formas de proporcionarle a un determinado dispositivo contenido específico para móviles. Sin embargo, es evidente que muchos de los sitios que los usuarios están mirando no están orientados a móviles.

Algunos sitios que hacen un fuerte uso de Ajax, multimedia y otras tecnologías que un simple teléfono no ofrece fácilmente, muestran una clara preferencia por la versión móvil. Facebook y YouTube entran en esta categoría.

Esta información no resulta necesariamente aparente en los informes del Estado de la Web Móvil, que pone juntos sitios diferentes para tales servicios, pero puede verse en estadísticas más detalladas estudiadas por el autor. Esto ocurre, en parte, porque algunos sitios reconocerán al navegador utilizado y ajustarán dinámicamente el contenido devuelto. De este modo, dividir las cuentas teniendo en cuenta si se ha accedido a un dominio específico para móviles no daría como resultado unas estadísticas más ajustadas.

Sin embargo, en el conjunto de datos disponibles para el autor hay un cierto número de sitios que muestran una preferencia por la versión "total" del sitio, incluso cuando se dispone de una versión para móviles. Parece que, siempre que es posible, los usuarios prefieren la funcionalidad que les resulta familiar en sus equipos de sobremesa en vez de optar por defecto, simplemente, por las versiones para móvil.

Otra categoría de "sitios para móviles" son aquellos que ofrecen productos o servicios específicamente diseñados para teléfonos móviles. Hay algunos de estos entre los diez primeros de varias listas, ofreciendo juegos para teléfonos móviles y cosas parecidas.

Curiosamente, los sitios que utilizan el dominio .mobi de alto nivel solo muy raramente aparecen entre los diez primeros, siendo mucho más comunes los indicadores como "wap" en el nombre de dominio, aunque algunos de tales sitios simplemente utilizan un nombre de dominio que no identifica el sitio como dirigido a móviles. En el primer informe, los sitios identificados como orientados a móviles eran responsables, en general, de casi una cuarta parte del volumen de datos transferido pero no se dispone de nuevos análisis de esta cantidad y no queda totalmente claro cómo se obtuvo esa cifra.

2.4. Contenido para adultos

Sí, en la web hay disponible contenido para

adultos. Posiblemente mucho, aunque este estudio no haga ningún intento para medirlo. Sorprendentemente, tal vez, es también popular en los navegadores para teléfonos móviles. Es raro que los contenidos para adultos alcancen los diez primeros sitios que aparecerían listados en el informe, pero la política de Opera en realidad es eliminarlos del mismo si eso llegara a ocurrir.

Menos sorprendentemente, los sitios de citas y contactos son también populares, más en general (por ejemplo, tanto en la lista global como por países, hasta el punto que aparecen ocasionalmente en las listas de los diez primeros sitios) que los sitios que ofrecen solamente contenido erótico. Esto coincide bien con la popularidad de los servicios de redes sociales. Dado que algunos de estos sitios están pobremente adaptados para navegadores móviles y en cambio están diseñados específicamente para sistemas de sobremesa, esto refuerza la idea de que el contenido es de capital importancia para los usuarios a la hora de escoger el tipo de sitios que visitan con sus móviles.

2.5. Diversidad y consistencia

Así como las tendencias mencionadas anteriormente se mantienen en los distintos países, hay algunos aspectos que están más localizados. Aunque los datos que interesan ya han sido publicados en los informes del Estado de la Web Móvil, no se ha puesto todo de relieve en esos informes.

2.5.1. Comunidades locales

En general, vemos que los sitios más populares son grandes marcas multinacionales como Facebook, Friendster, Yahoo y Yandex. Algunos servicios de correo y mensajería son también muy populares en unos pocos países, y noticias y deportes son temas comunes a un cierto número de países. Estos servicios están generalmente dominados por algunos grandes nombres. Por ejemplo, la BBC es particularmente popular en muchos países, tanto de habla inglesa como árabe.

Sin embargo hay unos pocos casos en los que vemos sitios locales que son excepcionalmente populares. En países tales como Eslovaquia [11], Polonia [12], Rusia [13] o China [14] hay una mayoría de dominios propios de ese mismo país incluso en la lista de los diez primeros. Ocurre muy frecuentemente que los países que prefieren contenidos locales son, principalmente, países del antiguo "Bloque Comunista" en Europa, o países que aún están gobernados por partidos comunistas, como Vietnam y China. Este estudio es muy superficial como para extraer conclusiones particulares de estos datos, pero parece ser una vía interesante a seguir en posteriores investigaciones.

2.5.2. Comercio electrónico

En general, los sitios de ventas bien conocidos

como Amazon y eBay no suelen aparecer entre los diez primeros (aunque sí lo hacen en los datos más completos que le han sido proporcionados a este autor). La excepción principal es Alemania, donde eBay ha aparecido con continuidad entre los diez primeros, generalmente alrededor del número 5.

Sin embargo hay un par de sitios de otros lugares que han tenido un impacto local. En el número de octubre de "Focus on South East Asia", las cifras muestran [15] que en Brunei Dar-es-salaam el sitio Bruneiclassified.com aparece en el quinto puesto, y de vez en cuando el sitio web de Ferrocarriles de la India, www.indianrail.gov.in, aparece entre los diez primeros en la India [16] (globalmente está entre los primeros 500 para los usuarios de Opera Mini). Más local todavía, capetown.gumtree.co.za, un sitio dedicado a la venta de artículos en una ciudad de Sudáfrica y sus alrededores, ha aparecido varias veces [17] entre los diez primeros de ese país. De nuevo, este estudio deja las razones de estas anomalías (o el por qué la gente está más interesada en los resultados de los deportes que en los precios de las acciones) como un tema para investigación posterior.

2.5.3. La "larga cola"

Aunque hay unas pocas docenas de sitios que aparecen repetidamente en las listas de diez primeros en muchos países, una cuestión importante para los desarrolladores es si su sitio será utilizado por gente con navegadores para móviles, y cómo. Casi a priori, la mayoría de los sitios no son los más populares, de modo que la pregunta para los desarrolladores es cómo de variada es la gama de sitios utilizados, en general.

Una rápida instantánea de la "larga cola"⁴ (Long Tail) [18] del informe de agosto muestra que entre los diez países del top ten solo hay dos en los que los 100 sitios más visitados representen más del 60% de las páginas vistas (ver figura 3). En Rusia, con el mayor número de usuarios de Opera Mini, los 100 sitios más visitados sólo suponen un 44% del tráfico, pero en cambio Indonesia, el segundo país, es uno de esos en los que los diez primeros sitios controlan una cantidad mayor (76% - la misma proporción que China).

En ese mes, globalmente, cerca de 15 millones de personas visitaron alrededor de 12 millones de sitios. Aunque 87 consumieron ellos solos más de la mitad del tráfico, y en Indonesia Friendster por sí solo acaparó el 50% del tráfico (lo que significa que los otros 99 sitios del top 100 contabilizaron menos de una cuarta parte), esto sugiere que la "larga cola" de la curva es, efectivamente, muy larga. Si se observan los gráficos, cuatro de cada cinco usuarios (*de media*) han escogido visitar un sitio que ningún otro usuario ha visto todavía en ese mes. Por tanto, la variedad de sitios

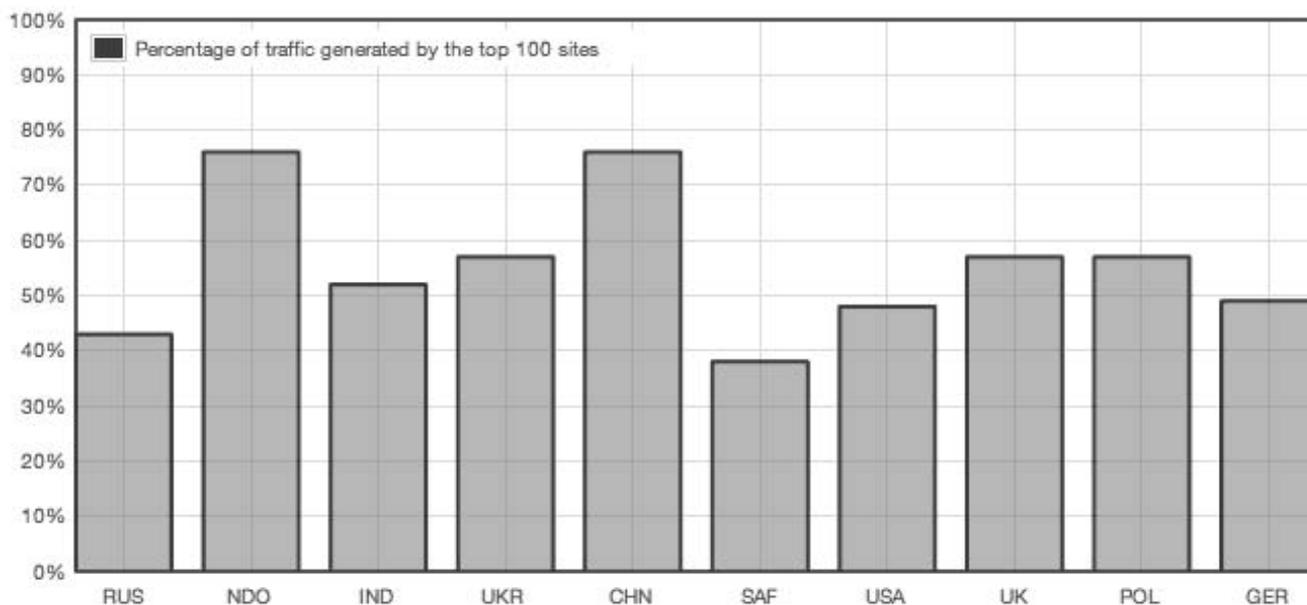


Figura 3. La "larga cola" – Proporción de vistas de páginas generada por los sitios del top 100, para cada uno de los países del Top 10 [18].

visitados es similar en amplitud a la variedad de usuarios. Debería indicarse, sin embargo, que parece haber cierta variación regional en la popularidad de los sitios más conocidos, por ejemplo, en la distribución precisa de la curva de "larga cola" de país a país.

3. Clientes (no) comunes

Es evidente que los usuarios de Opera Mini son en gran medida un reflejo del mercado de teléfonos. Así vemos que los teléfonos más conocidos aparecen como los dispositivos más populares en varias regiones, con Nokia como marca dominante globalmente, marcas chinas también populares en algunas regiones, Blackberry extendida en los EE.UU. y así sucesivamente. Estos resultados se ajustan a las expectativas más simples, pero hay un par de anomalías curiosas.

En el informe de septiembre se destacaba a África. Muchos de los resultados no constituyen ninguna sorpresa. Como en el resto del mundo dominan las búsquedas y las redes sociales, con noticias y entretenimiento (principalmente BBC en países de habla inglesa y árabe, con el periódico deportivo L'Équipe superando a Le Monde en el más francófono Costa de Marfil). Las grandes economías como Sudáfrica, Egipto y Nigeria son los países líderes en términos de usuarios en general, con Libia y otros mostrando un enorme crecimiento (de varios miles por ciento, en comparación a los cientos por cien de los países en cabeza de la lista).

Los dispositivos utilizados en los países de mayor nivel reflejan economías relativamente modernas con terminales actuales y, como para el resto del mundo, en muchos casos esto incluye dispositivos que tienen un navegador web moderno, razonablemente avanzado y capacitado para Ajax.

Pero hay una extraña anomalía en las cifras para Zambia [19] y Swazilandia [20]. Los dispositivos para esos países son generalmente los que cabría esperar, pero entre ellos (en el número 3 en Zambia y en el número 4 en Swazilandia) está el "equipo de sobremesa". Aparentemente, en esos países la gente está utilizando el *applet* de demostración (disponible para probar Mini) como primer navegador. Estos son países donde el número de páginas vistas sobrepasan en crecimiento al número de usuarios en un factor cercano a 3, por lo que a primera vista parece que no se trata simplemente de que mucha gente esté probándolo en una plataforma antes de usarlo.

De hecho, este *applet* solo necesita ejecutarse en un entorno Java Micro Edition. Aunque la demostración que ha estado disponible durante un par de años [21] está explícitamente diseñada para ejecutarse desde un navegador, desde comienzos de 2008, como mínimo, [22] se han venido publicando instrucciones acerca de cómo instalar el *applet* y ejecutarlo directamente desde un ordenador de sobremesa. En la práctica no es algo terriblemente complicado y parece que ha habido casos repetidos e independientes entre sí de gente describiendo este proceso y ofreciendo consejos. Los motivos para utilizar Opera Mini [23] con preferencia a un navegador para ordenador de sobremesa son, generalmente, economizar ancho de banda, a menudo para gente que utiliza conexiones por línea conmutada o que usa sus móviles como módem.

Las preguntas que la gente hace en los foros como consecuencia de este uso muestran que están utilizándolo en una amplia variedad de circunstancias, con explicaciones que se hacen públicas por todo el globo, desde Madagascar a Rusia (ver figura 4).

Con la reciente proliferación de "netbooks" (dispositivos baratos mucho más pequeños que un portátil estándar pero mayores incluso que los teléfonos de alta gama disponibles, y a menudo con potencia bastante baja para un notebook), es de esperar que este uso se haga cada vez más común, especialmente en países donde el ancho de banda todavía se cobra como consumible en vez de a precio fijo.

4. ¿Marcadores?

Por último, Opera Mini permite trasladar los marcadores (*bookmarks*) al dispositivo. Este parece ser un método bastante exitoso para generar tráfico. Mientras que el propio sitio de red social de Opera, My.Opera.com, estaba clasificado en el puesto 700 por Alexa.com a lo largo de la segunda mitad de 2008, para los usuarios de Opera Mini estaba clasificado entre los 100 primeros, y en muchos países aparecía incluso entre los 10 primeros.

Puede ser que los usuarios de Opera Mini sean más proclives a usar productos de Opera, en general (con un enlace Opera expresamente diseñado para dar soporte al movimiento de datos entre distintas versiones de Opera para un único usuario, esto es ciertamente una meta de Opera), pero de los datos proporcionados al autor se desprende que parte de este efecto se debe, más probablemente, a su presencia como uno de los marcadores por defecto en la página inicial de Opera Mini. Sin embargo, hace falta una investigación más detallada para establecer esto mejor y para determinar en qué medida vale para otros sitios que tienen esta presencia en su página frontal.

5. Tendencias en equipos de sobremesa

Alexa [24] es una referencia habitual para los sitios web más populares en general, y aunque



Figura 4. La web de la BBC. Captura de pantalla de Opera Mini ejecutándose en un ordenador de sobremesa a tamaño completo.

el uso de equipos que no son de escritorio está creciendo rápidamente, es todavía seguro asumir que el tráfico general representa en gran medida accesos efectuados desde equipos de sobremesa.

Las búsquedas y las redes sociales son claramente las aplicaciones clave en todas partes. Entre los diez primeros se encuentran siempre sitios similares, aunque hay algunas diferencias notables. Por ejemplo, como ya se ha dicho, My.Opera es bastante más popular entre los usuarios de Opera Mini de lo que lo es en general. Hay también algunos sitios específicos para móviles que puntúan notablemente más alto en Opera Mini de lo que lo

hacen en general (gamejump.com en Estados Unidos y wapos.ru en Rusia son ejemplos de esta clase), y los sitios de noticias generalmente puntúan igual de alto o más en Opera Mini de lo que lo hacen en la web en general.

Por otro lado, como se ha indicado anteriormente, los sitios de comercio electrónico no son tan populares, generalmente, entre usuarios de Opera Mini como lo son globalmente, y donde lo son, los sitios muy localizados son probablemente más populares que los generalistas. La **tabla 1** muestra que eBay es más popular en Estados Unidos que Craigslist, pero los datos a los que hemos

accedido para este artículo muestran que Craigslist es bastante más popular que eBay entre los usuarios estadounidenses de Opera Mini, con algunos subdominios de Craigslist en un par de grandes ciudades cada uno de los cuales maneja casi tanto tráfico como eBay.

Mirando más datos aparecen otras diferencias notables. Los sitios dedicados puramente a compartir fotos son bastante menos populares entre usuarios de Opera Mini que entre el resto, un efecto que es mucho más pronunciado en general que la popularidad relativa de los sitios para compartir vídeo. De manera similar, los sitios que se orientan primordialmente a los blogs y a la creación de contenido son menos populares entre los usuarios de Opera Mini, aunque los sitios como Wikipedia no muestran este efecto.

6. Conclusiones

Aunque hay algunas diferencias entre cómo se usan los teléfonos móviles y los equipos de escritorio para acceder a la Web, parece que los usuarios de Opera Mini están haciendo en sus teléfonos muchas cosas similares a las que hacen en un ordenador de sobremesa, especialmente en lo que se refiere al entretenimiento personal y a mantenerse en contacto.

Su uso para negocios parece que es menos evidente, sin embargo lo clasificamos, tal vez manteniendo el concepto de teléfono móvil, principalmente como un dispositivo personal para la comunicación.

Hay un claro deseo por parte de mucha gente de acceder a la web, por distintas razones y en distintas formas. La longitud de la larga cola nos muestra que aunque unos pocos "megasitios" son inmensamente populares, los datos de la parte baja de la escala de popularidad muestran que en general los desarrolladores web deben probablemente

	Rusia		EE.UU.	
Clasificación	Opera Mini [25]	Alexa [26]	Opera Mini [27]	Alexa [28]
1	vKontakte	vKontakte	Google	Google
2	Mail.ru	Yandex	Facebook	Yahoo!
3	Yandex	Mail.ru	Friendster	MySpace
4	Odnoklassiki	Odnoklassiki	MySpace	YouTube
5	Rambler	Rambler	Wikipedia	Facebook
6	Google	Google	Gamejump (Alexa: ~200,000 para EE.UU.)	Windows Live
7	My.Opera (Alexa: ~700 para Rusia)	Narod.ru	Yahoo!	MSN
8	Wikipedia (Alexa: 11 para Rusia)	Youtube	My.Opera (Alexa: ~1300 for EE.UU.)	eBay
9	Wapos (Alexa: ~20,000 para Rusia)	LiveJournal	YouTube	Wikipedia
10	Seclub (Alexa: ~7000 para Rusia)	Ucoz	NYTimes (Alexa: 25 para EE.UU.)	Craigslist (Opera Mini: top 50)

Tabla 1. Comparación de clasificaciones para Rusia y EE.UU.: Opera Mini y Alexa.

esperar que un número creciente de usuarios de una gama amplia de sistemas accedan a sus contenidos.

Saber si esto sugiere que deberían invertir en optimización específica para móviles (que algunas veces es bastante menos popular que el acceso al sitio "principal") o en un paradigma de diseño que permita más fácilmente adaptarse al cliente es una pregunta que una mirada superficial como ésta no puede responder, sino que por el contrario surge de ella.

7. Temas pendientes de estudio

Un artículo de investigación serio investigaría con mayor detalle los distintos hallazgos de este breve texto. En particular, las áreas que parecen más interesantes son el uso de navegación con móviles en sectores específicos, y la prevalencia en general de los sitios adaptados para su uso con móviles (y la preferencia o no de los usuarios por ellos).

El uso de navegadores móviles, en general, es un campo que se podría beneficiar de esa investigación y ofrecer una respuesta a toda una gama de preguntas acerca del uso de clientes específicos, de distintas clases de dispositivos (¿por qué la gente usa Opera Mini en el Nokia N95 en lugar del navegador incorporado? ¿Cuál es la tendencia de crecimiento para su uso en dispositivos de sobremesa?, etc.), podría ser muy valioso para la comunidad de desarrollo Web. La intención de este texto es poner de manifiesto algunas vías a seguir, con la esperanza de que nos conduzcan a descubrimientos útiles.

Referencias

- [1] WAP <<http://www.openmobilealliance.org/tech/affiliates/LicenseAgreement.asp?DocName=/wap/technical%5B1%5D.zip>>.
- [2] La Open Mobile Alliance, anteriormente conocida como Forum WAP, no publica una lista de miembros pero en <http://en.wikipedia.org/wiki/Open_Mobile_Alliance> se indican algunos de ellos bien conocidos.
- [3] Compact HTML for Small Information Appliances, <<http://www.w3.org/TR/1998/NOTE-compactHTML-19980209>>.
- [4] <<http://www.openmobilealliance.org/tech/DTD/xhtml-mobile12.dtd>>.
- [5] XHTML™ Basic 1.1, <<http://www.w3.org/TR/2008/REC-xhtml-basic-20080729/>>.
- [6] <<http://simon.html5.org/articles/mobile-results>>. Aunque es una investigación parcial muy incompleta, sus resultados pueden generalizarse con cierta fiabilidad.
- [7] Las estadísticas de navegadores son notoriamente poco fiables. Pero comparando <<http://www.liveinternet.ru/stat/ru/browsers.html?date=2008-12-31;period=month>> y <<http://www.liveinternet.ru/stat/ru/browsers.html?date=2007-12-31;period=month>> se muestra que el uso de Opera Mini ha crecido muy significativamente como una fracción del uso de la web general. (Esas cifras están sesgadas hacia contenidos en ruso, haciendo a Opera Mini bastante más visible que las estadísticas recogidas primariamente de contenidos en idioma inglés).
- [8] <<http://opera.com/smw/>>.
- [9] <<http://www.opera.com/smw/2008/12/>>.
- [10] <<http://www.opera.com/smw/2008/12/#global2>>.
- [11] <<http://www.opera.com/smw/2008/11/#slovakia2>>.
- [12] <<http://www.opera.com/smw/2008/11/#poland2>>.
- [13] <<http://www.opera.com/smw/2008/11/#russia>>.
- [14] <<http://www.opera.com/smw/2008/11/#china>>.
- [15] <<http://www.opera.com/smw/2008/10/#brunei>>.
- [16] Por ejemplo, véase <<http://www.opera.com/smw/2008/11/#india>> y <<http://www.opera.com/smw/2008/04/#india>>.
- [17] Por ejemplo, <<http://www.opera.com/smw/2008/08/#southAfrica>> y <<http://www.opera.com/smw/2008/07/#southAfrica>>.
- [18] <http://www.opera.com/smw/2008/08/#chart_longtail>.

- [19] <<http://www.opera.com/smw/2008/09/#zambia>>.
- [20] <<http://www.opera.com/smw/2008/09/#swaziland>>.
- [21] <<http://www.opera.com/mini/demo>>.
- [22] <<http://java4me.blogspot.com/2008/01/opera-mini-as-pc-browser-big-screen.html>>.
- [23] Véase también <<http://my.opera.com/ariespntn/blog/show.dml/2735233>> o <<http://lawand.wordpress.com/2009/01/12/opera-mini-for-windows/>> o <<http://helpforlinux.blogspot.com/2008/12/use-opera-mini-in-ubuntu.html>> y <<http://my.opera.com/chaals/blog/2009/01/22/opera-mini-on-mac-os>> para algunas instrucciones y discusiones adicionales.
- [24] <<http://www.alexa.com>>.
- [25] <<http://www.opera.com/smw/2008/12/#russia>>.
- [26] <http://www.alexa.com/site/ds/top_sites?cc=RU&ts_mode=country&lang=none>.
- [27] <<http://www.opera.com/smw/2008/12/#US>>.
- [28] <http://www.alexa.com/site/ds/top_sites?cc=US&ts_mode=country&lang=none>.

Notas

¹ Utilizo los términos "Web total" y "Web móvil" para distinguir los sitios web "de todos los días" diseñados sin tomar particularmente en consideración que los usuarios puedan estar accediendo desde un móvil del subconjunto de tales sitios que están específicamente orientados al usuario de teléfonos móviles.

² Midlet es una programación en lenguaje de programación Java para dispositivos embebidos (se dedican a una sola actividad), más específicamente para la máquina virtual Java MicroEdition (Java ME). Generalmente son juegos y aplicaciones que corren en un teléfono móvil. Está desarrollada bajo la especificación MIDP (perfil para información de dispositivo móvil) <<http://en.wikipedia.org/wiki/MIDlet>> [25-3-2009].

³ <<http://www.opera.com/privacy/#operamini>>.

⁴ La larga estela o larga cola (en el original en inglés *The Long Tail*) fue una expresión acuñada por Chris Anderson en un artículo de la revista Wired de Octubre de 2004 para describir determinados tipos de negocios y modelos económicos tales como Amazon.com o Netflix... En estas distribuciones una amplia frecuencia o gran frecuencia de población es seguida por una baja frecuencia o baja amplitud de la población que disminuye gradualmente. En muchos casos, los acontecimientos de baja frecuencia o escasa amplitud -la larga cola- pueden abarcar la mayor parte del gráfico [25-3-2009].

¿Estudiante de Ingeniería Técnica o Ingeniería Superior de Informática?

Puedes aprovecharte de las condiciones especiales para hacerte

socio estudiante de ATI

y gozar de los servicios que te ofrece nuestra asociación,

según el acuerdo firmado con la

Asociación RITSI

Infórmate en <www.ati.es>

o ponte en contacto con la Secretaría de ATI Madrid
secretmdr@ati.es, teléfono 91 402 93 91



Stephane Boyera
World Wide Web Foundation

<boyera@w3.org>

La utilización de tecnologías web y teléfonos móviles para el desarrollo social: W3C

1. Introducción

El W3C (*World Wide Web Consortium* [1]), desde su creación en 1994 por el inventor de la Web Sir Tim Berners-Lee, ha estado trabajando para hacer realidad su visión de alcanzar un acceso universal a la Web: desde cualquier lugar, para todo el mundo, en cualquier momento y sobre cualquier cosa. Durante este tiempo la Web ha crecido de forma exponencial hasta alcanzar los 1.500 millones de usuarios en el 2008 (fuente: Internet World Stats [2]), creando servicios, proporcionando información, conectando personas, creando nuevos trabajos y sectores de actividad.

A pesar de este enorme éxito en tan poco tiempo, hay todavía más de 5.000 millones de personas que no están obteniendo ningún beneficio de la Sociedad de la Información creada por la Web. No obstante, la Web, y las tecnologías informáticas en general, han sido reconocidas como herramientas que potencialmente pueden resolver la brecha histórica entre economías desarrolladas y no desarrolladas a través de una infraestructura para proporcionar servicios básicos (salud, educación, negocios, gobierno...) a comunidades rurales y poblaciones en situación de marginación.

Este es el motivo por el que se han puesto en marcha algunas acciones específicas en los últimos 20 años para tratar de resolver la denominada brecha digital. Desafortunadamente, algunas de estas acciones (frecuentemente centradas en la creación de "telecentros") han tenido poco éxito hasta el momento. Por ejemplo, el "modelo telecentro" se ha encontrado con algunas dificultades como consecuencia de las condiciones en las zonas más desfavorecidas, tales como la falta de electricidad, la falta de conocimientos sobre mantenimiento... y se han llevado a cabo muy pocos esfuerzos en esta área que se hayan mantenido en el tiempo y hayan continuado funcionando.

Desde hace 2 ó 3 años, están apareciendo nuevas oportunidades como consecuencia de la alta penetración de la telefonía móvil en países en desarrollo. Actualmente, una infraestructura básica (redes GSM) y dispositivos con capacidad inferior a la del ordenador (teléfonos móviles), están disponibles para millones de usuarios en sus bolsillos (o por lo menos en su entorno cercano), incluyendo el segmento más pobre de la población. Muchos de los países en desarrollo que no han podido sumarse a la revolución de la telefonía por falta de infraestructura e inversiones, han participado de forma directa en la revolución de los móviles.

Traducción: Encarna Quesada Ruiz (editora invitada de la monografía)

Resumen: este artículo presenta la nueva Iniciativa de Web Móvil para el Desarrollo Social del Consorcio World Wide Web (*Mobile Web for Social Development, W3C*) lanzada en mayo del 2008. Este nuevo grupo investiga cómo usar las tecnologías informáticas en teléfonos móviles como solución a la brecha digital y mecanismo que proporcione servicios básicos (salud, educación, gobierno, negocios...) a comunidades rurales y colectivos marginados en países en desarrollo. En la primera parte se presentan los motivos del lanzamiento del grupo, y en la segunda parte el punto de vista y la dirección que se está siguiendo actualmente, una agenda futura y una lista de entregables.

Palabras clave: brecha digital, comunidades rurales, desarrollo social, grupo de interés de web móvil para el desarrollo social, Iniciativa de Web Móvil, Mobile Web Initiative, países en desarrollo, Web móvil, W3C.

Autores

Stephane Boyera trabaja para W3C (*World Wide Web Consortium*) desde 1995. Desde el año 2001 ha dirigido el Grupo de Trabajo de Independencia de Dispositivo del W3C, habiendo sido un participante crucial en el desarrollo y lanzamiento de la Iniciativa de Web Móvil del W3C. Coordinó también hasta el año 2005 el Grupo de Trabajo de Descripción de Dispositivos. De forma simultánea, Stephane tomó también parte en la coordinación de las actividades multimodal y de voz. Desde el año 2006, ha sido el coordinador del trabajo de W3C relacionado con la Web Móvil para el desarrollo social, actividad que se centra en cómo ampliar las fronteras de la Web para hacerla más usable y útil en comunidades rurales y colectivos marginados en países en desarrollo. Desde su lanzamiento en enero del 2008, Stephane ha coordinado el proyecto EUPF7 del Digital Word Forum, centrado en el uso de las tecnologías informáticas como elemento de desarrollo económico en África y Latinoamérica. Desde mediados del 2008, Stephane participa en el lanzamiento de la Fundación World Wide Web, siendo el encargado de definir la Web para el programa social. Antes de formar parte del equipo del W3C Stephane trabajó en Inteligencia Artificial (adquisición de conocimiento y modelado) en INRIA (*Institut National de Recherche en Informatique et en Automatique*, Francia). Es licenciado en ingeniería por la ESSTIN (Escuela de Ingeniería en Telecomunicaciones y Redes en Sophia-Antipolis).

¿Puede repetirse esto en la Web? Hace 4 años el W3C, junto con la industria de las tecnologías móviles, lanzó la Iniciativa de Web Móvil (*Mobile Web Initiative* [10]) cuyo objetivo era convertir a los usuarios de teléfonos móviles en ciudadanos web de primera clase. Desde entonces, el número de personas que acceden a la Web desde dispositivos móviles está creciendo rápidamente. ¿Sería posible que, al igual que para la telefonía, algunos países en desarrollo se salten la revolución PC-Web y vayan directamente a la siguiente fase, la Web Móvil?

En mayo del 2008 el W3C puso en marcha el Grupo de Interés de Web Móvil para el Desarrollo Social (MW4D, *Mobile Web for Social Development* [9]) con el objetivo de investigar esta tendencia. Tanto su visión como objetivos, entregables y resultados esperados serán presentados en este artículo¹.

2. Grupo de Interés de Web Móvil para el Desarrollo Social

La misión de este grupo (MW4D), parte de la Iniciativa de Web Móvil, es explorar la poten-

cialidad de las tecnologías web en teléfonos móviles como una posible solución a la brecha digital, y proporcionar servicios informáticos a las comunidades rurales y poblaciones marginadas en países en desarrollo.

Los cuatro objetivos principales de este grupo de interés son:

1. Identificar retos: comprender los problemas de acceso a las tecnologías informáticas en comunidades rurales y poblaciones marginadas, conocer sus necesidades documentales, retos y casos de uso.
2. Definición del mapa de ruta: identificar para cada reto soluciones posibles.
3. Participación de la comunidad: crear un foro interdisciplinario para discutir cómo las tecnologías de teléfonos móviles pueden ayudar a reducir la brecha digital.
4. Contribuir al acceso universal de la Web: ayudar a que el W3C alcance su objetivo de hacer la Web accesible a todo el mundo.

El grupo tiene también el objetivo de identificar la forma en la que la comunidad W3C puede



"Hay todavía más de 5.000 millones de personas que no están obteniendo ningún beneficio de la Sociedad de la Información"



contribuir a reducir la brecha digital y a ayudar en la consecución de los Objetivos del Desarrollo del Milenio de la ONU. No obstante, entre los objetivos de este grupo están:

- Ofrecer información a otros grupos de trabajo de W3C sobre los retos concretos ocasionados por la brecha digital.
- Proponer nuevas áreas de trabajo a la comunidad del W3C (por ej. a través de la propuesta de nuevos grupos de trabajo).
- Comprender y proponer a los grupos que trabajan dentro del W3C formas de contribuir.

3. Web móvil para el desarrollo social

Sabemos que actualmente la mitad de la población mundial vive con menos de 2,5 dólares al día [3]. Este porcentaje de población carece de todo tipo de servicios (salud, gobierno, educación, finanzas...) lo cual impide que aumenten sus ingresos.

En los últimos años, la potencialidad de los servicios informáticos para proporcionar soluciones a estos problemas ha sido demostrada de forma clara. Por ejemplo, en la India, en el sector de la pesca, en concreto en Kerala, un estudio económico [4] ha demostrado que la adopción de un servicio móvil de información sobre envíos (necesidades y precios) desde diferentes áreas geográficas, permitiendo a los pescadores proporcionar mercancía a los mercados más apropiados, ha incrementado los ingresos de esos pescadores en un 9% (incluyendo el precio del teléfono móvil/suscripción), mientras que el precio total del pescado descendió en un 4% para los consumidores como consecuencia de la menor necesidad de eliminación de mercancía no vendida.

Experiencias similares y resultados parecidos se han visto en otras regiones y productos (ver por ejemplo, el mercado del cereal en Nigeria [5]).

Desafortunadamente, mientras estos experimentos muestran resultados impresionantes, el número de este tipo de servicios a nivel global es todavía muy bajo, y los sectores cubiertos son principalmente agricultura y banca, existiendo en cambio áreas con similar potencialidad en salud, educación, gobierno, etc.

En el Libro Blanco sobre Web Móvil para el Desarrollo Social [6], explicamos por qué la tecnología actualmente utilizada (SMS) es

claramente un factor limitante, dificultando el desarrollo, distribución y uso a gran escala de numerosas aplicaciones. A pesar de que hay muchas razones por las que SMS se utiliza actualmente de forma generalizada (disponibilidad en todos los teléfonos, coste conocido, facilidad de uso para los usuarios, recepción gratuita...), esta tecnología tiene limitaciones intrínsecas (conocimiento requerido, falta de localización, falta de mecanismos automáticos de descubrimiento, falta de estandarización, limitación a 160 caracteres...) lo cual impide el desarrollo y la distribución de servicios de bajo coste a gran escala.

Por otro lado, las tecnologías web han demostrado su potencialidad en estas áreas en concreto y, por lo tanto, la puesta en marcha de la nueva generación de aplicaciones móviles basadas en estas tecnologías (navegador móvil, tecnologías de voz, *widgets* para móviles...) podría ser una solución.

El objetivo del grupo de trabajo MW4D es investigar y comprender las características de un entorno que permita la adopción de esta nueva generación de aplicaciones. Esto dará como resultado la aparición de numerosos servicios que influirán de forma positiva en la vida del segmento más pobre de la población en los países en desarrollo.

Los proyectos con éxito normalmente siguen tres pasos:

1. "Alguien" que observa cómo trabaja y vive una comunidad, identifica la potencialidad de posibles servicios informáticos para ayudar a esa comunidad.
2. La idea es entonces implementada utilizando la tecnología.
3. Los usuarios objeto encuentran el servicio accesible, asequible, usable y útil, y hacen un esfuerzo por aprender y adoptar ese servicio.

En relación al primer paso, hay diferentes actores sociales:

- ONG/organizaciones de base/organizaciones sin ánimo de lucro.
- Gobierno/Administración pública.
- Emprendedores sociales individuales investigando cómo hacer negocio a base de ofrecer servicios sociales.

Cada uno de estos actores sociales tiene retos

y objetivos específicos, y para cada uno es necesaria una metodología específica. Como primer paso, MW4D se centra en ONG/organizaciones de base/organizaciones sin ánimo de lucro (ver el diagrama del MW4D en la referencia [7]).

En relación al segundo paso, MW4D está llevando a cabo investigaciones para identificar los problemas, barreras, necesidades y retos para proveedores potenciales de servicios orientados al desarrollo a la hora de ofrecer y distribuir estos servicios, así como a la hora de explotar la potencialidad de las plataformas móviles.

Como hemos mencionado previamente, en la actualidad hay múltiples formas (tecnologías) para desarrollar y distribuir contenidos en teléfonos móviles. Cada una de estas tecnologías tiene dominios de aplicación, requisitos (en los dispositivos de mano, en el operador...) y costes específicos. MW4D está analizando e identificando las diferentes dimensiones a tener en cuenta para hacer una selección apropiada basada en las especificaciones de la aplicación, el contexto, el usuario final, etc. El grupo está también investigando los vacíos que existen actualmente y las direcciones más adecuadas a seguir para resolverlos. Esto incluye las necesidades educativas específicas que ayudarían a las personas a enriquecerse permitiéndoles desarrollar aplicaciones que necesitan sin requerir experiencia externa.

En relación al tercer paso, es importante comprender los retos y barreras principales que los usuarios finales (población marginada en países en desarrollo) afrontan a la hora de acceder y utilizar servicios móviles. Esto incluye problemas de asequibilidad, usabilidad, analfabetismo, accesibilidad e internacionalización/localización. MW4D está trabajando en la identificación de estos retos y en formas de trabajar o caminos prometedores para encontrar el modo de afrontarlos en el futuro.

En términos de objetivos finales, el grupo de trabajo MW4D seguirá en activo hasta finales de mayo del 2009. Para entonces, se espera que el grupo haya desarrollado:

1. Un manual para cualquiera que desee desarrollar y distribuir servicios móviles orientados al desarrollo social. El manual describirá las tecnologías actuales que pueden usarse

con este propósito, su campo de aplicabilidad, requisitos, retos y aspectos específicos a tener en cuenta.

2. Un mapa de ruta que identifique las principales alternativas a explorar o las acciones más interesantes a poner en marcha con el objetivo de reducir las barreras a la hora de proporcionar o acceder a contenido y servicios en teléfonos móviles.

3. Un directorio de recursos relacionados con el uso de teléfonos móviles en desarrollo social [8].

En enero de 2009, se han creado algunos borradores de forma colaborativa [11]. MW4D, a diferencia de la mayoría de los grupos de W3C, está abierto a la participación pública.

Referencias

- [1] **World Wide Web Consortium (W3C)**. <<http://www.w3.org/>>.
- [2] **Internet World Stats**. *Internet Usage Statistics: World Internet Users and population stats*, junio 2008. <<http://www.internetworldstats.com/stats.htm>>.
- [3] **Anup Shah**. *Poverty Facts and Stats*. septiembre 2008. <<http://www.globalissues.org/article/26/poverty-facts-and-stats>>.
- [4] **R. Jensen**. The Digital Provide: Information (technology), Market, Performance, and Welfare in the south Indian Fisheries Sector. *The Quarterly Journal of Economics*, Vol. CXXII, agosto 2007 Número 3. <<http://www.mitpressjournals.org/doi/pdf/10.1162/qjec.122.3.879>>.
- [5] **J.C Aker**. *Does Digital Divide or Provide? The Impact of Mobile Phones on Grain Markets in Niger*. University de California, Berkeley, enero 2008, <http://www.cgdev.org/doc/events/2.12.08/Aker_Job_Market_Paper_15jan08_2.pdf>.
- [6] **Stephane Boyera**. *White paper on Mobile Web for Social Development*. Enero 2008, <http://www.w3.org/2006/12/digital_divide/ajc>.
- [7] **Ken Banks**. *Key Focus Areas for the W3C Mobile Web for Social Development Group*. Agosto 2008,

<http://lists.w3.org/Archives/Public/public-mw4d/2008Aug/att-0060/W3C_Mobile_Web_for_Social_Development_Group_-_Focus_Areas.pdf>

[8] **MW4D**. *Mobile Web For Social Development (MW4D) Wiki*. <<http://www.w3.org/2008/MW4D/wiki/>>.

[9] **MW4D**. Página principal del *Mobile Web for Social Development (MW4D)*.

<<http://www.w3.org/2008/MW4D/>>.

[10] **W3C**. *Mobile Web Initiative*. <<http://www.w3.org/Mobile/>>.

[11] **MW4D**. Borradores de Web móvil para el desarrollo social del W3C. <<http://www.w3.org/2008/MW4D/wiki/MW4D>>.

[12] **Digital World Forum**. Página principal del *Digital World Forum on Accessible and Inclusive ICT*. <<http://www.digitalworldforum.eu/>>.

Nota

¹ El trabajo presentado en este artículo, aunque está principalmente dirigido por W3C, es parte del proyecto EU FP7 del Digital Word Forum sobre accesibilidad e inclusión en las TIC (para obtener mas información consultar la referencia [12]).

XV JORNADAS DE ENSEÑANZA UNIVERSITARIA DE LA INFORMÁTICA

JENUI 2009

Barcelona, del 8 al 10 de julio 2009



El objetivo de estas Jornadas, promovidas por la Asociación de Enseñantes Universitarios de Informática (AENUI) y organizado por la Facultat d'Informàtica de Barcelona de la Universitat Politècnica de Catalunya, es favorecer el contacto e intercambio de experiencias docentes entre profesores universitarios de la informática, debatir sobre el contenido y los métodos pedagógicos empleados, y presentar temas y enfoques innovadores que permitan mejorar la docencia de la informática en las universidades españolas.

PLAZO DE PRESENTACIÓN DE RESÚMENES

Hasta el 6 de febrero de 2009

PLAZO DEFINITIVO DE PRESENTACIÓN DE TRABAJOS

Hasta el 13 de febrero de 2009

<http://jenui2009.fib.upc.edu/>

Organizado por:



UNIVERSITAT POLITÈCNICA DE CATALUNYA



AENUI

Asociación de Enseñantes Universitarios de la Informática

Bert Bos

W3C; inventor de las hojas de estilo en cascada (CSS)

<bert@w3.org>

1. El papel de los estándares de W3C en el desarrollo de la Web

La Web es un sistema de comunicación: permite a la gente intercambiar información. Sus características se basan en otros sistemas de comunicación, pero la combinación de ellos es lo que hace a la Web única. En particular, la Web muestra cómo es posible utilizar ordenadores para mejorar la comunicación. Y esto se debe a que la Web nos ha dado ideas de cómo convertir la semántica (es decir el significado) en sintaxis. W3C (*World Wide Web Consortium*) ha señalado la expresión de significado en lenguajes formales como el mecanismo principal para hacer la Web más útil.

Al igual que los libros, la Web permite a un autor llegar a muchas personas, que no necesariamente tienen que estar leyendo al mismo tiempo. Pero publicar en la Web es más barato y rápido, a lo que hay que añadir el hecho de que la información está en un formato electrónico, lo que permite una adaptación sencilla a diferentes entornos (si la codificación de la información sigue determinadas pautas). Por supuesto, existen entornos donde un libro es más fácil de manejar. Por otro lado, puede ser difícil conseguir el libro, mientras que es más fácil acceder a algún tipo de red para conectarse a la Web.

En el diagrama de la **figura 1** podemos ver una representación de cómo un autor (arriba) tiene una idea que desea comunicar. Este autor representa su idea en formato digital. Utiliza texto, imágenes, video, etc., y probablemente necesite comprobar y modificar su trabajo varias veces antes de que esté satisfecho con la presentación.

Esta representación digital de la idea es transportada a varios lugares, donde puede ser simplemente almacenada por algún tiempo o incorporada a otra información.

Un lector (abajo) visualiza la información digital en formato no digital e intenta comprenderla. Puede probar diferentes modos de mostrar la información: cambiar el tamaño de la fuente, imprimirla en papel, extraer sólo las partes interesantes, etc.

Quizá el lector quiera entonces convertirse en autor o incorporar sus propias ideas en la Web.

Al igual que en un correo electrónico y en las listas de correo, la Web permite interactuar con otras personas en un tiempo razonablemente corto. No tan corto como una comunicación cara a cara o por teléfono, pero sin

Convirtiendo la semántica en sintaxis

Traducción: Encarna Quesada Ruiz (editora invitada de la monografía)

Resumen: la Web es un sistema de comunicación con una combinación de características que permite el procesamiento automático, de tal forma que un lector no lea simplemente lo que narra el autor, sino una versión mejorada, la cual combine recursos y muestre información adaptada a éste. Para W3C, el potencial de la Web radica en representar la información a través de lenguajes formales que un ordenador pueda manipular. Ya existen lenguajes o están siendo desarrollados para aspectos de la comunicación tan diversos como los colores, los sonidos, las ubicaciones geográficas o las licencias de copyright. No se han alcanzado aún los límites, pero existen: cuanto mayor es la necesidad de una expresión precisa del significado, menor es el número de personas que son capaces de proporcionarlo. En algún momento, el ordenador tiene que enfrentarse con lo que queda implícito, en otras palabras, el complemento de la Web Semántica es la inteligencia artificial. Hay algunos aspectos prácticos que influyen también sobre los lenguajes que puede ser estandarizados. En concreto se trata de la escasez de recursos.

Palabras clave: CSS, EOT, estándares, gobierno electrónico, grupo de trabajo, HTML, ontologías, Web Semántica, W3C.

Autor

Bert Bos estaba realizando una investigación sobre interfaces de usuario gráficas en la Universidad de Groningen cuando se convirtió en co-inventor (junto con Håkon Wium Lie) de las hojas de estilo en cascada (CSS, *Cascading Style Sheets*) en 1994. En 1995, se unió al W3C para desarrollar CSS y dirigir la actividad de internacionalización del W3C. Desde entonces ha participado en la estandarización de HTML, XML y recientemente MathML. Es editor de varios (borradores) estándares relacionados con CSS y MathML, y coautor del libro "*Cascading Style Sheets – designing for the Web*" (Addison-Wesley Professional, 2005).

la presión del tiempo: la información espera al lector, es decir no se pierde cuando el que ha de escuchar no está disponible en el momento en el que el interlocutor habla.

La interacción no es tan simple como en un e-mail, o por lo menos de momento, porque

es necesario acceso de escritura a un servidor web, tanto si es propiedad del lector como si es de otra persona (por ejemplo, a través de *wiki*). Por otro lado, es más fácil hacer referencia a mensajes anteriores ya que todos tienen una dirección única en un servidor (las listas de correo también pueden tener un

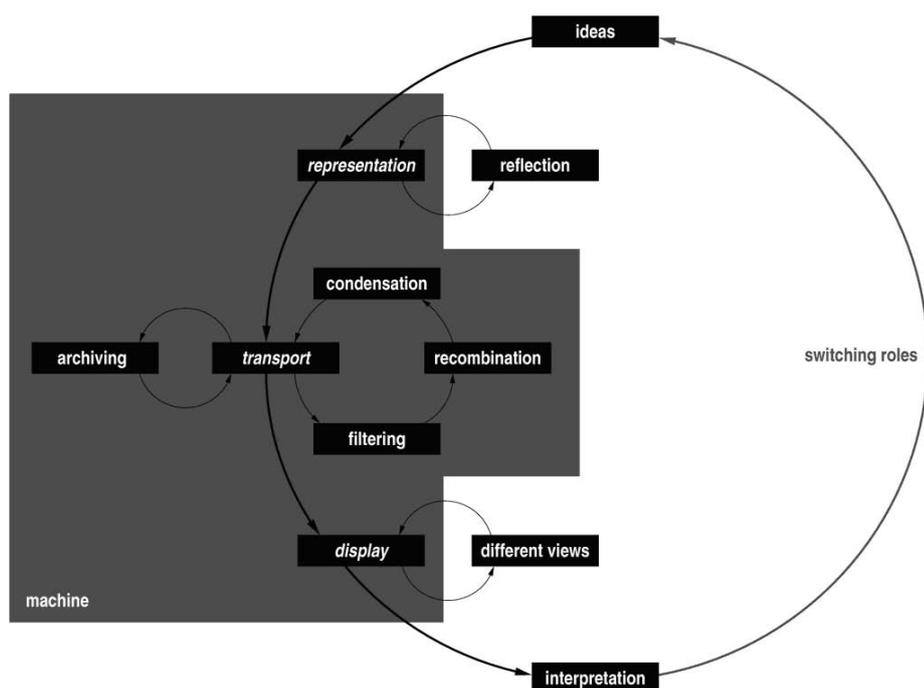


Figura 1. Diagrama semántico de flujo de información en la Web.



"La incorporación de información adicional requiere esfuerzo. No todo el mundo es capaz de hacerlo y es necesario un aprendizaje"



archivo web con este propósito).

La estabilidad de la información no puede garantizarse en la Web. La estabilidad radica en el mantenimiento de los datos que escriben los usuarios (por supuesto sin tener en cuenta accidentes como errores en los ordenadores, pero que en la práctica es difícil que causen la pérdida de la información).

La tecnología sobre la que se construye principalmente la Web (DNS, URL, diferentes protocolos y formatos), asume que la información permanece estable durante cierto tiempo, por lo que tiene sentido hacer referencias, pero la estabilidad no está realmente incorporada. La idea original de Ted Nelson para Xanadu [1] tuvo en cuenta la estabilidad, en el sentido de que versiones anteriores de información en ubicaciones concretas permanecían disponibles. El sistema entre iguales (*peer-to-peer*) Freenet [2] también tiene una noción de estabilidad: la información permanece siendo la misma, bit a bit, durante su ciclo de vida, y su ciclo de vida no depende del que publica, sino de alguien, en algún lugar, que esté interesado en esa información.

En la Web, la estabilidad es todavía un problema sin resolver. Por un lado, no queremos que información importante desaparezca cuando el que la publica no puede hacerlo por más tiempo desde su ordenador (Internet Archive [3] se ocupa de mantener en lo posible en sus ordenadores información en línea que no existe en otro lugar), pero por otro lado, también queremos que se olvide información especialmente privada.

Por el momento, la posición del W3C sobre la estabilidad es que se trata de un problema social (ver "*Cool URIs don't change*" de Tim Berners-Lee [4]). Mientras que la información se identifique a través de la ubicación (URL), en lugar de identificadores más indirectos y abstractos (URN), no hay mucho que podamos hacer.

Y de momento no sabemos cómo hacer un sistema lo suficientemente rápido y escalable basado en URN. Los libros, por supuesto, utilizan URN: se hacen referencia unos a otros a través del autor, el título y el año y algunas veces a través del ISBN. Para traducir esa identificación en una ubicación (edificio, sala, estantería), es necesario tiempo. El ya mencionado sistema Freenet utiliza URN, en

un sistema en el que la información se mueve donde es necesaria, por lo que no tiene una ubicación fija.

A pesar de las investigaciones y experimentos, esto todavía es muy lento, a no ser que la información ya esté cerca del lugar ("cerca" en términos de red). A diferencia de la Web, donde la referencia es igual a la ubicación y por lo tanto la información llega a través del enlace más directo, es decir en tres segundos para una página web simple, o unos segundos más para una más compleja... a no ser que la información haya desaparecido de esa ubicación, por supuesto.

La Web, cuando se usa de forma interactiva se presenta como un sistema de hipertexto. A diferencia de las referencias en libros, las cuales requieren cierto esfuerzo, algunas veces bastante esfuerzo para ser enlazadas, las referencias en los documentos de la Web no necesitan más que un simple clic para descargarse en segundos. Por supuesto, sólo si la referencia es a algo en la Web, pero la popularidad y el tamaño de la Web es tan grande que hay muchas posibilidades de que la información que se ha enlazado esté en línea. La interacción a través del hipertexto no es única en la Web, pero la combinación del hipertexto con esa gran y variada cantidad de información sí lo es.

La Web es un sistema modular en el sentido de que permite que diferentes modalidades de información (texto, sonido, imágenes, video...) sean transportadas de formas diversas (HTTP, FTP, peer-to-peer, con cable y sin cable...) y los diferentes componentes del sistema puedan ser mejorados o reemplazados individualmente sin necesidad de cambiar los otros componentes.

HTTP (*HyperText Transfer Protocol*), el protocolo de transporte más utilizado en la Web, no es irremplazable. Tiene muchas características útiles (negociación de contenido, *caching*, carencia de estado, autenticación opcional, etc.), pero otros protocolos entre iguales tienen otras ventajas también.

Debido a que las redes físicas se están haciendo cada vez más rápidas (con fibra de vidrio, por ejemplo), el tamaño de la información en bytes se está convirtiendo en un problema menor, pero los protocolos entre iguales pueden ayudar aún más acortando la distancia entre la información y la persona que la

solicita. Después de todo, con independencia de la capacidad de la banda ancha, la velocidad de la luz es finita y cada petición y respuesta enviada a través del Océano requiere un tercio de un segundo, mientras que enviarla a través de una ciudad, requiere menos de una décima. Si una información está compuesta por muchas piezas pequeñas, esos retrasos empiezan a acumularse.

Parece ser que velocidad y estabilidad están relacionadas. Los protocolos entre iguales que usan técnicas de *caching* pueden ayudar a aumentar la velocidad de la Web sólo para información estable. Las cookies no son de gran ayuda en este sentido, ni tampoco lo son los identificadores de sesión, pero la tecnología conocida como AJAX (*Asynchronous JavaScript And XML*) es probablemente uno de los peores culpables (a pesar de que AJAX está centrada en la usabilidad tanto para los usuarios como para los agentes inteligentes, es probablemente un problema mayor).

El texto tiene un lugar especial entre las diversas modalidades, porque es una forma muy eficaz de representar el lenguaje humano y porque puede ser codificado en forma digital. Sabemos cómo desarrollar programas para texto, mucho mejor que para otras formas de información.

Sin embargo, tanto para texto como para cualquier otra cosa, necesitamos mejorar la información con anotaciones que los ordenadores utilicen, no para otros seres humanos. Ello es debido a que nuestros programas (o programadores) no son lo suficientemente listos todavía para reconocer muchos de los mensajes implícitos en nuestra comunicación. Necesitamos indicar palabras clave para los motores de búsqueda, señalar texto con marcadores para los encabezados, listas, citas y párrafos. Los espacios, las comas y los puntos a los que estamos acostumbrados, no son suficientes.

Esta información extra es lo que normalmente se conoce como "Web Semántica", la información digital que permite que tengan lugar varios tipos de procesamiento automático, que no sería posible si el texto fuera sólo texto y la imagen fuera sólo imagen, sin nada más que lo que un ser humano necesitaría para comprenderlo.

La incorporación de información adicional

requiere esfuerzo. No todo el mundo es capaz de hacerlo y es necesario un aprendizaje. Pero la cuestión es si merece la pena el esfuerzo. Hacer que nuestros programas sean más inteligentes es otra solución, pero esto también requiere un esfuerzo, sin mencionar el tiempo invertido en investigar. Para simplificar un poco, podemos decir que esta última solución reflejaría la esencia de la Inteligencia Artificial.

Un ejemplo en el que el esfuerzo en codificación semántica es casi siempre efectivo es en la separación del texto entre HTML y CSS (*Cascading Style Sheets*). Codificamos el texto y su estructura (listas, encabezados, secciones, frases destacables, etc.) en HTML y los aspectos de presentación (colores, márgenes, tipos de letra, o en el caso de audio: voz, volumen, velocidad, etc.) en CSS.

No todo el mundo es capaz de separar mentalmente la presentación de la estructura, pero aquellos que pueden, encuentran que hay un beneficio en hacerlo, no sólo para sus lectores, sino para ellos mismos también. Principalmente, la ventaja es que el texto puede ser adaptado de forma sencilla a diferentes dispositivos (pantallas pequeñas, grandes, impresoras...), y además puede ser reutilizado en un contexto en el que sea necesaria una presentación diferente.

Visto de esta forma, la Web Semántica no es una meta concreta que alcanzaremos un día, sino una dirección en la que queremos ir, por ejemplo hacia el "Este". Y la *semántica* en este contexto se define mejor, y en cierto modo irónicamente, como *esa información de la que todavía no tenemos una notación formal*.

2. Un ejemplo: un calendario de eventos

Todo el mundo tiene un ejemplo favorito de lo que es la Web Semántica y lo que es capaz de hacer.

Aquí está uno de los míos.

Imagina que vives en la preciosa ciudad de Niza. Durante el fin de semana, deseas disfrutar de alguna de las atracciones de la ciudad: ir a un concierto, ver una película, ir al museo, etc. Visitas algunos sitios web en los que encuentras eventos culturales: la agenda de la ciudad, la agenda de la ópera, la agenda de un sitio internacional especializado en danza, la sección de cultura del periódico local...

La página parece que funciona bien. Hay muchos recursos de información y bastantes de los acontecimientos que van a tener lugar en la ciudad están anunciados. Gracias al estándar HTML toda esta información está al alcance de tus manos y todo lo que necesitas es un navegador web.

Pero lleva tiempo leer y comprender toda la

información. Todos estos sitios tienen su propia presentación y estructura, y es necesario manejar diferentes niveles de información, diferentes presentaciones, y probablemente hacer frente a diferentes formas de interacción cuando alguien decide utilizar JavaScript para hacer su página "más fácil" de usar (pero desafortunadamente, bastante diferente a otros sitios).

En este caso, un formato de calendario común, añadido a HTML, es lo que sería necesario para que la información pudiera ser comparada fácilmente mostrando los diferentes recursos en un calendario único. iCalendar [5] es el estándar de Internet para hacerlo. hCalendar [6], el microformato [7], ayuda a los autores a manejar la información del calendario permitiéndoles escribir el HTML necesario (siguiendo ciertas convenciones) y obtenerlo automáticamente como iCalendar cuando les es necesario (no pude encontrar un ejemplo para Niza, pero tengo uno para Besançon [8] y Birmingham [9]).

En este momento, en 2009, la mayoría de los eventos en Niza están disponibles en la Web en formato HTML pero una parte muy pequeña está en iCalendar (o hCalendar). Esto se debe en parte a que no se conocen o no se han visto los beneficios todavía, o bien porque algunos desarrolladores no *quieren* hacer que la búsqueda de información sea sencilla. Por ejemplo, porque quieren que la gente esté el mayor tiempo posible viendo los anuncios publicados en sus sitios web.

Pero vamos a dejar este tema para otro momento y asumir que algunos proveedores de información *han convertido* la semántica en sintaxis, por ejemplo en iCalendar o hCalendar.

Ahora ya tienes todos los eventos de todos los sitios web ordenados en un único calendario en tu pantalla. Gracias a un estándar común para algunos tipos de semántica nos hemos librado de los problemas de presentación y navegación y todo está ahora en un único programa con una única interfaz. Es mucho más fácil buscar y para una ciudad de tamaño medio como Niza podríamos detenernos aquí, pero quizás el ordenador nos pueda ayudar más.

Puede que quieras filtrar y ordenar los eventos siguiendo más criterios que sólo la fecha, como puede ser el tipo de evento o lo fácil que es llegar a él. Puede que quieras introducir en tu navegador preferencias tales como que estás interesado en películas subtítuladas, y no dobladas, o que estás interesado en eventos que estén a menos de 10 kilómetros de donde vive tu novia.

Esta información está disponible leyendo cuidadosamente la descripción de cada evento, aunque probablemente sólo de forma indirecta (la dirección se da, pero no la distancia,

porque ésta depende del lector). Por lo tanto, necesitamos estándares adicionales para que esa información esté disponible en un formato que puedan interpretar también las máquinas.

La ubicación (dirección) de un evento se puede mostrar en un formato (utilizando iCalendar o hCalendar) de tal manera que un programa pueda extraer la ubicación y pueda calcular la distancia a otro punto, por ejemplo utilizando algún planificador de rutas en línea.

El tipo de evento (concierto, película, teatro, danza, exposición o incluso categorías más específicas) no es todavía algo fácilmente expresable. Un estándar, aunque sea uno sólo, no existe todavía.

Por supuesto, podríamos discutir sin fin lo que debería haber en una taxonomía de eventos (¿sólo "película"? o ¿"película romántica"? o ¿"película romántica china"?), pero incluso una sola ya sería muy útil.

Quizás, incluso entonces no sería suficiente, porque hay un montón de información en el calendario y muchos de los eventos estarían más de una vez: el mismo concierto está anunciado en el sitio web de la ópera y en el sitio web de la orquesta. ¿Cómo eliminamos duplicados?, ¿necesitamos identificadores únicos para cada evento?

No, eso es llegar demasiado lejos. Para un autor es fácil etiquetar el tipo de evento y la dirección. Es algo parecido a cuando alguien recibe la información en un trozo de papel y la introduce en la base de datos del sitio web donde ya hay una distinción entre esos campos. Pero un identificador global para el evento necesita una coordinación global. Esto es ya demasiada complicación.

Y entonces ¿por qué no puede el ordenador hacerlo como nosotros lo hacemos? Reconocemos duplicados de forma sencilla y si nos saltamos unos pocos no es ningún problema, sólo son unas pocas entradas más en nuestro calendario.

Parece ser que hemos vuelto a alcanzar la intersección entre Web Semántica y la Inteligencia Artificial. El trabajo extra que un autor lleva a cabo para que la información sea comprendida por las máquinas tiene que ser razonable comparado con los beneficios que se esperan. En este caso, existe ciertamente un beneficio manejando un identificador único global para cada uno de los eventos, pero el esfuerzo de obtener uno es considerable, mientras que los ordenadores son ya capaces de reconocer duplicados razonablemente bien por su cuenta, por lo que la relación coste-beneficio es demasiado alta.

Por un momento vamos a volver a la pregunta de la taxonomía (vocabularios o algunas veces

ontologías). ¿Quién los desarrolla? W3C crea algunos estándares que expresan parte del significado (semántica) de la información: HTML, CSS, SVG, P3P, XML, etc. Otras organizaciones desarrollan otros (HTTP, Unicode, MIME, etc.) ¿Quién haría una categorización simple (¡o compleja!) de eventos culturales?

No es fácil responder a esa pregunta. Hacer un estándar requiere reunir varios prerequisites: expertos en el campo por supuesto; un proceso y las herramientas de soporte que ayuden a asegurar que el desarrollo se hace de forma sencilla y representa a todas las partes; dinero, para pagar a los expertos que invierten su tiempo y las herramientas que utilizan; autoridad y marketing, para estar seguros de que el resultado se hará público y se utilizará; mantenimiento, en el sentido de mantener el estándar disponible tanto tiempo como sea necesario y actualizado en el caso de que las circunstancias requieran modificaciones; herramientas de cumplimiento, normalmente en formato prueba, para ayudar a los implementadores.

3. El futuro inmediato

Las agendas culturales son una de las aplicaciones de la Web Semántica y donde el progreso se está realizando haciendo la información reutilizable, comparable y adaptable a las necesidades del usuario, pero hay otras. Una organización de estandarización como el W3C está siempre atenta a aplicaciones concretas con rasgos comunes, porque encontrar soluciones significa, por un lado, que hay un uso inmediato de esas soluciones y una buena oportunidad de que el trabajo sea reutilizable para otras aplicaciones futuras.

XML, y más recientemente las tecnologías EXI (*Efficient XML Interchange*), son buenos ejemplos. XML capturó los aspectos comunes de diferentes tipos de sistemas de documentación basados en SGML (*Standard Generalized Markup Language*), es decir, que contenían documentos multilingües estructurados con piezas ordenadas e independientes; lo que resuelve parte del problema de intercambio de documentos entre estos sistemas. XML se ha mostrado útil para otros muchos tipos de información para los cuales no existen todavía estándares de intercambio; incluso para algunos que no tienen nada que ver con documentos y que quizás incluso no contienen nada de lenguaje natural. EXI funciona de igual manera para información binaria estructurada, especialmente cadenas de datos estructurados.

Cada vez más, se le da más atención a los aspectos de seguridad, privacidad y derechos (tales como derechos de autor) en sistemas electrónicos, incluida la Web. De nuevo, organizaciones como el W3C buscan aspectos comunes a los diferentes problemas. Encontrar soluciones puede ser más complicado que hacerlo para XML o los calendarios, porque la propia solución del problema está

culturalmente definida. Lo que es normal en un país es impensable en otro (o, lo que es quizá aún más difícil: es casi lo mismo en otro país, pero no *exactamente*). Si esto hace el problema más difícil de resolver o más fácil es una cuestión de punto de vista: en un caso, la solución puede ser derivada matemáticamente, pero las matemáticas no son el fuerte de mucha gente; en el otro caso, se interactúa con personas y por lo tanto quizás se puede tratar de cambiar a la gente en su lugar...

Varias iniciativas del W3C muestran que se demanda el progreso en estas áreas y que probablemente tenga lugar en un futuro cercano: el Grupo de Interés llamado PLING (*Policy Languages Interest Group*) [10] se estableció como un foro general para discutir todos los tipos de "lenguajes de directivas", por ejemplo, varias formas de convertir la semántica de los derechos en sintaxis.

Hay varios lenguajes desarrollados, bajo estudio, o propuestos. El Grupo de Trabajo de Contexto de Seguridad Web (*Web Security Context Working Group*) [11], trata de establecer el (inicialmente pequeño) conjunto de aspectos relacionados con la seguridad que un navegador debería poder transmitir a los usuarios sobre la información que ve y sobre el servicio con el que está interactuando.

Una propuesta procedente de Creative Commons [12] apuesta por un lenguaje concreto para expresar subconjuntos de derechos de autor en formato comprensible para las máquinas. El grupo de trabajo de Anotaciones Multimedia [13] trata de definir un vocabulario común para algunos aspectos del vídeo, incluyendo los derechos de autor. Finalmente, hay una propuesta de tecnología para "incrustar" fuentes en documentos web [14], donde "incrustar" (*embed*) se refiere a un tipo particular de licencia de derechos de autor que es muy popular entre los proveedores de fuentes (*fonts*), más que poner físicamente un archivo dentro de otro. (Esta propuesta es interesante por otras razones, que no tienen nada que ver con los derechos de autor, o con convertir la semántica en sintaxis; pero ya hablaré más sobre esto más tarde).

Otro desarrollo, del que podemos esperar frutos muy pronto, es la armonización de datos de la administración pública. El W3C tiene un grupo de interés en eGovernment [15], debido a que los gobiernos se han dado cuenta de que los datos que manejan tienen que extenderse a través de los límites nacionales.

En la Unión Europea, por ejemplo, la gente viaja de un país a otro de forma frecuente para realizar transacciones. En muchas fronteras, las comunidades a ambos lados han descubierto que es más eficiente compartir servicios con las ciudades cercanas a las fronteras que con aquellas que están lejanas en la misma

zona, incluyendo hospitales, servicios de emergencia, transporte público, instituciones culturales, policía, etc. (La Euroregión más vieja, llamada Euregio [16], cerca de Enschede en el lado alemán-holandés, tiene 50 años, aunque algunas tengan menos de 15). Pero las fronteras generan problemas, y algunos de ellos pueden resolverse con estándares para la Administración Electrónica (eGovernment).

4. La práctica de la estandarización

Parece que sabemos dónde queremos ir, hacia la Web Semántica, y aunque no sabemos lo lejos que tenemos que ir, ya tenemos suficiente que hacer. Entonces, ¿por qué no hacemos más estándares, por qué no creamos el formato para las fuentes incrustadas, establecemos uno o dos formatos de vídeo, o desarrollamos un vocabulario para el sector farmacéutico que sea comprensible por las máquinas (otra área donde el W3C tiene un grupo de interés)?

Vamos a detenernos, por ejemplo, en ese estándar para las fuentes incrustadas. El problema a resolver es que los diseñadores pueden especificar de forma extendida (dentro de los límites de HTML, CSS, SVG, XSL y otros estándares relacionados) la forma en la que se muestra un documento, en relación a los colores, el formato, alto de línea, etc., pero están muy limitados en los tipos (de caracteres) que pueden usar. Técnicamente, pueden especificar cualquier tipo arbitrariamente e incluso dónde descargar el archivo de la fuente si no está disponible en el dispositivo del lector. Pero muchas de las fuentes están regidas por licencias que prohíben descargarlas. El resultado es que los diseñadores pueden hacer un documento PDF con casi cualquier fuente que tengan, pero no un documento HTML, porque no hay un estándar web para incrustarlas.

Algunas veces, la resolución de un problema supone una investigación técnica complicada, formación de los programadores sobre nuevas técnicas y un periodo largo de pruebas y detección de errores hasta que un software es lo suficientemente bueno para soportar esas nuevas técnicas. CSS es un ejemplo de esto. Han sido necesarios muchos años antes de que los programadores controlaran de forma avanzada el modelo, porque para muchos de ellos era diferente de lo que habían estado utilizando en el manejo de documentos. En concreto, la noción de (la cual HTML heredó de SGML y ahora es parte de XML) que un documento puede ser representado como un árbol de estructuras anidadas, más que como una cadena de cambios de estado, fue algo ajeno a muchos programadores de los primeros navegadores.

Pero para las fuentes incrustadas la dificultad no es realmente el problema. Alguien en Microsoft, nadie recuerda quién exactamente, tuvo la idea hace diez años del equivalente

al "incrustado" en la Web.

En un documento compuesto tradicional, todas las piezas están en un archivo grande y por lo tanto, no hay dudas sobre el documento al que pertenecen. Los documentos compuestos en la Web, por otro lado, consisten en partes separadas descargadas individualmente: el texto en HTML, algunas imágenes en JPEG, una hoja de estilo en CSS, etc. Hay un enlace desde HTML a la hoja de estilo, pero no al revés, por lo que si se comienza por la hoja de estilo, no se sabe de qué documento es parte. (Esto es frecuentemente una ventaja y la Web la explota a gusto). Pero si, como es el caso con muchas fuentes, la licencia necesita que una fuente sea distribuida sólo como parte de un documento (es decir, un documento escrito por el titular) y no por sí sola, entonces ese modelo no funciona. La solución más inteligente es hacer dos enlaces: un enlace de retorno desde la fuente indica el documento al que pertenece la fuente.

Microsoft implementó esto en un formato llamado EOT (*Embedded Open Type*) [17], derivado de OpenType, un estándar ISO para fuentes digitales. Pero EOT es un formato propietario, no un estándar. CSS permite utilizarlo, pero como nadie más que Microsoft lo puede implementar, nunca será muy popular.

Tenemos por lo menos dos soluciones que permiten que un número extenso de fuentes estén disponibles para documentos web: cambiar la práctica de las licencias para que las fuentes se puedan utilizar sin incrustar, o estandarizar una tecnología que enlace una fuente a un documento (o varios documentos al mismo tiempo, como lo hace EOT, para mayor eficacia).

Lo primero no es probable que tenga éxito a corto plazo, porque nadie conoce todavía la forma en la que los creadores de fuentes ganarían dinero si sólo venden cada fuente una sola vez. De forma que ha de inventarse un modelo económico que no esté basado en la venta de licencias. Nos queda entonces una solución técnica: hacer un formato de fuente con un enlace interno, como Microsoft hizo con EOT, pero hacerlo estándar. (De hecho, podría ser EOT. Microsoft está dispuesto a donarlo aunque con diez años de experiencia sería fácil mejorarlo). El software de las aplicaciones, como navegadores que muestran o imprimen documentos web, miraría el enlace para comprobar si la fuente realmente pertenece al documento e ignorarlo si no.

Pero todavía no está pasando nada. Los creadores de fuentes han dicho que la solución les gusta. No hay ningún DRM (*Digital Rights Management*) relacionado, no hay intención de hacer que sea difícil usar una fuente, sólo haría falta esta anotación (el enlace) que hace que la licencia sea lo suficientemente com-

preensible para las máquinas para que el software sepa lo que significa; y los creadores creen que es suficientemente bueno (y encaja con el *zeitgeist* o "espíritu" de esta época).

Los diseñadores quieren utilizar en la web las fuentes que compraron para otros documentos (como PDF). Los piratas de software no tienen tampoco problema porque quitar o cambiar el enlace es sencillo. Por supuesto, hay todavía problemas técnicos que investigar (¿están todos los tipos de *caching* y *proxying* cubiertos?, ¿y sobre documentos con URL aleatorias, como ID de sesión?, ¿se pueden incrustar fuentes en documentos enviados por e-mail?) pero hasta el momento ninguno parece particularmente problemático.

Y todavía hay fabricantes de navegadores que han dicho que no van a implementar una solución como esa, o por lo menos no ahora. ¿Por qué? Aquí estamos tocando problemas de estandarización.

Es una cuestión de recursos: sólo hay algunas horas en el día, sólo algunos programadores capaces de hacer ciertas cosas y algunas personas que pueden trabajar en el mismo proyecto al mismo tiempo. El desarrollo de un estándar para fuentes incrustadas necesita gente: para tratar el tema, escribir, implementar y probar la tecnología. Esta gente no puede entonces trabajar en otros proyectos. Y hay muchos otros proyectos en marcha al mismo tiempo. Se trata de una cuestión de prioridades: estimar la urgencia de los proyectos y el orden más eficaz para ellos.

Y, como manejamos en gran medida proyectos con aspectos comerciales, están también frecuentemente las directivas del marketing: ¿por cuánto tiempo es mejor hacer algo diferente a la competencia en lugar de llegar a un acuerdo sobre un estándar? ¿cuál es el mejor momento para pedir apoyo para un estándar? ¿es posible vender una solución dos veces, una vez como una solución incompleta y otra vez cuando cumpla con el estándar? ¿por cuánto tiempo es posible demorar la ratificación de un estándar de tal forma que los competidores que han invertido en él no pueden rentabilizar su inversión todavía?

Y esto parece ocurrir con la cuestión de las fuentes incrustadas, así como con otras tecnologías que son buenas ideas, pero que tienen que esperar a que llegue su turno.

Veremos soluciones parciales de algunos implementadores, intentos por parte de aquellos que no dominan todavía la tecnología de retrasar y cambiar el estándar, de forma que las inversiones de aquellos que van en cabeza pierdan algo de valor. Un retraso de dos años o más no es de extrañar.

Una organización de estandarización como el

W3C tiene que enfrentarse a eso, tener paciencia y suficiente visión como para vislumbrar el futuro a pesar de las políticas de corto plazo. También necesita ser fuerte ya que cuando el desarrollo de las cosas va despacio, la gente se queja a los autores de los estándares. W3C se enfrenta a esto especialmente porque es muy conocido y abierto: cualquiera puede unirse a la lista de correo del W3C o enviar un correo a los autores de un estándar, porque todos los documentos están firmados con los nombres y direcciones de correo electrónico de la gente que los han desarrollado, algunas veces incluso con una foto. (Eso se debe a que las contribuciones y los comentarios positivos pesan más que los negativos).

5. Conclusiones

El éxito de la Web se debe en gran parte al hecho de que ésta permite a los usuarios de la información adaptarla a sus necesidades. No son muy dependientes de las elecciones particulares que el autor ha hecho, y pueden combinar diferentes fuentes y extraer información importante mientras que ignoran el resto. El W3C piensa que el futuro radica en hacer la información cada vez más reusable.

Muchos autores podrían hacer un trabajo mejor del que están haciendo en la actualidad y hacer más fácil la utilización de la información en formas que no han imaginado, pero incluso un poco de HTML y CSS, hace posible aplicaciones interesantes. (Además del uso incorrecto del HTML, los *scripts* y las *cookies* son impedimentos importantes para desarrollar aplicaciones nuevas para la Web).

Nuevos estándares están desarrollándose y desplegándose, haciendo posible una mayor conversión de semántica en sintaxis, y permitiendo a los ordenadores ayudar más a los usuarios para darle más sentido a la información en la Web.

No obstante, la pregunta que permanece es ¿dónde están los límites de la sintaxis? ¿cuánto del significado implícito puede una persona hacer explícito?

Estos límites son diferentes para cada persona y además dependen de su formación. Dependen también de su motivación: ¿el beneficio merece el esfuerzo? No obstante, está claro que mucho del significado de cualquier mensaje permanecerá implícito, escondido en representaciones que en apariencia parecen expresar algo más. La Web Semántica necesitará la contribución de la Inteligencia Artificial si los ordenadores tienen que ayudar a los usuarios a encontrar, catalogar y comparar ciertos tipos de información.

Sin embargo, todavía no hemos alcanzado los límites y además hay mucho que puede ser mejorado en la forma en la que representamos digitalmente la información.

Se necesitan más estándares. Pero los estándares necesitan un tiempo. La gente necesita entender el problema, llegar a un acuerdo sobre una solución a pesar de los conflictos de intereses, y otras personas necesitan aprender los conceptos que hay detrás de las soluciones y acumular experiencia en su aplicación.

Algunas áreas en las que podremos ver nuevas aplicaciones interesantes en los próximos dos o tres años serán la gestión de calendarios, la información basada en la ubicación, la gestión de los derechos de autor, la seguridad y la tipografía digital.

Referencias

- [1] **Wikipedia**. Proyecto Xanadu <http://en.wikipedia.org/wiki/Project_Xanadu>.
- [2] **The Free Network Project**. <<http://freenetproject.org/>>.
- [3] **Internet Archive**. Universal access to human knowledge. <<http://www.archive.org/index.php>>.
- [4] **Tim Berners-Lee**. *Cool URLs don't change*. <<http://www.w3.org/Provider/Style/URI.html.en>>. Traducido al castellano en: <<http://www.w3.org/Provider/Style/URI.html.es>>.
- [5] **F. Dawson, D. Stenerson**. *Internet Calendaring and Scheduling Core Object Specification (iCalendar)*. RFC 2445. <<http://tools.ietf.org/html/rfc2445>>.
- [6] **Tantek Çelik, Brian Suda**. *hCalendar*. Microformats Specification. <<http://microformats.org/wiki/hcalendar>>.
- [7] **Microformats**. <<http://microformats.org/>>.
- [8] *Le portail officielle du Grand Besançon*. Besançon, Francia. <http://www.besancon.fr/index.php?p=44&art_id=141>.
- [9] *What's on - All events*. Birmingham, Reino Unido. <<http://www.thsh.co.uk/events>>.

- [10] **W3C**. *W3C Policy Languages Interest Group - PLING*. <http://www.w3.org/Policy/pling/wiki/Main_Page>.
- [11] **W3C**. *Web Security Context Working Group*. <<http://www.w3.org/2006/WSC/>>.
- [12] **W3C**. *W3C Member Submission Request: ccREL - the Creative Commons Rights expression Language*. <<http://www.w3.org/Submission/2008/02/>>.
- [13] **W3C**. *Media Annotations Working Group*. <<http://www.w3.org/2008/WebVideo/Annotations/>>.
- [14] **W3C**. *Fonts Activity Proposal*. <<http://www.w3.org/2008/09/fonts-activity-v2>>.
- [15] **W3C**. *eGovernment at W3C*. <<http://www.w3.org/2007/eGov/>>.
- [16] *EUREGIO - Living without borders*. <<http://www.euregio.de/cms/publish/content/showpage.asp?themeid=23>>.
- [17] **W3C**. *Embedded OpenType (.EOT) Font Format Submission Request to W3C*. <<http://www.w3.org/Submission/2008/01/>>.



XIV Jornadas de Ingeniería del Software y Bases de Datos

San Sebastián, 8-11 de septiembre de 2009



Las Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2009) solicitan contribuciones de trabajos de investigación originales y de calidad, o de trabajos de carácter profesional describiendo experiencias industriales de interés, así como propuestas de tutoriales, talleres, demostraciones de herramientas científicas, y de trabajos relevantes ya publicados durante el año 2008 en revistas o actas de congresos de prestigio.

JISBD 2009	Recepción de contribuciones	Notificación de aceptación	Versión definitiva
Artículos Demostraciones de herramientas Divulgación de trabajos ya publicados	27 de Abril	12 de Junio	24 de Junio
Talleres y Tutoriales	2 de Marzo	23 de Marzo	



Contacto: jisbd2009@eps.mondragon.edu
<http://www.mondragon.edu/jisbd2009/>



Raúl García Castro, Andrés García Silva

Ontology Engineering Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid

<rgarcia@fi.upm.es>
<hagarcia@delicias.dia.fi.upm.es>

1. Introducción

En los últimos años el contenido de la Web ha evolucionado, pasando de una Web formada por páginas HTML enlazadas entre sí a una Web repleta de contenidos multimedia, servicios y aplicaciones, los cuales son generados, clasificados y evaluados por los usuarios.

La búsqueda, acceso, comprensión y uso de tal cantidad de contenidos puede ser facilitada mediante la expresión de la semántica de estos recursos, para lo cual es necesario definir metadatos que los describan y asignar dichos metadatos a los recursos. Esto es lo que se conoce como el proceso de anotación semántica de recursos Web.

El concepto de "anotación semántica" podría describirse como caleidoscópico, pues adopta diferentes formas según la perspectiva desde la que se observe [1]. Explicitar el significado, es decir, la semántica de las palabras o de un documento y realizar su anotación se ha llevado a cabo de diversas maneras, dependiendo del área de investigación. En consecuencia, también son múltiples los términos acuñados en los trabajos sobre este tema para referirse a esta actividad, a veces sin diferencia aparente. En inglés, lengua en la que se publica prioritariamente, algunos de los términos utilizados son: *semantic annotation*, *semantic markup*, *semantic tagging* o *semantic labelling*.

Las distintas áreas que tratan la anotación semántica proporcionan diferentes puntos de vista sobre el proceso. En el campo de la lingüística computacional, el objetivo de la anotación es identificar y describir el significado de los textos, mediante etiquetas que corresponden a los diferentes niveles que aportan significado (morfosintáctico, semántico, discursivo, pragmático, por ejemplo). En la Web Semántica y en la Web 2.0, el objetivo de la anotación es poder incluir información no sólo de textos sino de cualquier contenido multimedia que pueda estar disponible en la Web. Mientras que en el caso de la Web Semántica se trata de anotar a partir de modelos de dominio formales (llamados ontologías), en el caso de la Web 2.0 se anota utilizando descriptores (llamados etiquetas).

Todas estas perspectivas de la anotación han de tenerse en cuenta para lograr una explotación completa del contenido de la Web, desde la agilidad del etiquetado de la Web 2.0 a la semántica formal de las ontologías. Lograr fusionar la Web Semántica y la Web 2.0 (lo

Anotación de contenidos en la Web del futuro

Resumen: en este artículo presentamos el estado y las tendencias actuales de los enfoques y herramientas de anotación de recursos Web en el contexto de las áreas de la Web Semántica, Web 2.0 y lo que se ha denominado Web 3.0, es decir, la fusión de la Web Semántica y la Web 2.0. Uno de los principales retos en la anotación de recursos Web está relacionado con la gran variedad de contenidos presentes actualmente y en el futuro, incluyendo no sólo textos sino grandes cantidades de contenidos multimedia, servicios y aplicaciones que son generados, clasificados y evaluados por los usuarios.

Palabras clave: anotación, Web Semántica, Web 2.0, Web 3.0.

Autores

Raúl García Castro es Doctor en Ciencias de la Computación e Inteligencia Artificial por la Universidad Politécnica de Madrid (UPM) y trabaja como investigador en el grupo de investigación *Ontology Engineering Group* de la Facultad de Informática de la UPM. Después de trabajar durante tres años como ingeniero de software, desde que se licenció en Informática (2003) ha estado trabajando en la UPM dentro del *Ontology Engineering Group*, donde defendió su tesis doctoral en 2008. Ha trabajado en varios proyectos de investigación nacionales y europeos centrándose en la evaluación y el *benchmarking* de tecnologías de la Web Semántica. Participa habitualmente en los comités de programa de las conferencias y talleres más relevantes del área y ha organizado talleres y conferencias internacionales.

Andrés García Silva es candidato a doctor en Inteligencia Artificial por la Universidad Politécnica de Madrid (UPM). Previamente obtuvo los títulos de Máster de investigación en Inteligencia Artificial por la UPM (2008), de Especialista en Administración de Empresas por la Universidad ICESI de Cali (Colombia, 2005), de Ingeniero de Sistemas por la Universidad del Valle de Cali (2001) y de Profesional certificado en Oracle por Oracle Corporation (2003). Tiene especial interés en la Web Semántica, la Web 3.0 y sus áreas relacionadas, incluyendo el desarrollo de ontologías. Es miembro del *Ontology Engineering Group* perteneciente al Laboratorio de Inteligencia Artificial de la Facultad de Informática de la UPM. Además ha participado en los proyectos de investigación europeos NeOn y SEEMP, en el proyecto español Geobuddies y en el proyecto CENIT España Virtual.

que se ha denominado la Web 3.0) permitirá obtener lo mejor de los dos mundos e integrar las etiquetas de la Web 2.0 con la semántica formal de las ontologías.

En este documento se resumen los distintos enfoques adoptados en la anotación en los campos de la Web Semántica y la Web 2.0, y la Web 3.0, y se identifican tendencias actuales de trabajo.

2. Anotación en la Web Semántica

En la comunidad informática el término *marcado semántico* (*semantic markup*) se empleó inicialmente para denotar la descripción de los contenidos de una página Web adjuntando a ciertos elementos del texto una etiqueta que identifica de alguna manera su referente en el mundo real (en este caso, en términos computacionales, se emplea lo que se conoce como *ontologías*). Para codificar el significado se utilizaban los denominados lenguajes de marcado, tales como XML [2], RDF(S) [3] u OWL [4].

Según Shea [5], el marcado de una página

Web puede ser de dos tipos, marcado para la presentación (*presentation markup*) y marcado semántico (*semantic markup*). El *marcado para la presentación* consiste en etiquetar una página Web con información sobre las características formales del documento mediante *HTML* (es decir, cómo debe mostrarse al lector, con negrita, mayúsculas, colores, etc.). El *marcado semántico*, en cambio, incluiría información sobre el contenido de la página Web en cuestión (documento, imagen, URL, etc.) y estaría implementado en *XML* o en cualquier otro lenguaje de marcado basado en él.

En cualquier caso, la preparación de documentos para su publicación en la Web Semántica mediante anotaciones que expliquen su contenido y que sean procesables por un ordenador conlleva al menos tres posibles fases: (1) la elección del lenguaje que se utilizará; es decir, ha de ser uno de los lenguajes basados en *XML* y desarrollados para la Web Semántica, como *RDF*, *RDF(S)* u *OWL*, entre otros; (2) la adición, en el lenguaje elegido, de comentarios o anotaciones semánticas (metadatos sobre los documen-



"Una buena parte de la investigación en el campo de la Web Semántica se dirige hacia la anotación de datos multilingües"



tos o páginas Web que aportan cierto significado) a los recursos que desean publicarse en la Web Semántica; y (3) el razonamiento con estos metadatos semánticos. Una condición fundamental para que estos metadatos tengan valor y permitan el razonamiento es su especificación de forma consensuada para que pueda ser compartida por los distintos agentes participantes. Precisamente para alcanzar este consenso sobre el significado nacen las ontologías. Una *ontología* se define como una "especificación formal y explícita de una conceptualización compartida" [6]. De forma informal, se puede decir que una ontología es un modelo de conocimientos consensuado en un determinado dominio, que proporciona un vocabulario compartido y controlado, que posibilita la comunicación entre personas y aplicaciones.

Las ontologías aparecen como la espina dorsal de la anotación de documentos en las aplicaciones previas a la Web Semántica, como el proyecto SHOE, la iniciativa (KA)² y el proyecto Planet-Onto, entre otros. Con la aparición de la Web Semántica, la anotación basada en ontologías ha sido el foco de numerosos proyectos y aplicaciones, dado que la existencia de contenido anotado es uno de los retos más importantes para conseguir hacerla realidad. Entre estos proyectos y aplicaciones se pueden citar los proyectos europeos Esperonto¹ y AceMedia², la Red de Excelencia Europea SCHEMA³, o el proyecto estadounidense MindSwap⁴. Todos ellos comparten la premisa de proporcionar herramientas y marcos de trabajo para anotar distintos tipos de contenidos (HTML, bases de datos, documentos multimedia, etc.) con distintos niveles de detalle. El estado de la cuestión sobre anotación semántica ha sido analizado en [7] y [8]⁵.

Existen diversos grados de formalidad en los que se pueden realizar las anotaciones de un documento. En líneas generales se puede distinguir entre anotaciones que utilizan vocabularios como *Dublin Core* (que describen en líneas generales el documento sin entrar en detalles sobre su contenido concreto), anotaciones basadas en tesauros y listas de términos (que describen el contenido de los documentos relacionándolo con términos que aparecen en dichos vocabularios), y anotaciones basadas en ontologías (que se centran en la creación de instancias de clases y relaciones de ontologías formalmente especificadas

en algún lenguaje como RDF Schema u OWL y que, por tanto, son las de mayor precisión).

Una clasificación detallada de los distintos tipos de anotación que pueden aplicarse a páginas de la Web Semántica puede encontrarse en [9]. Por su parte, en [10] se incluye una discusión acerca del grado de formalización que alcanzan los distintos tipos de anotación posibles.

2.1. Anotación multimedia

La anotación de documentos multimedia todavía sigue siendo un área en el que se está trabajando intensamente, y existen grandes proyectos en ejecución donde uno de sus objetivos principales es el de proporcionar herramientas que permitan realizar estos tipos de anotaciones, tal como X-Media⁶ [11], y otros proyectos pasados en los que también se ha trabajado en esta área, tales como Acemedia⁷ [12], MUMIS⁸ o Esperonto [13]. También hay un grupo activo asociado al consorcio W3C que trabaja proporcionando recomendaciones⁹.

Algunas de las propuestas consisten en la anotación de los textos que acompañan a una determinada imagen en un documento, dado que normalmente dichos textos tratarán sobre la imagen en cuestión y pueden proporcionar información interesante sobre la misma. Sin embargo, este tipo de anotación de archivos multimedia no proporciona suficiente calidad, dado que no siempre el texto da suficiente información sobre los archivos adjuntos.

En otro sentido, uno de los principales problemas que debe abordarse al realizar anotación de archivos multimedia, especialmente en el caso de imágenes y vídeos, es lo que se conoce como la *brecha semántica*. Por un lado, las técnicas de análisis de imágenes permiten obtener una gran cantidad de descriptores de una imagen, basadas en las características de sus píxeles, y permiten incluso identificar objetos y su posición con respecto a la imagen. Sin embargo, aún es difícil pasar de ahí a la anotación de los objetos concretos que se encuentran en la imagen y, sobre todo, la relación entre ellos (por ejemplo, "esta foto trata sobre un niño que está en la playa"). Este es precisamente el foco de los proyectos AceMedia o X-Media, entre otros.

2.2. Tecnología actual de anotación

De manera muy básica, las herramientas de

anotación basadas en ontologías permiten poblar ontologías con instancias procedentes de distintos tipos de recursos. Debido a la heterogeneidad de estos recursos, estas herramientas son muy heterogéneas. Algunas cubren todo el ciclo de vida de la anotación y otras sólo parte del mismo y, aunque gran parte de ellas trabajan sobre documentos de texto, algunas ofrecen la posibilidad de anotar contenido de la Web o recursos multimedia.

La mayor parte de las herramientas de anotación permiten la anotación manual o semiautomática de contenidos, estas últimas utilizando técnicas basadas en extracción de información o aprendizaje automático. La anotación manual de documentos es una tarea costosa y que puede llevar a gran número de errores, como se ha demostrado en la mayor parte de las iniciativas de la Web Semántica y de las anteriores a ésta. Por tanto, el esfuerzo en anotación debería centrarse en la automatización de esta tarea, dejando a los usuarios la tarea de supervisión de las anotaciones propuestas. Para ello, la calidad de la anotación automática debe también mejorar, con el objetivo de hacer más productivo todo este proceso.

Además de tratar con la calidad de las anotaciones generadas, es importante que las herramientas de anotación gestionen adecuadamente las inconsistencias que se pueden producir en los datos anotados de manera distribuida, el ciclo de vida de las anotaciones y de sus vocabularios relacionados (por ejemplo, la gestión de la evolución de los vocabularios en los cuales las anotaciones están basados), la existencia de sistemas de gestión de anotaciones para realizar consultas, almacenamiento, razonamiento, etc. Todo esto son líneas que se están tratando actualmente.

2.3. Tendencias

Hoy en día, una buena parte de la investigación en el campo de la Web Semántica se dirige hacia la anotación de datos multilingües, con el fin de que las búsquedas de información recuperen los documentos relevantes sea cual sea la lengua en la que estén escritos. También se trabaja en el tratamiento de la calidad de las anotaciones generadas.

Por otro lado, en el área de anotación semántica de documentos multimedia todavía queda mucho trabajo por realizar, y la mayor parte de los esfuerzos se centran en el proble-

ma de la brecha semántica, que se puede definir como el problema de pasar de las representaciones basadas en análisis de imágenes y vídeos a representaciones semánticas de las mismas, en las que se dice que un objeto concreto es por ej. una mesa o una silla, y que dicho objeto está relacionado con otro de los que aparecen en la imagen o vídeo de alguna manera (está encima de él, una persona está sentada en una silla, etc.).

3. Anotación en la Web 2.0

En los últimos años se ha pasado de una Web donde el contenido era generado por los creadores de los sitios Web a una Internet más abierta y social donde los usuarios ya no son sólo consumidores de información sino también creadores. Esta nueva etapa, conocida como la Web 2.0¹⁰, trajo consigo una variedad de aplicaciones como *wikis*, *blogs*, redes sociales y sitios Web para compartir direcciones en Internet, fotos, vídeos o música. Toda esta nueva gama de aplicaciones conlleva nuevos contenidos para los cuales los métodos tradicionales de anotación y recuperación de información no son suficientes. Surge entonces la anotación mediante la asignación de palabras clave a recursos, los cuales pueden posteriormente ser recuperados utilizando dichas palabras clave. Estas palabras clave se conocen como etiquetas y el proceso de asignación de etiquetas se denomina etiquetado.

Tener el contenido de la Web etiquetado permite clasificarlo, adaptarlo y relacionarlo con otros contenidos. Cualquier usuario puede etiquetar el contenido de la Web asignando las etiquetas que quiera y cuantas quiera, lo que resulta más fácil y más flexible que encajar la información en carpetas o categorías ya establecidas. A la colección de etiquetas generada por la anotación colaborativa de recursos se le denomina *folksonomía*. La forma más utilizada para visualizar las etiquetas de una *folksonomía* es por medio de las nubes de etiquetas, donde las etiquetas más recientes y más usadas se presentan en diferentes tamaños de fuente, donde una mayor fuente indica una mayor frecuencia de uso.

La anotación colaborativa ha tenido éxito ya que los usuarios no necesitan habilidades específicas para llevarla a cabo. Los usuarios no tienen ninguna restricción sobre las etiquetas que pueden asignar. Sin embargo, esta libertad en el uso de las etiquetas y la carencia de relaciones explícitas entre etiquetas hace que las *folksonomías* presenten problemas que afectan el proceso de recuperación de información. Por ejemplo, debido al uso de etiquetas polisémicas un usuario puede recuperar información que no es de su interés. O, debido a variaciones morfológicas de una palabra o al uso de sinónimos o de palabras compuestas, el usuario puede no encontrar la información que requiere. Es por estos pro-

blemas que las *folksonomías* hoy en día se consideran más adecuadas para la exploración de la información con el objetivo de encontrar nuevos contenidos que para la búsqueda de información en sí.

Por otra parte, los sitios Web que utilizan etiquetado colaborativo no siguen un estándar para el almacenamiento ni para la publicación de las *folksonomías*. Esto hace que estos sitios sean islas donde un usuario tiene que crear un conjunto de etiquetas para anotar sus fotos y otro independiente para anotar su blog, siendo probable que exista una relación entre las fotos que ha cargado y el blog que está escribiendo, como podría ser el caso de una descripción de un viaje con sus respectivas fotos. Permitir el intercambio de información de la *folksonomía* entre diferentes redes sociales es uno de los retos que afronta la Web 2.0 actual.

3.1. Modalidades de etiquetado

El etiquetado se puede analizar desde múltiples dimensiones como las que se han identificado en [14] y que se refieren al tipo de soporte a la anotación que ofrece la aplicación de etiquetado colaborativo, al modelo de agregación que utiliza, y a quien puede realizar la anotación.

■ **Soporte a la anotación.** Este factor distingue tres categorías: a) anotación ciega, cuando el usuario no ve las etiquetas que otros usuarios han asignado al mismo recurso; b) anotación visible, cuando el usuario ve las etiquetas que otros han asignado; y c) anotación sugerida, cuando las etiquetas se sugieren en base a etiquetas existentes del mismo usuario o a etiquetas asignadas al mismo recurso por otros usuarios, entre otros casos.

■ **Modelo de agregación.** El sistema permite o no asignar múltiples veces la misma etiqueta a un mismo recurso por diferentes usuarios. Se distingue entre el modelo de bolsa que admite etiquetas duplicadas como en Delicious, y el modelo de conjunto en el que no se admiten etiquetas duplicadas como en Flickr.

■ **Derechos de anotación.** Distingue quién etiqueta el recurso: sólo el dueño del recurso (auto-etiquetado), a quien él le haya dado permiso implícitamente por ser miembro de su red social o explícitamente (con permiso), o cualquier usuario puede etiquetar cualquier cosa (libre para todos).

Algunos sistemas no solicitan las etiquetas por medio de una interfaz de usuario convencional sino que exigen que se incluyan como elementos de código especiales que luego son procesados automáticamente. Por ejemplo, si un usuario quiere que las etiquetas asociadas a una entrada sean recopiladas junto a ésta, debe escribir las anotaciones en XML dentro del código de la entrada siguiendo un formato predefinido. Así, cuando Technorati encuentra la entrada, la analiza y, si encuentra las etiquetas

dentro del código, las carga en su sistema y las pone a disposición de los usuarios.

3.2. Recomendación de etiquetas

Los sistemas de anotación colaborativa generalmente incluyen algún tipo de mecanismo de recomendación que ayuda a los usuarios a encontrar las etiquetas más apropiadas para anotar un recurso. Etiquetar un recurso con un buen grupo de etiquetas ayuda en el descubrimiento y la recuperación de información. Una etiqueta específica ayuda a identificar fácilmente un recurso pero puede ser menos útil para que otros usuarios descubran nuevos objetos. Por el contrario, una etiqueta genérica es útil para descubrir recursos, pero no es efectiva para recuperar recursos específicos. Además, la recomendación de etiquetas es importante porque ayuda a estabilizar rápidamente el vocabulario utilizado dentro de la *folksonomía*.

ZoneTag [15] es una aplicación móvil para la captura y anotación de fotos que son cargadas a Flickr. Esta aplicación recomienda etiquetas teniendo en cuenta el contexto de la foto. El contexto para ZoneTag se compone de la localización y del tiempo. Las etiquetas recomendadas se agrupan en tres categorías que reflejan su origen: (1) etiquetas locales, que son creadas por el usuario, por miembros de su red social o por otros en su localización actual, (2) etiquetas recientes, son las usadas por el usuario las últimas veinticuatro horas y (3) todas las etiquetas. Además, ZoneTag incorpora etiquetas de otras fuentes, como lugares y eventos, basándose en la localización física.

En Flickr existe la opción *Elige de tus etiquetas* que despliega en orden alfabético las etiquetas que ya han sido utilizadas por el usuario para anotar las fotos. Para dotar a Flickr de una recomendación de etiquetas más sofisticada en [16] se presenta un enfoque basado en la co-ocurrencia de etiquetas. El sistema genera una lista de etiquetas candidatas que co-ocurren con las etiquetas que el usuario ha asignado inicialmente, luego estas etiquetas son agrupadas y ordenadas.

Además, para recomendar etiquetas se han utilizado técnicas de aprendizaje de máquina y filtrado colaborativo. En [17] se han usado técnicas de aprendizaje de máquina y procesamiento de lenguaje natural con el objetivo de encontrar páginas previamente anotadas por el usuario similares a la página que se desea anotar y luego sugerir las etiquetas de las páginas similares. El filtrado colaborativo también se ha utilizado para recomendación de etiquetas. En [18] se propone adaptar la estructura tripartita de las *folksonomías* a un espacio bidimensional y así poder utilizar las técnicas de filtrado colaborativo tradicionales.

3.3. Tendencias

Actualmente las aplicaciones de la Web 2.0

han empezado a generar funcionalidades que permiten reducir la ambigüedad de las etiquetas y proveer de estructura a la información que codifican las etiquetas. Para reducir la ambigüedad de las etiquetas que pueden tener más de un significado, Delicious permite que los usuarios asignen una descripción a cada etiqueta. De la misma manera, se permite crear grupos de etiquetas con el fin de organizarlas mejor.

Por otra parte, los usuarios de Flickr cada vez empiezan a utilizar más las llamadas etiquetas de maquina como "*flora:tree=coniferous*", "*medium:paint=oil*" y "*flickr:user=straup*". Estas etiquetas constan de un espacio de nombres, seguido de un predicado y un valor. El espacio de nombres define una clase o categoría a la cual la etiqueta pertenece y el predicado es una propiedad del espacio de nombres a la cual se le asigna el valor. Es posible realizar consultas como "quiero recuperar todas las fotos que tengan un título", utilizando una expresión como "*machine_tags*" => "*title*".

4. Anotación en la Web 3.0

En la Web 3.0 se utilizan combinaciones de técnicas de anotación de la Web Semántica y de la Web 2.0 con el objetivo de minimizar problemas de ésta como son la ambigüedad y la falta de relaciones explícitas entre las etiquetas que forman parte de una *folksonomía*. Además de ofrecer portabilidad a los datos de la Web 2.0, las tecnologías de la Web Semántica pueden mejorar el proceso de anotación ofreciendo etiquetas más acordes al dominio y al contenido de los recursos, utilizando ontologías que modelen el conocimiento del dominio en cuestión. Por último, formalizar la información de etiquetado mediante lenguajes de la Web Semántica permite que dicha información sea procesable automáticamente y que esté disponible en buscadores semánticos como Swoogle y Watson.

Los lenguajes de la Web Semántica como RDF son útiles porque favorecen el compartir información y su procesamiento automático, dos características de las que carece actualmente la Web 2.0. Además, las ontologías permiten hacer explícita y formal la semántica de las etiquetas que forman parte de las *folksonomías*, con lo cual se mejoran los procesos de recuperación de información. De la misma manera, las ontologías permiten incluir información semántica de los usuarios o la descripción de los recursos a los procesos actuales de filtrado colaborativo, lo que permite recomendaciones más precisas.

En [19] se afirma que las ontologías para la anotación son útiles porque contribuyen a:

- Sofisticar la representación del conocimiento. Una ontología de anotación puede representar robustamente entidades y relaciones que dan forma a las actividades de anotación, estableciendo explícitamente la estruc-

tura de conocimiento de los datos de la anotación y facilitando el enlace de datos de la anotación en la Web.

- Facilitar el intercambio de conocimiento. Las ontologías habilitan el intercambio de conocimiento entre diferentes usuarios y aplicaciones al proveer construcciones reutilizables. Así, una ontología de anotación puede ser compartida y utilizada para distintas actividades de anotación en diferentes plataformas.
- Hacer la información procesable automáticamente. Las ontologías y la Web Semántica en general exponen el conocimiento humano a las maquinas de tal manera que pueda ser enlazada e integrada con otras fuentes de información automáticamente.

4.1. Ontologías para etiquetados colaborativos

Algunos modelos para representar, identificar y formalizar una conceptualización de la actividad de etiquetado a nivel semántico han sido últimamente propuestos. Describimos seguidamente las ontologías más relevantes para modelar el proceso de etiquetado colaborativo, descentralizado y espontáneo.

TagOntology [20] es una ontología para el proceso de anotación que incluye conceptos para modelar el recurso a etiquetar, la etiqueta que se utiliza, y quien realiza el etiquetado. Además se incluye el concepto de polaridad con el que se puede valorar positiva o negativamente la anotación. La Tag Ontology¹¹ de Newman tiene una estructura similar a la anterior. En particular la ontología de Newman define el anotador como un agente FOAF (*Friend Of A Friend*).

La ontología SCOT¹² *Social Semantic Cloud of Tags* adiciona información a las etiquetas incluyendo su frecuencia, la frecuencia de coocurrencia con otras etiquetas e información de equivalencia con otras etiquetas. También incluye el concepto *TagCloud* (nube de etiquetas) para definir conjuntos de etiquetas.

La ontología MOAT¹³ es la única que permite definir el significado de la etiqueta. El significado se define por medio de la asignación de URIs de recursos de la Web Semántica (tales como URIs de DBPedia o GeoNames).

Por último, en [21] se ha evaluado la posibilidad de formalizar el proceso de anotación utilizando SKOS¹⁴. SKOS es un vocabulario útil para expresar sistemas de representación de conocimiento como tesauros o taxonomías sencillas. Modelar *folksonomías* por medio de SKOS presenta varias dificultades: Modelar las etiquetas de una *folksonomía* mediante un *skos:Concept* no es semánticamente correcto porque la clase *skos:Concept* debe referirse a una idea o noción abstracta, es decir, a un concepto bien definido formalmente con un significado de consenso para todos sus usuarios. Esta clase

permite declarar que un recurso es conceptual. Un símbolo de una *folksonomía* sin embargo, no llega a serlo porque cada usuario le asigna uno o varios significados.

Hay una propuesta para resolver el problema del modelado de propiedades léxicas en SKOS que garantiza la compatibilidad con OWL-DL. Se trata de extender SKOS con la posibilidad de definir etiquetas léxicas para un *skos:Concept* o un recurso. Se adiciona la clase *skos:Label* (y sus correspondientes propiedades *skos:prefLabelResource*, *skos:altLabelResource*, *skos:hiddenLabelResource*) que se puede utilizar como vínculo entre el concepto y su etiqueta. Además se introduce una nueva propiedad *skos:labelRelation* para expresar relaciones entre etiquetas, de manera que se pueda especializar para definir relaciones particulares.

4.2. Tendencias

La investigación actual y futura en temas de la Web 3.0 incluye, entre otros temas, los derivados de la carencia de una semántica explícita en las *folksonomías*. Se han presentado algunos enfoques que buscan identificar automáticamente la semántica oculta en el conjunto de etiquetas de una *folksonomía* y, posteriormente, formalizar ésta información por medio de ontologías. Estos enfoques utilizan técnicas estadísticas o de minería de datos para encontrar grupos de etiquetas, o técnicas lingüísticas y de gestión de conocimiento cuando se trata de identificar explícita y formalmente el significado de cada etiqueta. A partir de los primeros resultados obtenidos han surgido nuevos problemas como por ejemplo el de la evolución conjunta de la *folksonomía* y la ontología resultante del proceso de identificación semántica, y también el sentido contrario, es decir, si la ontología resultante cambia como se vería afectada la *folksonomía*.

5. Conclusiones

Como se muestra en este artículo, la investigación en el campo de la Web Semántica se dirige hacia la anotación de datos multimedia (textos, imágenes, vídeo, etc.), de forma que se pueda proporcionar información acerca de distintos contenidos de forma integrada, así como hacia la anotación multilingüe, con el fin de que las búsquedas de información recuperen los documentos pertinentes sea cual sea la lengua en la que estén escritos.

Por otro lado, la investigación en el campo de la Web 2.0 se dirige a la integración de datos de distintas redes sociales mediante la identificación de relaciones entre *folksonomías*, para permitir un acceso integrado a los datos a lo largo de las distintas aplicaciones, y a la anotación de etiquetas para dotarlas de significado y poder procesarlas automáticamente.

No obstante, aún existen numerosos problemas abiertos que deberán ser tratados en mayor o menor medida. Uno de estos problemas es

debido a la integración de datos de distintas fuentes. En el entorno abierto de la Web, esta integración provocará inconsistencias e información incompleta que habrá que manejar, tanto a nivel de los datos como de los modelos de dichos datos (ontologías o *folksonomías*).

Por otro lado, el incremento de información multimedia en la Web requiere dar soluciones para representar la información en los distintos documentos multimedia y para relacionar contenidos multimedia de distinta naturaleza (por ej., relacionar las imágenes y vídeos que aparecen en un documento y su texto). Estas soluciones son abordadas parcialmente en la Web Semántica y la Web 2.0, pero una solución general del problema requiere una aproximación híbrida.

Otra situación relevante es la del manejo del ciclo de vida de las anotaciones (evolución de las anotaciones, gestión de cambios, diferentes versiones, etc.) que en muchos casos está relacionada con la procedencia de dichas anotaciones y de los datos que anotan.

Además, hay que destacar que las áreas de investigación tratadas en este documento son muy dinámicas y se espera que su evolución traiga nuevos avances tecnológicos. Actualmente las herramientas de anotación se dividen en herramientas orientadas a usuarios expertos, no usables por un público en general, y herramientas de dominio específico, difíciles de integrar con otras.

En el futuro se espera pasar de las herramientas actuales de anotación, manuales o semiautomáticas, a herramientas de anotación automáticas en las que el papel del usuario sea el de supervisar y evaluar la calidad de las anotaciones. Por otro lado, se espera en el futuro que las herramientas de anotación estén parcial o totalmente integradas con herramientas y sistemas de uso general, tales como herramientas de comunicación personal (chats, mensajería instantánea, correo electrónico, etc.), herramientas de escritorio, herramientas ofimáticas, etc.

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el proyecto CENIT "España Virtual", subvencionado por el CDTI dentro del programa Ingenio 2010. Muchas gracias a los siguientes miembros del *Ontology Engineering Group* por sus contribuciones a este artículo: Guadalupe Aguado de Cea, Inmaculada Álvarez de Mon Rego, Óscar Corcho y Antonio Pareja Lora.

Referencias

- [1] G. Aguado de Cea, I. Álvarez de Mon, A. Pareja-Lora. *Una visión interdisciplinar de la anotación semántica*. Amparo Alcina (ed.). Terminología y Sociedad del conocimiento. Peter Lang, Berna 2008.
- [2] T. Bray, J. Paoli, C. Sperberg. *Extensible Markup Language (XML) 1.0*. W3C Recommendation, 1998. <<http://www.w3.org/TR/REC-xml>>.
- [3] F. Manola, E. Miller. *RDF Primer W3C Recommendation*. 10 Febrero 2004. <<http://www.w3.org/TR/rdf-primer/>>.
- [4] D.L. McGuinness, F. van Harmelen. *OWL: Web Ontology Language Overview*. (W3C Recommendation, 10 February 2004). <<http://www.w3.org/TR/owl-features/>>.
- [5] D. Shea. *Who Cares about Semantics Anyway?* <http://www.mezzoblue.com/archives/2005/05/30/who_cares_ab/>. [Accedido en octubre, 2008]
- [6] R. Studer, VR. Benjamins, D. Fensel. Knowledge Engineering: Principles and Methods. *IEEE Transactions on Data and Knowledge Engineering* 25(1-2): pp.161-197. 1998.
- [7] O. Corcho. Ontology based document annotation: trends and open research problems. *International Journal of Metadata, Semantics and Ontologies* 1(1), 2005.
- [8] S. Handschuh, S. Staab. Annotation for the Semantic Web. IOS Press, 2003.
- [9] S. Bechhofer, L. Carr, C. Goble, S. Kampa, T. Miles-Board. The Semantics of Semantic Annotation. ODBASE: *First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, 2002. Irvine, California.
- [10] M. Uschold. Where are the Semantics in the Semantic Web? *AI Magazine*, 24 (3), 2003: 25-36. Menlo Park (California): American Association for Artificial Intelligence (AAAI).
- [11] Kosmas Petridis, Dionysios Anastasopoulos, Carsten Saathoff, Norman Timmermann, Yiannis Kompatsiaris, Steffen Staab. M-OntoMat-Annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Features *KES 2006*.
- [12] Ajay Chakravarthy, Vitaveska Lanfranchi, Daniela Petrelli, Fabio Ciravegna. Semantic Enrichment of Multimedia Document. *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*. 2006.
- [13] T. Declerck, J. Contreras, O. Corcho, C. Crispi. Text-based Semantic Annotation Service for Multimedia Content in the Esperanto Project. *European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies (EWIMT2004)*. London, United Kingdom.
- [14] Cameron Marlow, Mor Naaman, Danah Boyd, Marc Davis. HT06, tagging paper, taxonomy, Flickr, academic article, to read. *HYPERTEXT '06: 17th conference on Hypertext and hypermedia*, pp. 31-40. ACM Press.
- [15] Morgan Ames, Mor Naaman. Why We Tag: Motivations for Annotation in Mobile and Online Media. *Actas de la CHI 2007*, San Jose, CA, USA.
- [16] Börkur Sigurbjörnsson, Roelof van Zwol. Flickr tag recommendation based on collective knowledge. *17th International Conference on World Wide Web (WWW 2008)* pp. 327-336.
- [17] P. Basile, D. Gendarmi, F. Lanubile, G. Semeraro. Recommending Smart Tags in a Social Bookmarking System. *Proc. of International Workshop on Bridging the Gap between Semantic Web and Web 2.0. Actas de la 4th European Semantic Web Conference (ESWC'07)*.
- [18] Robert Jäschke, Leandro Marinho, Andreas Hotho, Lars Schmidt-Thieme, Gerd Stumme. Tag

Recommendations in Folksonomies. Knowledge Discovery in Databases: *PKDD 2007*.

[19] H.L. Kim, S. Scerri, J.G. Breslin, S. Decker, H.G. Kim. The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies. *The 8th International Conference on Dublin Core and Metadata Applications*, Berlin, Germany, Septiembre 2008.

[20] Thomas Gruber. *TagOntology, a way to agree on the semantics of tagging data*, 2005.

[21] Emilio Rubiera, Sergio Fernández. *Morfeo Project, Entregable 5.3.1*. Integración de folksonomías y ontologías, 2007.

Notas

- ¹ <<http://www.esperanto.net/>>.
- ² <<http://www.acemedia.org/>>.
- ³ <<http://www.schema-ist.org/>>.
- ⁴ <<http://www.mindswap.org/>>.
- ⁵ <<http://annotation.semanticweb.org/>>.
- ⁶ <<http://www.x-media-project.org/>>.
- ⁷ <<http://www.acemedia.org/>>.
- ⁸ <<http://lands.let.kun.nl/TSPublic/MUMIS/>>.
- ⁹ <<http://www.w3.org/TR/swbp-image-annotation/>>.
- ¹⁰ <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>.
- ¹¹ <<http://www.holygoat.co.uk/projects/tags/>>.
- ¹² <<http://scot-project.org/>>.
- ¹³ <<http://moat-project.org/>>.
- ¹⁴ <<http://www.w3.org/2004/02/skos/>>.

Ossi Nykänen

Laboratorio de Hipermedia, Universidad de
Tecnología de Tampere (Finlandia)

<ossi.nykanen@tut.fi>

Web Semántica para espacios de conocimiento evolutivos entre iguales

1. Introducción

Las visiones de un espacio de conocimiento unificado han prevalecido desde el comienzo de la evolución sociocultural. Intuitivamente, un espacio de conocimiento moderno denota una aplicación de la gestión de la información y la comunicación a gran escala que va más allá de las responsabilidades de los sistemas de información orientados a documentos. En última instancia, el objetivo es dar soporte a las interacciones a nivel de conocimiento. Sin embargo, hasta hace poco, la falta de tecnología apropiada (almacenamiento eficaz y duplicidad de datos, indexación y búsqueda, distribución y acceso remoto, y cálculo automático) planteaba el principal obstáculo a la hora de conseguir este objetivo.

Hoy en día, una forma práctica para entender la amalgama de las distintas técnicas de computación y comunicación es percibir las a través del concepto unificado de tecnologías de Web. De hecho, la Web puede ser incluso percibida *per se* como el espacio de conocimiento interoperable, o al menos como la tecnología que lo permite.

Sin embargo, cuando se supone que los ordenadores trabajan para el beneficio de los humanos, las posibilidades de alcanzar un espacio de conocimiento unificado se incrementan introduciendo semánticas que puedan ser calculadas automáticamente. La Web Semántica del World Wide Web Consortium (W3C)¹ tiene como objetivo abordar esta necesidad. Proporciona un marco común que permite compartir y reutilizar datos traspasando los límites de aplicaciones, compañías y comunidades. Sin embargo, no todos los sistemas modernos comparten las asunciones de las aplicaciones Web de cada día. En particular, las emergentes redes específicas y los sistemas entre iguales (P2P, *Peer to Peer*) retan a las aplicaciones Web cliente-servidor en parte en diseño y en parte en el puro volumen de datos [1].

En este artículo, resumimos los conceptos básicos de la gestión del conocimiento y de los sistemas de compartición de datos. Posteriormente, reflejamos y analizamos los retos técnicos de la novedosa información entre iguales (P2P) y la gestión de la evolución de la información vistos a través de la Web Semántica. Un buen marco para este análisis lo proporciona un reciente proyecto de Red de Excelencia *Open Philosophies for Associative Autopoietic Digital Ecosystems* (OPAALS). Entre otras cosas, el proyecto tiene como

Traducción: Mercedes Montes Rubio (Grupo de Trabajo de Lengua e Informática de ATI).

Resumen: la Web Semántica proporciona la tecnología fundamental que permite la realización de aplicaciones que exploten los datos con semánticas legibles automáticamente. Se introduce un área de aplicación particularmente interesante por medio del concepto de espacios de conocimiento. En este artículo destacamos los conceptos básicos de la gestión del conocimiento y los sistemas de compartición de datos. Posteriormente reflejamos y analizamos los retos técnicos de la novedosa información entre iguales (*Peer-to-peer*, P2P) y la gestión de la evolución de la información a través de la óptica de la Web Semántica. Un reciente proyecto de Red de Excelencia cuyo objetivo es la realización de su particular versión de un Espacio de Conocimiento Abierto, proporciona un buen marco para este análisis. La principal aportación de este artículo consiste en esbozar los tres dominios complementarios del conocimiento y los sistemas de compartición de datos, Web Semántica, espacios de conocimiento, así como discutir el modelado P2P y la evolución en este contexto.

Palabras clave: espacio de conocimiento abierto, OKS, OPAALS, OWL, procesos entre iguales, P2P, RDF, semántica legible automáticamente, Web Semántica, W3C.

Autor

Ossi Nykänen, nacido en 1972, recibió su Máster en Informática por la Universidad de Jyväskylä (Finlandia) en 1997. Es Ingeniero en Matemáticas Aplicadas (2001) y Doctor en Matemáticas (2007) por la Universidad de Tecnología de Tampere (TUT). Trabaja como investigador especial en el Laboratorio de Hipermedia de la TUT. Es responsable de Estudios Avanzados en Hipermedia en la Universidad y contribuye en varios proyectos de investigación a nivel nacional e internacional, desempeñando distintas funciones. Es también Director de la Oficina de la W3C (World Wide Web Consortium) en Finlandia. Sus temas de investigación incluyen aplicaciones de lógica en informática, inferencia mecánica y aprendizaje automático, educación y matemáticas apoyadas por ordenador, visualización y modelado de la información y las aplicaciones (Web) relacionadas. Este trabajo tiene como resultado aplicaciones como por ejemplo documentos estructurados, gestión del conocimiento y sistemas inteligentes. Ha sido autor de unas 30 publicaciones científicas y varios libros sobre estas áreas.

El Laboratorio de Hipermedia de la Universidad de Tecnología de Tampere (TUT) opera como parte del Instituto de Matemáticas y fue fundado en 1994. Su equipo de investigación multidisciplinar estudia y desarrolla entornos de aprendizaje basados en Web, modelado y producción de contenido digital, tecnologías Web y de Web Semántica, el estudio basado en simulador y los métodos de evaluación de enseñanza y aprendizaje basados en Web. El Laboratorio participa tanto en investigaciones a nivel nacional como internacional, I+D y proyectos de excelencia de red (por ej. Aprendizaje a distancia en Redes Multimedia, Entornos de Aprendizaje Abierto, Entorno de Aprendizaje basado en Simuladores de Máquinas Forestales, DBE, OPAALS) así como comunidades expertas y científicas (por ej. PEG, EDeAN). Además el Laboratorio de Hipermedia organiza formación de empresa, siendo el responsable de la formación en hipermedia en la TUT (junto con el Instituto de Sistemas de Software). Realiza también las funciones de Oficina del W3C en Finlandia.

<<http://matriisi.ee.tut.fi/hypermedia/en/>>.

objetivo la realización de su particular versión de un espacio de conocimiento, el *Open Knowledge Space* (OKS). Presentamos un esbozo de las tecnologías de Web Semántica en este contexto de investigación, resumiendo las técnicas básicas y los componentes técnicos de aplicaciones típicas. Posteriormente introducimos y discutimos dos novedosos retos de sistemas emergentes, a saber, información semántica en P2P y evolución de la información en P2P.

La principal aportación de este artículo recae en resumir los tres dominios complementa-

rios del conocimiento y de los sistemas de compartición de datos, Web Semántica, el OPAALS OKS, así como discutir el modelado P2P y la evolución en este contexto. Debido a las restricciones de espacio, simplemente asumimos muchas de las definiciones relacionadas. Para discusiones, aplicaciones y referencias más técnicas, rogamos consulten las referencias [2] y [3].

El resto del artículo se organiza de la siguiente forma: la **sección 2** resume los temas relativos al conocimiento y a los sistemas de compartición de datos e introduce brevemente

te el dominio concreto OPAALS OKS para este trabajo; la **sección 3** caracteriza las tecnologías de Web Semántica del W3C y los tópicos relacionados; la **sección 4** analiza dos retos identificados, la información en sistemas P2P y su evolución. Finalmente, la **sección 5** concluye el artículo.

2. Hacia un Espacio de Conocimiento Abierto (Open Knowledge Space)

Desde el punto de vista contemporáneo, el conocimiento no es algo que simplemente pueda ser almacenado y transmitido, más bien se produce y se consume en colaboración, utilizando los sistemas de información apropiados.

2.1. Gestión del Conocimiento a través de los Sistemas de Compartición de Datos.

En una organización de creación de conocimiento, se pueden reconocer los siguientes procesos y roles clave en la gestión de recursos del conocimiento [4]:

- Externalización (transformar el conocimiento de tácito a explícito, por ej. comunicación reflectiva escrita en pequeños grupos)
- Combinación (transformar el conocimiento de explícito a explícito, por ej. integrando la producción de los grupos a nivel de organización).
- Internalización (transformar el conocimiento de explícito a tácito, por ej. explorando y estudiando la memoria de la organización)
- Socialización (transformando el conocimiento de tácito a tácito, por ej. comunicación existencial "cara a cara")

Los retos prácticos, no solamente en parte técnicos, de la gestión del conocimiento deben ser también reconocidos. Cuando una actividad de gestión del conocimiento falla, no es debido necesariamente a problemas teóricos en la gestión del conocimiento abstracto, sino que más bien puede ser debido a la no comprensión o captura del proceso existente (por ej. motivación humana y política de participación) y a las pequeñas deficiencias en los sistemas de información diarios (por ej. la falta de conocimiento explícito, problemas de gestión). Parece que sin un ejército de voluntariosas y habilidosas (quizá artificiales) secretarías personales no es posible evitar la tediosa gestión de información.

También, la compartición es un asunto a tener en cuenta. Cuando los intereses y los valores difieren, un modelo de colaboración demasiado idealista podría no funcionar. Considerando los retos prácticos de la gestión de la información podemos analizar (simplificando) el escenario además en términos de Sistemas de Compartición de Datos (*Data-Sharing Systems*, DSS). Según Smith, Seligman y Swarup [5], las actividades fundamentales de los DSS incluyen:

- Descubrimiento (por ej. aprender que hay "ahí").
- Acceso (por ej. ser capaces de recuperar porciones de información interesantes).
- Comprensión (por ej. ser capaces de traducir/contextualizar los datos recuperados en algo localmente aplicable).
- Creación de políticas (por ej. tener claras las normas de los términos de uso y dar y recibir crédito).

De nuevo, mientras las propias herramientas informáticas son obviamente una necesidad, las tecnologías constituyen sólo una parte del panorama general. En particular, mientras que quienes abogan por las "aplicaciones semánticas" típicamente destacan por ej. búsqueda semántica (y sus aplicaciones relacionadas más sofisticadas como los sistemas de recomendación), la búsqueda semántica por sí sola no proporciona un "espacio de conocimiento". El análisis de los sistemas de compartición de datos actuales sugiere que los factores sociales y la creación de políticas de organización desempeñan un papel decisivo en aplicaciones de éxito.

Por lo tanto, debido a sus fundamentos basados en la compartición de información, la gestión del conocimiento, y consecuentemente los espacios de conocimiento, deben identificar y apoyar activamente las actividades clave tanto de la gestión del conocimiento como de los sistemas de compartición de datos. Como de costumbre, los grandes retos recaen en la complejidad del sistema socio-técnico.

2.2. OPAALS OKS: Un enfoque de investigación

En 2006 un proyecto de Red de Excelencia Europea (NoE, *European Network of Excellence project*) *Open Philosophies for Associative Autopoietic Digital Ecosystems* (OPAALS)³ realizó un esfuerzo en la definición de su versión del espacio de conocimiento unificado, el *Open Knowledge Space* (OKS). El OKS está previsto como la memoria a largo plazo de la comunidad OPAALS (acceso, recuperación, reutilización, colaboración, visibilidad, diseminación, identidad de comunidad). En sentido abstracto, OKS proporciona el modelo socio-técnico que se utilizará para crear, gestionar y dejar disponible el conocimiento para la comunidad OPAALS. Sin embargo, en lugar de presentar el OKS como un espacio de conocimiento ideal, debería percibirse como un escenario del mundo real para el análisis de las tecnologías de Web Semántica y sus retos relacionados.

De acuerdo con su descripción de trabajo, los dos objetivos primordiales de OPAALS son construir una comunidad de investigación interdisciplinar en el área emergente de los ecosistemas digitales (*Digital Ecosystems*, DE) y desarrollar una base teórica integrada para investigación de ecosistemas digitales

abarcando tres dominios disciplinarios totalmente diferentes: sociología, informática y ciencias naturales. Estos dos objetivos juntos darán como resultado una Red de Excelencia global en ecosistemas digitales.

La principal reivindicación de OPAALS es que para alcanzar ecosistemas de negocio digital sostenibles en PYMEs y componentes de software adecuados necesitamos entender en profundidad los procesos de colaboración y las tecnologías informáticas que sostienen la continua creación, formalización y compartición del conocimiento en forma de modelos de negocio, infraestructura de software para transacciones de comercio electrónico y nuevos lenguajes formales y semiformales. La estrategia de *networking* (establecimiento de contactos) se basa en el desarrollo del OKS, un espacio *online* en el que los investigadores pueden libre y fácilmente aportar e intercambiar conocimiento así como desarrollar funciones de apoyo a sus investigaciones.

El consorcio está coordinado por el *London School of Economics and Political Sciences* (Reino Unido) e incluye otros socios colaboradores como *T6 Società Cooperativa* (Italia), *Salzburg University of Applied Sciences* (Austria), *The University of Surrey* (Reino Unido), *Waterford Institute of Technology* (Irlanda), *Tampere University of Technology* (Finlandia), *TechIDEAS Asesores Tecnológicos* (España), *The University of Dundee* (Reino Unido), *University of Limerick* (Reino Unido), *Create-Net* (Italia), *Universität Kassel* (Alemania), *Indian Institute of Technology Kanpur* (India), *Instituto de Pesquisas em Tecnologia da Informação* (Brasil), *Kigali Institute for Science and Technology* (Ruanda) y *University of Central England* (Reino Unido).

El consorcio en sí mismo ya ha mostrado de muchas formas un OKS orientado a investigación nada trivial. Sin embargo, la realización de un ecosistema de conocimiento autosostenible todavía supone un reto importante.

El OPAALS OKS difiere de los espacios de conocimiento que siguen la tendencia principal en dos decisiones de diseño que merece la pena identificar aquí. En primer lugar, el núcleo de su arquitectura sigue las convenciones de los sistemas P2P. En segundo lugar, la evolución es adoptada de distintas formas, desde tendencias de actividades hasta lenguajes de representación, un marco de trabajo evolutivo por tanto para que se reconozcan las necesidades del lenguaje. Resulta que estas dos características del OKS están relacionadas.

3. Web Semántica para desarrolladores

En dos palabras, la Web Semántica W3C es un esfuerzo de colaboración dirigido por

W3C con participación de un gran número de investigadores y socios industriales. Su fundamento estándar es establecido por el *Resource Description Framework* (RDF) y sus especificaciones relacionadas.

3.1. Motivación y aplicaciones

Es fácil encontrar motivación para la Web Semántica. Existen grandes cantidades de datos en la Web, en las redes internas de organizaciones, así como en otras bases de datos. Estos datos incluyen información sobre documentos y otros expedientes, artefactos de software, objetos físicos, gente y estructuras conceptuales.

Sin embargo, los datos no están siempre formateados ni descritos semánticamente para poder ser integrados fácilmente. Los estándares de la Web Semántica tienen como objetivo resolver el problema de la descripción semántica y la integración introduciendo un modelo de datos semántico común, una gramática interoperativa para la publicación de descripciones semánticas, medios para el rediseño de vocabularios semánticos reutilizables y las técnicas comunes relacionadas. Además de "mapear" e integrar las aplicaciones existentes, el modelado común permite diseñar nuevas aplicaciones de manera que se puedan aprovechar los beneficios de la integración semántica.

Las áreas de aplicación citadas más a menudo para aplicaciones de Web Semántica comprenden: ingeniería del software, configuración, medicina, bibliotecas digitales y sistemas de información basados en Web. Las áreas más destacadas de trabajo incluyen aplicaciones para portales web, colecciones multimedia, webs corporativas, documentación de diseño, agentes y servicios, así como computación ubicua. Las visiones de las aplicaciones modernas varían desde la publicación de reglas de negocio hasta el descubrimiento del potencial de colaboraciones empresariales, pasando por agregar los álbumes de fotos de la familia.

3.2. Resumen de las técnicas básicas

Mientras que la Web Semántica se presta a distintas aplicaciones, sus técnicas básicas son mejor percibidas como tecnologías facilitadoras.

Considerando el modelado, las propiedades clave de las tecnologías de Web Semántica se pueden resumir como se indica a continuación:

1. La postura de modelado básica es que el mundo se puede modelar como cosas con relaciones con otras cosas (por ej. Un "empleado" puede "tener" un "nombre de puesto de trabajo". Las cosas que se referencian se llaman recursos y se pueden identificar con referencias URI (*Uniform Resource Identifier*).
2. Los recursos pueden referenciarse utilizando múltiples (diferentes) referencias URI.

Sin embargo, cada referencia URI identifica exactamente una cosa (esta cosa puede ser por ejemplo una entidad portafolio que contiene más de un recurso).

3. La información semántica se establece mediante declaraciones. Cada declaración incluye tres partes (y es por lo tanto a menudo llamada un "triple"): sujeto ("la cosa descrita"), predicado ("la propiedad de la que habla la descripción") y objeto ("el valor de la propiedad para la cosa que es descrita").

4. Además, mediante referencias URI, las "cosas" sujeto y objeto en declaraciones pueden referenciarse localmente de forma anónima (con identificadores locales o temporales). Es más, los valores literales (cadenas con tipos de datos opcionales e identificadores de lenguaje) pueden utilizarse como objetos en declaraciones.

5. El marco de modelado sigue el principio de diseño "cualquiera puede realizar declaraciones sobre cualquier recurso".

6. Mientras que las declaraciones arbitrarias son posibles, el marco de modelado es insensible a la multiplicidad de declaraciones idénticas.

7. Las asociaciones semánticas se producen por la unificación directa y deducida de identificadores (esta es la "Web Semántica"). En particular, se anima a la reutilización de los identificadores de predicados utilizados comúnmente. Son posibles las declaraciones sobre declaraciones por medio del mecanismo de reificación⁴.

8. La interpretación lógica del mero modelo de datos RDF es bastante mínima. En particular, un conjunto de declaraciones (también llamado gráfico RDF) incluye cada uno de sus subconjuntos de declaraciones base y un conjunto de sentencias donde las referencias URI de sujetos y objetos han sido reemplazados por identificadores temporales de objetos anónimos.

9. Cuando es necesario, se pueden utilizar implicaciones lógicas más ricas mediante la introducción de nuevas capas semánticas de descripciones. En particular, se pueden utilizar los vocabularios recomendados del esquema RDF y del Lenguaje Ontológico de Web OWL (*Web Ontology Language*) como ontologías de representación del conocimiento, introduciendo nuevos vocabularios de forma estándar.

10. No toda aplicación RDF tiene como objetivo seguir fielmente las semánticas RDF formales. Por ejemplo, un caso de utilización trivial de RDF es simplemente la comunicación sintáctica en términos de declaraciones, sin calcular en realidad las clausuras semánticas de los gráficos RDF, etc. (Considere por ejemplo aplicaciones de sindicación simples).

11. Además de los componentes ontológicos, las aplicaciones pueden incluir componentes de reglas (típicamente en forma de cláusulas de Horn). Sin embargo, en el momento de redactar este artículo, los esfuerzos por estandarizar una norma de lenguaje de Web Semántica están todavía por desarrollar.

12. Se encuentran disponibles prácticas razonablemente bien probadas, documentando tanto los patrones de buen diseño como las soluciones provisionales (*workarounds*) para trabajar con las limitaciones de las recomendaciones (por ej. representar clases como valores de propiedad para superar las limitaciones intrínsecas del OWL DL).

13. Se pueden poner en práctica aplicaciones de Web Semántica Abstracta siguiendo el patrón de diseño básico de los sistemas de bases de datos: Una base de datos lógica de Gráficos RDF es interrogada utilizando el lenguaje de consulta estandarizado SPARQL (*SPARQL Protocol and RDF Query Language*). En un principio, se pueden ampliar los procesadores de consultas para reconocer ontologías de representación del conocimiento (y reglas), proporcionando un apoyo de razonamiento básico (p. ej. incluyendo triples virtuales deducidos por vía del razonamiento transitivo).

14. Para añadir nuevas declaraciones lógicas se utilizan adaptadores o herramientas de autorización específicas de aplicación. En particular, el lenguaje de interrogación estándar no incluye (aún) la equivalencia de la funcionalidad SQL INSERT INTO.

15. Las consideraciones de versionado y seguridad se tratan normalmente fuera del núcleo RDF lógico, a menudo con soluciones de herramientas específicas que se aprovechan de los estándares y herramientas disponibles cuando éstos se pueden aplicar. (por ej. Seguridad XML y WebDAV).

Existen múltiples formas de representar o serializar declaraciones lógicas. La familia de especificaciones RDF fomenta la utilización de RDF/XML o la sintaxis de serialización Turtle. Esto permite transmitir gráficos RDF en términos de documentos de texto.

3.3. Componentes del desarrollo de aplicaciones

Desde el punto de vista de los desarrolladores de aplicaciones, los estándares de la Web Semántica significan interfaces y soporte de herramientas integradas. Las categorías de herramientas básicas para el desarrollo de aplicaciones incluyen:

- Transformadores y adaptadores (por

ej. para mapear bases de datos relacionales o documentos legibles automáticamente al modelo de datos RDF y viceversa).

- Analizadores sintácticos y repositorios de datos (por ej. servicios de bases de datos con APIs lógicas para E/S básicas, autenticación, etc.).
- Editores (por ej. herramientas específicas de aplicación para usuarios humanos).
- Componentes de control y validación de la calidad de los datos: (por ej. para excluir datos erróneos o potencialmente dañinos).
- Componentes de consulta (por ej. filtrado, transformación y razonamiento con los datos).
- Acceso programático a repositorios de datos (por ej. para desarrollar cálculo semántico más allá de los componentes del editor y de consulta).
- Componentes específicos de aplicación (por ej. lógica de aplicación, quizá basada en una arquitectura en *pipeline*⁵ con una interfaz de usuario gráfica).

Nótese que mientras que los estándares incluyen el muy necesario lenguaje de consulta SPARQL, la, a menudo citada "búsqueda semántica", debe ser implementada en la mayoría de los casos como una aplicación de Web Semántica. Esto es debido a que son necesarios varios componentes de software no trivial (conceptos de apoyo al usuario final, cálculo semántico eficaz, temas de calidad de los datos...).

Además, las aplicaciones no sólo tratan datos lógicos. En principio, el "cálculo semántico" puede tener en cuenta cualquier dato inducido (por ej. clasificaciones debidas a análisis estadísticos) o cálculos arbitrarios (del exterior, por ej. del área de las descripciones lógicas). Debido también al uso extensivo de adaptadores y transformadores, las aplicaciones se asocian normalmente con un amplio espectro de entradas de datos, lo que proporciona importantes retos con respecto al control de calidad de los datos. Sin embargo, mientras que la utilización por ejemplo de *proxies* puede solucionar ciertos problemas, la validación puede requerir en última instancia la moderación por parte de una persona.

Para optimizar el caro proceso de escribir descripciones semánticas, podemos identificar tres principales fuentes de datos:

1. Descripciones realizadas por personas (por ej. las decisiones y las clasificaciones realizadas por usuarios finales o expertos).
2. Descripciones contextuales introducidas por máquinas (por ej. la fecha, la herramienta de autoría, el tipo MIME, así como el autor de un documento, recuperados de forma automática por el editor de software).
3. Repositorios de descripción de datos y software preexistente (por ej. archivos adaptados de propósito general y otras aplicacio-

nes interoperables).

Es de sospechar que la motivación de la aplicación de herramientas de Web Semántica radique en su integración con los datos existentes, herramientas y aplicaciones. Técnicamente, esto requiere a menudo la implementación de adaptadores en las arquitecturas en *pipeline* específicas de aplicación. Por otra parte, los usuarios necesitan procesos bien establecidos, políticas, herramientas e instrucciones, por no mencionar una comunidad activa.

3.4. Arquitectura Web y temas relacionados

Las especificaciones de la Web Semántica se basan en una visión del W3C de la Arquitectura Web [6]. La buena noticia es que la integración Web se realiza sin problemas. En particular, cuando se utilizan los nombres URL para identificar cosas, la referenciación existente, la resolución de nombres y los mecanismos de recuperación de representaciones son aplicables directamente cuando se necesitan. Esto también significa que los procesos básicos de la aplicación y el modelado semánticos se entienden bien y se implementan razonablemente.

La noticia de algún modo preocupante es que si el papel por ejemplo de los nombres globales es cuestionado, como ocurre en algunas aplicaciones P2P, la extensión en el uso de tecnologías relacionadas con Arquitectura Web podría suponer algún tipo de replanteamiento. Esto no quiere decir de ninguna manera que la Arquitectura Web no sea aceptable, sino que los estándares de Web Semántica deben aplicarse con cuidado, de manera que los sistemas resultantes no se limiten a sistemas cliente-servidor con nombres globales.

En particular, aunque los términos "cliente" y "servidor" en principio se refieran solamente al papel desempeñado por un programa para una conexión en particular en las especificaciones Web abstractas, las aplicaciones prácticas son a menudo diseñadas *per se* como sistemas cliente-servidor. Las razones son en parte técnicas y en parte económicas.

3.5. La Web Semántica en sentido estricto y en sentido amplio

Considerando aplicaciones de conocimiento, es importante darse cuenta de que en el sentido estricto (por ej. los estándares y las herramientas básicas) la Web Semántica facilita o contribuye al desarrollo de la gestión del conocimiento o de sistemas de compartición de datos. En cambio, los estándares de la Web Semántica como tales no establecen directamente ni describen en su totalidad el espectro de aplicaciones y actividades socio-técnicas requeridas. Esto es perfectamente entendible, considerando el ámbito del establecimiento de los estándares universales.

Sin embargo, cuando se utiliza el término Web Semántica en sentido amplio (por ej. incluyendo las aplicaciones potenciales), existe el peligro de asumir que la gestión de información práctica, gestión del conocimiento y la creación de políticas podrían "resolverse" simplemente mediante (la migración a) una Web Semántica. Por desgracia, mientras que la Web Semántica contribuye de forma significativa al entendimiento del cálculo semántico (y en su núcleo se orienta a poco más que a eso), el desarrollo de aplicaciones puede aún requerir esfuerzos considerables.

4. El conocimiento en sistemas evolutivos entre iguales

Para un examen más detallado seleccionemos dos retos interesantes para la Web Semántica, como los motivados por el marco de investigación OPAALS OKS. El primero de los retos lo proporciona la arquitectura distribuida en tiempo de ejecución P2P de OKS y el segundo la evolución de los sistemas de información.

4.1. Información en sistemas entre iguales

Por definición, un sistema P2P es un organismo autoorganizado de entidades autónomas e iguales (*peers*) cuyo objetivo es la utilización compartida de recursos distribuidos en un entorno de red evitando los servicios centrales [1].

El P2P difiere del cliente-servidor en un aspecto crucial: los servicios y los datos ya no son direccionados a través de la dirección del servidor (ubicación) sino a través de los servicios y datos en sí mismos (identidad o contenido). Sin embargo, en lugar de copiar los recursos reales al sistema P2P, se puede utilizar almacenamiento indirecto vía referencias de localizador. En este caso, el sistema P2P simplemente proporciona un índice distribuido, que referencia recursos "externos".

Es más, debido a la dinámica asumida de la red P2P, la disponibilidad de los nodos individuales no está garantizada. En su lugar, el objetivo típico de la red se centra en garantizar los servicios a nivel de la red P2P en su totalidad. Por tanto, el poner nombre a los recursos con respecto a nodos individuales P2P no es recomendable; en su lugar los recursos se identifican con identificadores P2P especiales, como por ejemplo claves *hash*.

Además, una función de búsqueda global podría no estar garantizada. Por tanto, mientras que siempre sería posible realizar una búsqueda de un recurso individual basado en su identificador P2P asignado, la búsqueda de todos los datos podría no estar garantizada.

El papel de las tecnologías de Web Semántica depende ahora de los servicios y de la transparencia del sistema P2P. Merece la pena destacar ahora los tres tipos básicos de sistemas

P2P (o niveles de responsabilidad de red P2P):

1. La red P2P aparece simplemente como un sistema de búsqueda y compartición de ficheros distribuidos (recuperando recursos de entre los identificadores P2P conocidos) que no tiene conocimiento de las capas tecnológicas que se encuentran por encima.
2. La red P2P también implementa una búsqueda por palabra clave y es capaz de devolver algunos recursos asociados con dicha palabra clave, basándose en un conjunto restringido de palabras clave.
3. La red P2P implementa una búsqueda basada en un esquema semántico de algún tipo.

Nótese que el problema obvio de la mera funcionalidad de búsqueda (tipo 1 descrito más arriba) es que mientras que se proporciona una distribución eficaz de los recursos de P2P, se necesita mantener un índice externo de identificadores P2P para recuperar un recurso en concreto.

Además de los retos de implementar una búsqueda semántica factible, P2P tiene implicaciones en el modelado semántico. En concreto, el nombrar los recursos de P2P se convierte en un hecho crucial ya que nombrar recursos con URLs sin pensarlo bien puede ser problemático, al menos en el peor de los casos: en un sistema distribuido la utilización de direccionamiento global o nombres de dominio podría considerarse que introduce un riesgo de punto único de fallo (técnico o socioeconómico).

Cuando se trata de evitar puntos únicos de fallo (*single points of failure*)⁶ en descripciones semánticas, una solución potencial es describir recursos indirectamente, por medio de propiedades que caractericen suficientemente (por ej. identificando un documento por el asunto, autor/es, título, editor y fecha de publicación, en lugar de darle únicamente un nombre explícito de URI o de identificador P2P). En este caso, las descripciones semánticas podrían incluir un componente común de información de metadatos, suficiente para identificar recursos y las distintas versiones de éstos. Ahora, dependiendo de los servicios de búsqueda proporcionados por la red P2P (tipos 2 y 3 descritos anteriormente), pueden buscarse los recursos bien utilizando un conjunto fijo de palabras clave de metadatos o bien mediante búsquedas semánticas más flexibles.

Se ha planteado otro nivel de retos por el papel que desempeñan los esquemas semánticos. Los sistemas basados en palabras clave y los basados en esquema (tipos 2 y 3 mencionados anteriormente) introducen un dominio sutil pero concreto para puntos únicos de fallo: el requerimiento de términos y vocabulario semántico común en descripciones y

metadatos (por ej. los nombres de los prediados semánticos, elementos de metadatos o terminología). Incluso si los localizadores específicos de dominio pudieran no ser técnicamente necesarios para la recuperación de datos, podrían surgir problemas debido a inconsistencias y razones normativas, al menos conceptualmente. Una posible solución a este problema consiste en el reconocimiento de niveles de información y evolución del conocimiento en un sistema P2P.

4.2. Evolución en sistemas de información

Todos los sistemas de información P2P se someten a algunos cambios durante su periodo de vida. De forma similar, todos los sistemas de información son vulnerables a los cambios que no fueron anticipados en sus diseños.

La hipótesis de trabajo abstracto en muchos de los sistemas de conocimiento o información a gran escala, incluyendo OPAALS OKS, es que su utilización natural orientada al objetivo produce como resultado el desarrollo útil y gradual del sistema. Por tanto, el desarrollo de sistemas de información no tiene lugar en un simple paso específico de implementación y diseño.

En su lugar, una vez que se especifica y diseña un sistema facilitador o de "arranque" (*bootstrapping*) adecuado, el sistema de información es desarrollado (y se desarrolla en sí mismo) de acuerdo con su uso real. Este desarrollo incremental tiene lugar en un proceso que puede describirse empleando los términos *evolutivo*, *autoorganizado* o *autosostenible* (es decir *autopoiesico*, -ver la nota 2 de este artículo- [7]).

Mientras que la fuerza motriz de sistemas físicos podría ser capturada en términos de las leyes de la física (fenómenos físicos), o energía, la fuerza motriz de un sistema de conocimiento podría ser capturada en términos de las leyes de los sistemas sociales, socioeconómicos, etc. (comportamiento de la gente) o de la información.

Para simplificar la discusión, es útil centrarse en el concepto de cambio. Trivializando el escenario un poco, se podrían reconocer dos áreas principales de cambio: cambios en los datos y cambios en los metadatos. Nótese que esta formulación idealista sugiere que las necesidades de información dominarán el desarrollo de herramientas en sistemas de información. Quizá otra postura igualmente justificada sería decir que el desarrollo de herramientas (de legado) domina en los sistemas de información.

Consideremos los cambios con respecto a las siguientes seis capas de representación de información en metadatos (o descripciones

semánticas):

- 6. Interpretación.
- 5. Datos de contenido.
- 4. Esquema(s) semánticos de dominio específico.
- 3. Lenguaje de Representación del Conocimiento (KR).
- 2. Sintaxis.
- 1. Conceptualización.

Intuitivamente, la conceptualización denota el marco de modelado de referencia y dirige todo el proceso de captura de semánticas (por ej. entidades confusas vs. entidades concisas, asunción de mundo abierto o cerrado⁷, casos de uso abstractos, etc).

Se requiere sintaxis para capturar las descripciones, por ejemplo declaraciones lógicas. El lenguaje de KR (*Knowledge Representation*) es necesario para la introducción de estructuras de dominio específico en términos de conceptos fundamentales comunes y consecuencias bien establecidas (por ej. clases y propiedades, implicación o deducción lógica, etc.).

Los esquemas de dominio específico son necesarios para llevar las semánticas específicas de aplicación (por ej. estructuras aceptadas y vocabularios tales como el Dublin Core o RSS). Obviamente también se requieren los datos de contenido, grabando las descripciones semánticas declaradas con respecto a los esquemas seleccionados (por ej. una fecha concreta en la que fue publicado un documento en concreto, utilizando vocabulario bien definido). Por último, la interpretación se refiere a como los esquemas y las descripciones semánticas disponibles se utilizan en las aplicaciones (por ej. utilizando una fecha de edición en una aplicación de calendario, quizá sin seguir las consecuencias formales de las capas subyacentes al modelo).

La relación con las tecnologías de la Web Semántica es obvia. También es obvio que, en principio, los cambios pueden tener lugar en cualquiera de las 6 capas anteriormente mencionadas. Además, el versionado puede tener lugar igualmente en todas estas capas, tanto en datos como en metadatos. Sin embargo, en el sentido más amplio, el versionado requiere un sistema de información completo para sí mismo. El simple hecho de introducir un "vocabulario de versionado" particular requiere la adecuación de las capas 1 a 3.

Está claro que los típicos sistemas de información no soportan bien los cambios arbitrarios. Los motivos están relacionados en parte con el trabajo necesario para la implementación de herramientas adecuadas y en parte relacionadas con el hecho de que no todos los cambios son archivados sistemáticamente con información de versionado completa. Las aplicaciones prácticas se orientan (si es que real-

mente lo hacen) hacia el versionado del contenido, utilizando distintas estrategias de versionado. El versionado de esquemas, los lenguajes de representación del conocimiento y la sintaxis es mucho más difícil y normalmente ocupa un lugar destacado únicamente en aplicaciones muy importantes. En muchos casos los motivos son altamente prácticos: la voluntad de modificar la capa X es en comparación inversamente proporcional a la complejidad, número y calidad de los datos y herramientas existentes en esta capa.

Mientras que en el peor de los escenarios se tiene conocimiento de la complejidad de los cambios potenciales, en concreto, cambios en la conceptualización y el lenguaje de representación del conocimiento, parece que en un escenario razonable la evolución tiene lugar principalmente en las capas 3-6. Además, los lenguajes de representación del conocimiento y los vocabularios de dominio específico que cumplen con los requerimientos técnicos y socioeconómicos (por ej. políticas de licencias) de la comunidad P2P pueden considerarse que evolucionan muy lentamente. Estas observaciones parecen apoyar firmemente los estándares de la Web Semántica W3C.

Como consecuencia, restringir la evolución para que tenga lugar en los niveles 4-6 proporciona la oportunidad de introducir vocabularios comunes. Está claro que es así como han evolucionado las aplicaciones. En un entorno P2P, se pueden utilizar vocabularios comunes para implementar sistemas eficaces basados en palabras claves y sistemas P2P basados en esquemas, describiendo al menos los recursos de P2P desde la perspectiva de la utilización de una búsqueda semántica en concreto.

Además, el tratamiento de la evolución es aparentemente más sencillo si se pueden utilizar semánticas de versionado común. En un principio esto permite también capturar (y quizás explicar) tanto la evolución de vocabularios específicos de dominio como la evolución de contenido dentro de un único sistema de versionado común. Sin embargo, se necesitan políticas comunes para la gestión de la evolución (en concreto soporte para herramientas e interpretación).

4.3. Conocimiento

Hasta ahora hemos hablado en su mayoría sobre la información ¿Y que hay del conocimiento?

Según la caracterización descrita anteriormente, el conocimiento aparece de forma implícita y explícita en la gente, en su entorno, en documentos en papel, (etc.) y en sistemas computerizados. La gestión del conocimiento requiere tanto herramientas para grabar, describir y transformar el conocimiento explícito, como políticas visionarias y lugares de socialización e internalización fructíferos.

Además se debe ser consciente también de los requerimientos de los sistemas de compartición de datos.

Debido a fundamentarse en la información, nuestra discusión anterior sobre la evolución de la información se aplica directamente sobre el conocimiento intermediado. De hecho, un purista en epistemología puede afirmar que el conocimiento externalizado sólo existe como información, quizá por tanto anulando el concepto de conocimiento externalizado formal. Sin embargo, para caracterizar de forma constructiva el ascenso desde la información hasta el conocimiento, consideremos brevemente una aplicación "amigable" que exista en la frontera difusa.

Los esquemas conceptuales proporcionan una aproximación simple y no trivial de las estructuras de conocimiento explícito (por ej. *folksomias* o tesauros). Intuitivamente, el nivel y la aplicabilidad de la información se incrementa hilando algunos de los fragmentos individuales de descripciones conceptuales consistentes en estructura/s común/es. Integrando las necesidades de entendimiento y descubrimiento, los esquemas de concepto también establecen una estructura práctica y extensible para ejecutar búsquedas semánticas basadas en temas.

Desafortunadamente, para alivio de los puristas en epistemología, la utilización de esquemas conceptuales en un sistema P2P se enfrenta de nuevo a los mismos retos y a las mismas fuentes de punto único de fallo a medida que lo hace la evolución de la información. Concretamente, mientras que podría ser técnicamente sencillo acordar un esquema común para la representación de esquemas de concepto (por ej. SKOS, *Simple Knowledge Organization System*), la organización de los conceptos nos conduce básicamente a la gestión de información sin un diseño centralizado.

De hecho, toda esta discusión nos indica que en un sistema centralizado es necesario tolerar las inconsistencias y los solapamientos.

4.4. Hacia espacios de conocimiento compartido: búsqueda semántica revisada

Una estrategia atractiva para abordar las inquietudes de las vistas desparejadas, los cambios evolutivos y la falta de una visión global, es considerar los espacios de conocimiento en términos de sistemas de información aportada con una memoria de versión a largo plazo. La idea básica es fundamentalmente que un espacio de conocimiento evolutivo no es un sistema de información estrictamente integrado y bien definido con semánticas perfectas e integridad referencial. En su lugar, aparece como una colección de sistemas de información de contribución que no coinciden exactamente con los esquemas, datos de con-

tenido o interpretaciones (o incluso las otras capas de información subyacentes).

En consecuencia, el papel de la búsqueda semántica destaca de nuevo. En un sentido técnico muy abstracto, la interfaz de la búsqueda semántica representa más o menos la totalidad del espacio de conocimiento, denotando quizá en sí mismo el artefacto de unión: si no puede encontrarse una referencia a un objeto por medio de la interfaz de búsqueda, ese objeto no existe en sentido útil en el espacio de conocimiento.

En un sistema distribuido, la red P2P es responsable idealmente de declarar esta interfaz. Además, para gestionar tendencias a nivel del sistema, los cambios anteriores también se deben recordar. En general, esto puede requerir actividades de archivo y versionado en varias capas de información. Considerando la Web Semántica, esto destaca la importancia de nombres de espacio apropiados y de la gestión de vocabulario semántico.

Una estrategia complementaria alternativa es el reconocimiento del contexto de afirmación y el contexto de interpretación de descripciones semánticas. En principio, esto habilita las perspectivas orientadas a aspectos para el espacio de conocimiento compartido. El que esto sea factible en la práctica está probablemente relacionado con el grado de automatización en la gestión y la codificación en la información contextual (ver **sección 3.3**).

Tratar con contextos de interpretación y afirmación podría considerarse como responsabilidad de la interfaz de búsqueda, o simplemente dejado por el usuario (por tanto reconociendo los contextos simplemente como búsqueda "extra" y criterios de traducción de conocimiento).

5. Conclusión

En este artículo hemos resumido tres dominios complementarios: sistemas de compartición de datos y conocimiento, Web Semántica y OPAALS OKS, así como presentado y analizado brevemente dos retos relacionados, a saber: información P2P y evolución de la información. Hasta cierto punto, la discusión podría resumirse diciendo que mientras que la búsqueda semántica parece caracterizar el verdadero concepto de un espacio de conocimiento unificado y que evoluciona, y que por tanto es claramente necesario, esto por sí sólo no es suficiente. Concretamente, políticas y factores sociales se deben tener también en cuenta.

El resumen y análisis presentados deberían sin embargo ayudar a comprender los tres dominios que se solapan, señalando las áreas destacadas susceptibles de futuras investigaciones y desarrollo. Por ejemplo, la observación de la

capa de interpretación (ver **sección 4.2**) es bastante interesante. Resulta que esta capa hasta cierto punto puede formalizarse, por ejemplo en términos de lógicas de interpretación [8]. Esto proporciona una estrategia concreta (aunque abstracta) para la implementación de una interfaz de búsqueda unificada en un sistema de conocimiento compartido.

Obviamente quedan muchos retos. A pesar del abrumador tamaño del problema, se puede y debe realizar un esfuerzo continuo en pequeños pasos. Sin embargo, el problema fundamental de los espacios de conocimiento, incluyendo la Web Semántica, parece recaer en sus propiedades de (eco)sistema. Los experimentos controlados y los análisis no pueden realizarse durante todo el proceso, en algún momento las cosas simplemente necesitan probarse a gran escala.

Agradecimientos

La red de excelencia IST (*Information Society Technologies*) del proyecto OPAALS del sexto programa marco ha sido apoyada por la Comisión Europea (número de contrato: FP6-034824).

Referencias

[1] R. Steinmetz, K. Wehrle. What is This "Peer-to-Peer" About? En R. Steinmetz, K. Wehrle (Eds): *Peer-to-Peer Systems and Applications*. Lecture Notes on Computer Science, Springer, 2006. ISBN-10: 354029192X.
 [2] O. Nykänen, J. Salonen, J. Huhtamäki, M. Haapaniemi. *OKS Data Model, version 1.01*. A milestone specification for the OPAALS project (Contract number IST-034824), WP10: Sustainable Research Community Building in the Open Knowledge Space <<http://matriisi.ee.tut.fi/hypermedia/julkaisut/20080701-oks-dm-v1-01.pdf>>.

[3] O. Nykänen, J. Salonen, M. Haapaniemi, J. Huhtamäki. A Visualisation System for a Peer-to-Peer Information Space. *Proceedings of OPAALS 2008*, 7-8 Octubre 2008, Tampere (Finlandia). <<http://matriisi.ee.tut.fi/hypermedia/julkaisut/20080807-nykanen-et-al-p2pviv.pdf>>.
 [4] I. Nonaka, H. Takeuchi. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, USA, 1995. ISBN-10: 0195092694.
 [5] K. Smith, L. Seligman, V. Swarup. Everybody Share: The Challenge of Data-Sharing Systems. *IEEE Computer Magazine*, September 2008.
 [6] I. Jacobs, N. Walsh (Eds.). *Architecture of the World Wide Web, Volume One*. W3C Recommendation 15 December 2004 <<http://www.w3.org/TR/webarch/>>
 [7] P. Dini. Notes on Relational Biology and Elementary Category Theory. *Proceedings of OPAALS 2008*, 7-8 Octubre 2008, Tampere, Finlandia. <<http://matriisi.ee.tut.fi/hypermedia/julkaisut/opaals2008-proceedings.pdf>>
 [8] O. Nykänen. Interpretation Logics. *Proceedings of the 1st OPAALS conference*, 26-27 Noviembre 2007, Roma, Italia. <<http://matriisi.ee.tut.fi/hypermedia/julkaisut/2007-nykanen-ilogics.pdf>>.

Notas

¹ Actividades en Web Semántica del W3C, <<http://www.w3.org/2001/sw/>>.
² La autopoiesis (del griego auto, "sí mismo", y poiesis, "creación" o "producción"), es un neologismo propuesto en 1971 por los biólogos chilenos Humberto Maturana y Francisco Varela para designar la organización de los sistemas vivos. Una descripción breve sería decir que la autopoiesis es la condición de existencia de los seres vivos en la continua producción de sí mismos... Según Maturana y Varela son autopoieticos los sistemas que presentan una red de procesos u operaciones (que lo define como tal y lo hace distinguible de los demás sistemas), y que pueden crear o destruir elementos del mismo

sistema, como respuesta a las perturbaciones del medio.
³ Página principal de la red de excelencia investigadora multidisciplinaria OPAALS, <<http://www.opaals.org/>>
⁴ Reificación es un proceso a través del cual un objeto computable o direccionable (un recurso) es creado en un sistema como representante de un objeto no computable o direccionable (Wikipedia en inglés) <[http://en.wikipedia.org/wiki/Reification_\(computer_science\)](http://en.wikipedia.org/wiki/Reification_(computer_science))> [23-3-2009].
⁵ La arquitectura en pipeline (basada en filtros) consiste en ir transformando un flujo de datos en un proceso comprendido por varias fases secuenciales, siendo la entrada de cada una la salida de la anterior. Esta arquitectura es muy común en el desarrollo de programas para el intérprete de comandos, ya que se pueden concatenar comandos fácilmente con tuberías (pipe) <<http://es.wikipedia.org/wiki/Pipeline>> [23-3-2009].
⁶ Un punto unico de fallo (SPOF, *Single Point of Failure*) es una parte de un sistema que si falla detiene el funcionamiento de todo el sistema. No son deseables en ningún sistema cuyo objetivo sea la alta disponibilidad... (Wikipedia en inglés) <http://en.wikipedia.org/wiki/Single_point_of_failure> [23-3-2009].
⁷ En lógica formal, la asunción de mundo abierto (*open world assumption*) es la asunción de que la verdad o falsedad de una declaración es independiente de si ésta es conocida por cualquier observador o agente. Por el contrario, la asunción de mundo cerrado (*closed world assumption*) sostiene que cualquier declaración que no se reconoce como verdadera es falsa. (Wikipedia en inglés) <http://en.wikipedia.org/wiki/Open_World_Assumption> [23-3-2009].

	<p align="center">IX Jornadas sobre Programación y Lenguajes (PROLE) I Taller de Programación Funcional (TPF) San Sebastián, 8-11 de septiembre de 2009</p>										
<p>Las Jornadas sobre PROgramación y Lenguajes (PROLE) constituyen un marco propicio de reunión, debate y divulgación de resultados para los grupos españoles que investigan en temas relacionados con la programación y los lenguajes de programación. Este año se complementa con un Taller específico sobre Programación Funcional (TPF), como actividad independiente. Se solicitan trabajos relacionados con los lenguajes de programación, en diferentes categorías, en las fechas que se indican:</p>											
<table border="1"> <thead> <tr> <th data-bbox="272 1839 651 1906">TPF'09 / PROLE'09 Fecha de celebración</th> <th data-bbox="651 1839 943 1906">Recepción de Resúmenes / Trabajos</th> <th data-bbox="943 1839 1150 1906">Notificación de aceptación</th> <th data-bbox="1150 1839 1305 1906">Versión definitiva</th> </tr> </thead> <tbody> <tr> <td data-bbox="272 1906 651 2000">Día 8 / días 9 al 11 Septiembre</td> <td data-bbox="651 1906 943 2000">1 de Abril / 11 de Abril</td> <td data-bbox="943 1906 1150 2000">6 de Junio</td> <td data-bbox="1150 1906 1305 2000">20 de Junio</td> </tr> </tbody> </table>				TPF'09 / PROLE'09 Fecha de celebración	Recepción de Resúmenes / Trabajos	Notificación de aceptación	Versión definitiva	Día 8 / días 9 al 11 Septiembre	1 de Abril / 11 de Abril	6 de Junio	20 de Junio
TPF'09 / PROLE'09 Fecha de celebración	Recepción de Resúmenes / Trabajos	Notificación de aceptación	Versión definitiva								
Día 8 / días 9 al 11 Septiembre	1 de Abril / 11 de Abril	6 de Junio	20 de Junio								
	Contacto: paqui.lucio@ehu.es http://www.mondragon.edu/prole2009/										

María Jesús Fernández Ruiz, Víctor Morlán Plo
Ayuntamiento de Zaragoza

<mjferuiz@zaragoza.es>, vmorlan@gmail.com>

La Web del Ayuntamiento de Zaragoza como servicio de Atención al Ciudadano

1. Introducción

La Web juega un papel cada vez más importante en el ámbito de los servicios educativos, profesionales, económicos, políticos y sociales, convirtiéndose en un medio habitual para el intercambio de información y conocimiento entre los miembros de la sociedad. En este ámbito la Web del Ayuntamiento de Zaragoza se ha convertido en un canal importante de servicios de Administración Electrónica (AE) para los ciudadanos (cualesquiera personas físicas, personas jurídicas y entes sin personalidad que se relacionen, o sean susceptibles de relacionarse, con las Administraciones Públicas)¹.

Los ciudadanos tienen a su disposición un número cada vez mayor de servicios (solicitud de Licencias Urbanísticas, Pago de Impuestos, etc.) y disfrutan de un acceso más sencillo a la información de interés general, incluyendo la legislación y la información reguladora, datos sobre las actividades del Ayuntamiento, los debates sobre la política pública, así como la información esencial sobre los recursos de su ciudad, como, por ejemplo, datos geoespaciales. Esta vía de comunicación facilita el ahorro de costes y tiempo, tanto a los interesados particulares en determinados servicios, como al ayuntamiento encargado de prestarlos.

Se conoce como Administración Electrónica: *"el uso intensivo de las Tecnologías de la Información y de la Comunicación y, en particular, de Internet para la realización de las actividades de la Administración Pública con el objetivo de mejorar o transformar la relación con el ciudadano, las relaciones entre las administraciones, la prestación de servicios públicos, la eficiencia interna y la participación política"*²

Por ello es esencial que todas las personas, independientemente de sus condiciones culturales, situaciones geográficas y capacidades físicas puedan acceder a los servicios contenidos en los sitios web.

La evolución de la Web del Ayuntamiento de Zaragoza viene marcada tanto por la evolución de la Informática como por la mayor versatilidad de los servicios móviles, el incremento de servicios demandados por el ciudadano y, en particular, por el papel clave que la administración debe tener en el desarrollo de la Sociedad del Conocimiento.

La Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios

Resumen: La Web, que se ha convertido en un mecanismo cada vez más habitual para el intercambio de información y conocimiento entre los miembros de la sociedad, es para el Ayuntamiento de Zaragoza un canal crítico en el proceso de interacción con el ciudadano. El objetivo principal de la Web Municipal es ofrecer a través de un punto de acceso único un servicio integral a los ciudadanos y a las empresas. La información facilitada a través de la Web, debe ser integral, actualizada, fiable y presentarse en un formato sencillo y acorde con su naturaleza y el perfil del usuario al que va dirigida, sin presentar obstáculos técnicos y al menor coste de tiempo. En este artículo se aborda la forma en la que el Ayuntamiento de Zaragoza ha creado un importante canal de servicios de Administración electrónica para los ciudadanos.

Palabras clave: accesibilidad, administración electrónica, ciudadano, independencia de dispositivo, Web Móvil, Web Semántica, W3C.

Autores

M^a Jesús Fernández Ruiz es responsable de la Web del Ayuntamiento de Zaragoza. Es Licenciada en Filosofía y Letras, División de Geografía e Historia, Sección Geografía, por la Facultad de Filosofía y Letras de la Universidad de Zaragoza (1984). Obtuvo el Diploma de Estudios Avanzados en el Programa de Doctorado de Ingeniería de Sistemas e Informática del Departamento de Informática e Ingeniería de Sistemas, así como la Suficiencia Investigadora en el Área de Conocimiento de Ciencias de la Computación e Inteligencia Artificial (2004). Actualmente es Responsable de la Unidad de Gestión de la Web del Ayuntamiento de Zaragoza desde su inicio 1994. Esta unidad gestiona el proyecto "Zaragoza Ciudad Móvil" (que recibió el Premio e-Mobility SIMO 2004 como "la institución u organismo público que ha desarrollado o contribuido al desarrollo de una o varias aplicaciones móviles y/o inalámbricas que han ayudado a mejorar las condiciones de vida de los ciudadanos") y de las sedes electrónicas www.zaragoza.es, www.ayto-zaragoza.es y para el trabajador municipal Intranet.ayto-zaragoza.es, así como de la gestión de la información necesaria para dar una respuesta clara y concreta a las demandas que llegan del ciudadano a través del servicio 010. Es miembro de la Comisión Técnica para la implantación de la Administración Electrónica en el Ayuntamiento de Zaragoza.

Víctor Morlán Plo, nacido en Huesca en 1978, es Ingeniero Técnico en Informática de Gestión por la Universitat de Lleida (2002). Trabaja en la Oficina Web del Ayuntamiento de Zaragoza desde el año 2002.

Públicos, define en su Capítulo Primero la Sede Electrónica como "dirección electrónica disponible para los ciudadanos a través de redes de telecomunicaciones cuya titularidad, gestión y administración corresponde a una Administración Pública funcionando con plena responsabilidad respecto de la integridad, veracidad y actualización de la información y los servicios a los que puede accederse a través de la misma".

2. Objetivo de la Web del Ayuntamiento de Zaragoza

El objetivo principal de la Web Municipal (ver figura 1) es ofrecer a través de un punto de acceso único un servicio integral a los ciudadanos y a las empresas. Es decir, ofrecer acceso a los servicios de la Administración Electrónica, a pesar de las restricciones asociadas a la organización administrativa y política, veinticuatro horas al día los siete días de la semana, debiendo ser accesible, útil y fácil de usar para el mayor número de personas, con independencia de su entorno social, cultural y tecnológico.

La información facilitada a través de la Web debe ser integral, actualizada, fiable y presentarse en un formato sencillo y acorde con su naturaleza y el perfil del usuario al que va dirigida, sin presentar obstáculos técnicos y al menor coste de tiempo.

3. Servicios de la Web del Ayuntamiento de Zaragoza

Una forma de conseguir el objetivo anteriormente expuesto es:

■ Proporcionando a los ciudadanos información, vía telemática, sobre distintos aspectos de la organización municipal y la ciudad, como:

- Bajo la etiqueta "Ayuntamiento" información sobre: la Estructura y Normativa Municipal; Oferta de Empleo y Tablón Municipal; Perfil del Contratante (información sobre los contratos públicos); Ayudas y Subvenciones; Convocatoria de Premios y Cursos. En el "Portal de Centros Municipales" se muestra información sobre los Centros y Edificios Municipales: información de



Figura 1. Página principal de la Web Municipal de Zaragoza.

sus instalaciones, de los servicios que prestan, una imagen de su fachada principal, su ubicación en el plano e información sobre las líneas de autobuses que facilitan el acceso a los ciudadanos a los mismos, etc.

- Bajo las etiquetas "**Ciudad**"; "**Cultura**" y "**Turismo**" podemos encontrar: Agenda Zaragoza, Información Turística, Callejero, Incidencias en la Vía Pública, Educación, Deporte, así como Información Medioambiental y Urbanística. A través del "Portal de Medio Ambiente" una información veraz, actualizada y accesible sobre la política medioambiental del ayuntamiento, tanto de las actuaciones en curso como de la toma de decisiones, que favorezca la participación del ciudadano (entendiendo por ciudadano cualquier ente social) en estos temas. Ejemplos: Información diaria sobre la calidad del aire desde 1999; Información y Documentación a través de los servicios y catálogo en línea del Centro de Documentación; etc.

- El usuario a través de la carpeta Ciudadano, utilizando el correspondiente certificado digital, puede consultar sus datos fiscales y situación de empadronamiento.

- En el portal IDEzar (*Infraestructura de Datos Espaciales de Zaragoza*), la Web proporciona al ciudadano todo un conjunto de servicios que le son útiles en función de su referente geográfico estable (generalmente domicilio) dentro de la ciudad, y su deambu-

lar por la misma: callejeros, mapas de interés, rutas, etc. (ver figura 2). Frente a una aproximación clásica muy disjunta a la hora de proporcionar estos servicios, basada en aplicaciones específicas orientadas a ofrecer únicamente un conjunto limitado de funcionalidades, y generalmente poco relacionadas, en el Ayuntamiento de Zaragoza se ha hecho una apuesta por sistemas más abiertos. Tomando como referencia la filosofía que proponen las Infraestructuras de Datos Espaciales (IDEs) [1], se ha puesto en funcionamiento un *middleware* de servicios sobre el que se están desarrollando aplicaciones que soportan las funcionalidades ofertadas al ciudadano. Esta aproximación no solo permite un desarrollo más homogéneo de los sistemas que se están creando (con la consecuente mejora en tasas de reutilización y la propia producción de servicios), sino que permite aprovecharse de servicios de información que se están poniendo en marcha por otras administraciones públicas en virtud de la puesta en marcha de la Directiva Europea INSPIRE [2]. El conjunto de servicios propios puestos en funcionamiento, más las aplicaciones que sobre estos servicios propios y otros "externos" se han construido, constituyen lo que se ha denominado Infraestructura de Datos Espaciales de Zaragoza (IDEZar) [3]. Dicha infraestructura <<http://www.zaragoza.es/idezar/>> se encarga de explotar el carácter geográfico de la información urbana para hacerla accesible al ciudadano a través de distintas aplicaciones y dispo-

sitivos maximizando su usabilidad y la comprensión de la misma. Entre los servicios al ciudadano más representativos, y que mejor aprovechan la aproximación IDE, caben destacar los siguientes: Servicio de Callejero; Incidencias en Vía Pública; Planifica tu visita; Callejero Móvil, etc.

■ Facilitando la realización de servicios, de trámites, como pueden ser la solicitud y/o ejecución de trámites, documentación o expedientes. Los servicios ofrecidos deben ser ante todo útiles y accesibles para el ciudadano, demandados por éstos, transparentes y eficientes. Esto se lleva a cabo a través de herramientas como:

- "**Ayuntamiento en Casa**" que pretende ser una herramienta eficaz al reunir todas las informaciones útiles para realizar un número importante de solicitudes y trámites con el Ayuntamiento. Dirigida tanto al ciudadano como a las Asociaciones, Colegios Profesionales, Entidades Empresariales, Sindicatos y Federaciones vecinales. Informaciones sobre subvenciones y ayudas, contratos públicos, premios y concursos, otras noticias de interés, en definitiva cuanta información pueda ser útil para promover y respaldar la administración electrónica. Una ventanilla única que tiene como objetivo ofrecer información, trámites y servicios de todo el Ayuntamiento y de otras administraciones más demandados por el ciudadano. Información sobre más de 583 trámites administrativos y la posibilidad de

complimentar más de 100 impresos. El Portal también permite realizar tramitación electrónica, (Solicitud de Certificado de Empadronamiento, Inscripción a cursos de la Casa de la Mujer, o la tramitación del recurso de una sanción de tráfico, etc.). Un total de 150

Web", servicio creado en 2007 que tiene como objetivo responder didácticamente a dudas ciudadanas en materia informática, para favorecer la utilización correcta de todos los servicios que se difunden a través de la Web y el conocimiento básico de los diferentes dis-

gidos en el articulado de la Ley 11/2007:

a) Derecho a la protección de datos de carácter personal en los términos establecidos por la Ley Orgánica 15/1999, de Protección de los Datos de Carácter Personal.

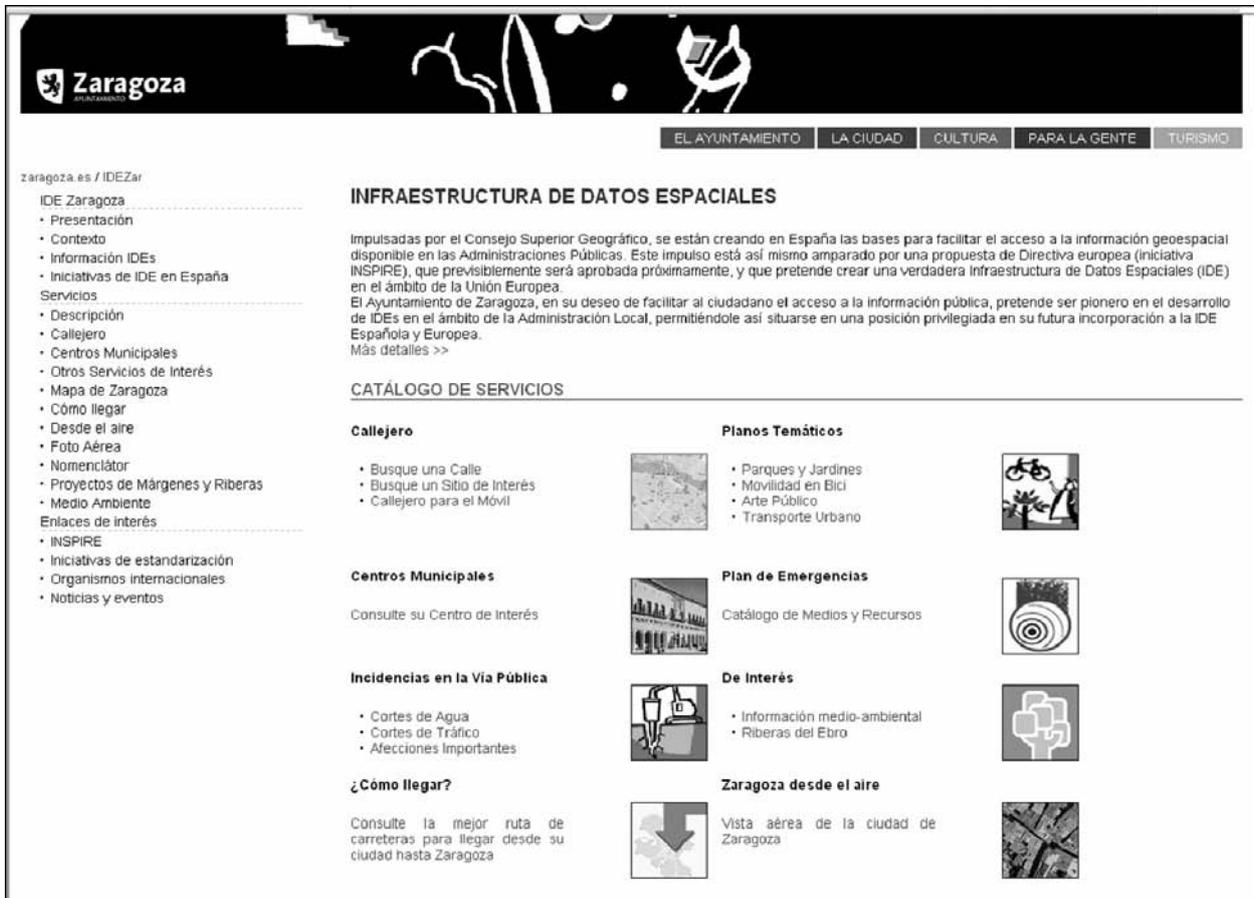


Figura 2. Portal IDEzar (Infraestructura de Datos Espaciales de Zaragoza).

trámites. Algunos de estos trámites también se pueden realizar a través de teléfono móvil.

- **PASURBAN:** Tutorización en línea de las solicitudes de licencias urbanísticas. Está dirigido tanto al ciudadano como a los profesionales y colectivos relacionados con urbanismo: Colegios Profesionales, Entidades Empresariales, Sindicatos y Federaciones vecinales. Accesible desde el "Portal de Urbanismo"

■ Favoreciendo el desarrollo de movimientos y redes sociales y nuevas formas de gobierno electrónico, poniendo a disposición de los ciudadanos distintos servicios de participación. A través de servicios como "Noticias" y "Agenda", el ciudadano puede participar en la publicación de contenidos y crear redes sociales (Blog Ciudadano³ cuenta con unos 1.500 blogs alojados; Weblog Educación Ambiental; Weblog Zaragoza Sin Barreras, espacio para el intercambio de vivencias de lo bueno, de lo menos bueno, para realizar preguntas, para dar respuestas, sobre la movilidad en los espacios urbanos y edificio públicos de nuestra ciudad). Destacaríamos el weblog "Educa

positivos que facilitan el acceso a los mismos.

■ Proporcionando al ciudadano un servicio de Quejas y Sugerencias para que puedan comunicar al Ayuntamiento todas sus inquietudes, intereses y reclamaciones, con la confianza de recibir en toda ocasión una respuesta personalizada. A través de este Servicio, el ciudadano puede saber si se registra el estado de su queja o de su sugerencia.

Es importante, y un indicador de calidad, informar a los ciudadanos de los servicios que se prestan a través de las Webs Municipales y de los compromisos que el ayuntamiento ha adquirido con todos ellos a la hora de prestarlos. Las Cartas de Servicio pueden ser un instrumento de mejora y de calidad.⁴

Los servicios prestados por el Ayuntamiento de Zaragoza a través de la Web tienen en cuenta las limitaciones establecidas por la Constitución y el resto del ordenamiento jurídico, respetando el pleno ejercicio por los ciudadanos de los derechos que tienen reconocidos, y ajustándose a los principios reco-

b) Principios de Igualdad y de Accesibilidad a la Información y a los Servicios por Medios Electrónicos.

c) Libre Acceso a la Información: Transparencia; Publicidad.

d) Efectividad y Eficiencia: Simplificación y Cooperación, transmisión de datos entre administraciones.

e) Responsabilidad y calidad en la veracidad y autenticidad de las información y servicios ofrecidos por las Administraciones Públicas a través de medios electrónicos.

f) Privacidad y Seguridad.

g) Personalización.

h) Principio de Neutralidad Tecnológica; Multicanalidad.

4. Política de accesibilidad y adaptación de contenidos de la Web Municipal

La Sociedad del Conocimiento es una fuente de oportunidades porque permite ofrecer nuevos servicios interactivos a los colectivos con necesidades concretas superando barreras que antes eran difícilmente franqueables, pero para ello es necesaria la implementación

“Los ayuntamientos, como toda administración pública, tienen la obligación de garantizar la igualdad de oportunidades en el acceso a los nuevos servicios y productos digitales”

de una política de accesibilidad e independencia de dispositivo en la gestión de las sedes electrónicas de la Administración.

La Web, como una vía de comunicación de los servicios electrónicos, conduce de forma natural a una concepción, cada vez más extendida, de la AE como derecho subjetivo, como una parte central de los "nuevos" derechos de los ciudadanos en la sociedad del conocimiento, junto con otros derechos sociales relacionados con el acceso, disponibilidad y formación en el uso de las nuevas tecnologías (para suavizar la llamada "fractura digital", o los derechos en materia de protección de datos y comunicaciones personales)⁵. Todo el mundo debe tener la oportunidad de acceder a la red en igualdad de oportunidades y de condiciones eliminando cualquier riesgo de exclusión, y en este sentido, los ayuntamientos deben velar por la accesibilidad de todos a la red.

En España, las Administraciones Públicas tienen el deber de garantizar la igualdad de oportunidades de los ciudadanos en el acceso a los servicios y productos digitales, procurando que ningún colectivo social quede excluido (Ley 34/2002, de 11 de julio; Ley 51/2003, de 2 de diciembre; Ley 11/2007, de 22 de junio).

Bajo esta premisa, la accesibilidad es un tema esencial ya que trata de eliminar o minimizar las barreras con las que se encuentran determinados grupos de usuarios al navegar por Internet e intentar acceder a información. La accesibilidad es un elemento esencial para favorecer la igualdad de oportunidades de las personas con discapacidad, permitiendo el ejercicio del derecho reconocido constitucionalmente como es el acceso a la cultura, el ocio y el tiempo libre.

Nuestra hipótesis de trabajo es que los ayuntamientos, como toda administración pública, tienen la obligación de garantizar que la implantación de la Sociedad del Conocimiento se desarrolle de forma armónica. Es decir, tienen la obligación de garantizar la igualdad de oportunidades de los ciudadanos en el acceso a los nuevos servicios y productos digitales, que ningún colectivo social quede excluido y procurar una adecuada prestación de los servicios públicos a través de diferentes vías de comunicación: el ordenador, Internet, SMS, Dispositivos Móviles, Televisión Digital Terrestre Interactiva. Hay que tener en cuenta los canales y servicios que existen en la actual-

idad o que puedan existir en el futuro. También es importante destacar que estas vías de comunicación tienen que coexistir con el canal presencial y telefónico.

Esa garantía puede materializarse debido a que los propios avances de la Informática están acercando los servicios al ciudadano, ofreciendo accesos desde diferentes ubicaciones mediante el uso de variados dispositivos. La velocidad con la que se desarrollan e implementan servicios puede conducirnos, y de hecho está ocurriendo, a que una parte importante de los ciudadanos quede excluida de estos servicios por ofrecerse accesos no adaptados a sus limitaciones físicas, culturales, etc.

Sin embargo, al mismo tiempo que los propios avances de la tecnología pueden conducirnos a la construcción de servicios no accesibles, los mismos avances nos proporcionan medios que, utilizados adecuadamente, permiten la construcción de servicios con un alto grado de accesibilidad.

En este orden de ideas hay que resaltar la labor de estandarización que está desarrollando el consorcio W3C (*The World Wide Web Consortium*) y, en lo que respecta a nuestra comunicación, destacar sus propuestas sobre CSS, independencia de dispositivos, Web Semántica y tecnologías XML. En estos estándares, el concepto de accesibilidad web (expresado inicialmente por Tim Berners-Lee, creador del World Wide Web y cofundador del consorcio W3C en 1994) es fundamental.

El Ayuntamiento de Zaragoza desde 1994 tiene presencia en Internet y en 1997 ha tenido entre sus objetivos facilitar el acceso al mayor número de personas, independientemente de sus características personales y de la plataforma tecnológica elegida, mejorando día a día la capacidad y rapidez de la Web Municipal <<http://www.zaragoza.es/>>. Desde el principio en el sitio web del Ayuntamiento existe una gran preocupación por conseguir que todas sus páginas y servicios sean accesibles y usables; es decir, que el sitio web pueda ser navegado y leído por cualquiera, sin tener en cuenta aspectos como la localización, la experiencia del usuario, o la plataforma de acceso (dispositivos, software, tipo de conexión) empleada.

Por lo tanto, es muy importante la implantación de una política de accesibilidad y adaptación de contenidos en el diseño y desarrollo de

una sede electrónica, así como el potenciar la Red Ciudadana.

La web del ayuntamiento de Zaragoza obtuvo en 2007 la Certificación Oficial en Accesibilidad TIC concedida por AENOR (*Asociación Española de Normalización y Certificación*) según Norma UNE 139803 "Requisitos de accesibilidad para contenidos Web" con Nivel de accesibilidad AA, en cumplimiento de la Ley 56/2007, de 28 de diciembre, de Medidas de Impulso de la Sociedad de la Información (LISI).

Y en junio de 2008 la auditoria realizada por AENOR confirma que la web del ayuntamiento cumple con todos los criterios de calidad que exige esta Asociación para conceder la Certificación Oficial en Accesibilidad TIC. Lo que ha conllevado en la Gestión de la Web una serie de ventajas adicionales como son:

- Simplificación del desarrollo.
- Ahorro de costes.
- Mejora de la indexación.
- Mejora del acceso en general.
- Facilita la independencia de dispositivos.
- Aumento de la usabilidad.

El resultado de la certificación ha sido la implantación de un Sistema de Gestión de Accesibilidad basado en 4 elementos fundamentales:

- Gestión de los recursos (personal, formación, herramientas).
- Elaboración y mantenimiento del sitio web lo que implica la planificación, desarrollo y documentación de los procesos necesarios para el diseño, desarrollo, gestión de contenidos y mantenimiento, así como los métodos de control para asegurar la accesibilidad del sitio Web.
- Gestión de proveedores.
- Tratamiento de reclamaciones de clientes.

La Política de Accesibilidad del sitio web ha sido recogida desde 1997 en diferentes documentos: la "Guía de Estilo", la "Política de Accesibilidad y Adaptación de Contenidos de la Web". Este último documento fue elaborado en 2005 y está actualmente en estado de revisión. Se edita en colaboración con la Fundación CTIC (*Centro Tecnológico de la Información y de la Comunicación*), y en él se establece la accesibilidad en relación al sitio web completo, que no es más que la suma de la accesibilidad de todas y cada una de sus páginas.

Los criterios marcados a la hora de la elaboración de contenidos consisten básicamente en el uso intensivo de los estándares del W3C, seguido de una evaluación continua de los contenidos.

Existe una separación total entre el contenido, la estructura y la presentación de la información. Utilizamos para el contenido xHTML 1.0 DTD Strict salvo cuando es necesario mostrar información geográfica, momento en el que utilizamos xHTML 1.0 con DTD Transitional Para la presentación se utiliza CSS 2.1.

La estructura de navegación se encuentra separada de los contenidos y de la presentación, de forma que en el momento en el que el usuario realiza una petición, el servidor se encarga de fusionar los contenidos con la plantilla que alberga la estructura y se la devuelve al cliente.

Por otro lado cada una de las páginas alojadas debe contener:

- Metainformación, como mínimo, <title>, (meta) encoding, (meta) description, (meta) keywords.
- Relaciones entre documentos, utilizando <link>, como mínimo rel="home" rel="index", rel="search", rel="copyright", rel="author".

tabulares (que nunca se utilizan para maquetar la información), marcado de los cambios de idioma y del idioma principal del documento, etc.

- Textos claros y sencillos con el objetivo de facilitar la comprensión.

Como hemos comentado, la aplicación de las pautas de accesibilidad ayuda a conseguir que los sitios sean más usables, pero para obtener una información más fidedigna en 2006 se firmó un Convenio de Colaboración con el Laboratorio de Usabilidad del Gobierno de Aragón para realizar todos los años un estudio de usabilidad de la Web. Las pruebas que el Laboratorio realiza son: Evaluación Heurística, test de usuario y Focus Group.

El test de usuario se realiza con 20 usuarios, en 4 sesiones de 5 personas cada una. Para calcular la muestra se han utilizado los datos recogidos en los estudios "Internet en Aragón: Estudio sobre Hogares (2005-2006) y Personas (2004-2006)" del Observatorio Aragonés de la Sociedad de la Información, tomando como marco poblacional los usuarios de Internet en Zaragoza capital.

Un paso más en la mejora de la accesibilidad es facilitar el acceso a la Web desde diferentes dispositivos móviles de una forma tan fácil, sencilla y apropiada como en el caso de los dispositivos de escritorio. En el portal "Ciudad Móvil" (ver **figura 3**) se describen las dos

dados pueden relacionarse con el Ayuntamiento a través de SMS desde el teléfono móvil, con servicios como petición de certificado de empadronamiento (SMS al nº 5010), envío diario y gratuito de los niveles de polen en época de alergias, Oferta de Empleo o descarga del callejero. Se han creado asimismo juegos virtuales formativos, de educación vial o de recorridos por Zaragoza mediante adivinanzas.

- www.zaragoza.mobi. La Web puede consultarse desde cualquier dispositivo (PC, móvil, PDA,...) adaptándose a las características de éste para que se pueda observar en su totalidad. Es importante implementar la independencia web de dispositivo basada en el paradigma de una Web Única para todos (uniendo a esta solución el seguimiento de las pautas de las Web Content Accessibility Guidelines 1.0 de W3C) [4]. El enfoque es facilitar la adaptación de los contenidos de la Web a dispositivos móviles, a partir de un contenido que previamente está siendo ofrecido para navegadores web de escritorio (Microsoft Internet Explorer, Mozilla Firefox, Opera, Safari, Konqueror, etc.) [5].

La "Web Única" no significa que la misma información deba estar disponible exactamente de la misma manera desde una URI dada. Debe ser tenido además en cuenta el contexto en el que un recurso web vaya a ser

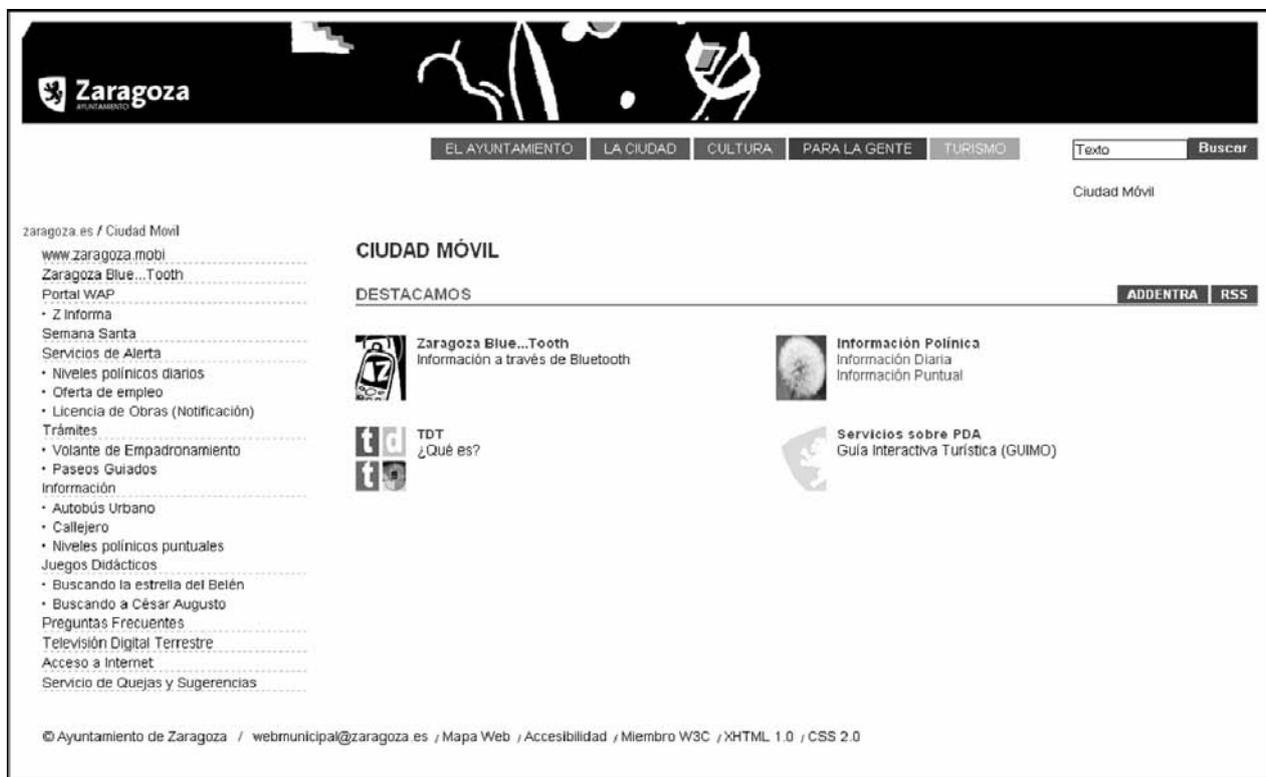


Figura 3. Portal del proyecto "Zaragoza Ciudad Móvil".

- Contenido semántico, utilizamos la especificación xhtml correctamente, es decir, uso de elementos de encabezado, párrafos, listas, tablas

líneas de trabajo de la Web en este campo:

- Servicios "ad hoc" para móviles: Los ciu-

mostrado; a causa de esto, el concepto de "capa de adaptación" se convierte en un elemento integral (y transparente, en la medida

de lo posible) de la Web.

La Web del Ayuntamiento de Zaragoza está entre las implementaciones de referencia de las Buenas Prácticas para la Web Móvil 1.0 (Recomendación Candidata), publicadas por la *Iniciativa para la Web Móvil* (MWI) del W3C [6].

Con el objetivo de mejorar la multicanalidad de la Web, actualmente algunos de los servicios son accesibles a través de la Televisión Digital Terrestre (TDT) Interactiva y de Bluetooth. A la hora de mostrar los servicios nos basamos en la reutilización de contenidos que son fácilmente procesables por la plataforma de TDT al poder ser exportados a formato RSS (*Really Simple Syndication*).

5. Servicios Personalizados, utilización de Tecnología Semántica

Uno de los portales más visitados en el sitio web del Ayuntamiento de Zaragoza es: "Turismo". Sin embargo, los turistas se enfrentan en la Web al mismo problema que cualquier otro usuario: la sobrecarga de información y la falta de tiempo o de conocimiento para analizarla correctamente.

La tecnología de la Web Semántica es un compendio de lenguajes y herramientas para extender la Web actual de manera que el acceso a la información relevante para el usuario pueda ser agilizado y, en muchos escenarios, automatizado. Esta tecnología se basa en la descripción formal de los recursos, donde recurso es toda entidad que pueda ser nombrada en la Web. Sobre estas descripciones se pueden construir aplicaciones y servicios que exploten esta semántica.

En la Web del ayuntamiento podemos destacar:

- **"Planifica tu Visita":** es una aplicación web, que utiliza tecnología semántica, en forma de reglas y ontologías, para que, aprovechando los contenidos existentes en la sede electrónica del ayuntamiento de Zaragoza, el usuario obtenga una ruta por la ciudad adaptada a sus preferencias y circunstancias personales⁶.

Esta aplicación permite a los turistas definir su perfil de visita a la ciudad y obtener como resultado una propuesta de ruta para sus días de estancia en Zaragoza. A través de un interfaz rico en información y a la vez sencillo de usar, los visitantes pueden obtener detalles sobre los monumentos, los eventos disponibles para cada día de estancia, restaurantes de acuerdo a sus gustos gastronómicos, zonas comerciales y, por supuesto, de ocio. "Planifica tu visita" se ocupa de organizar una propuesta turística completa, personalizada y única.

Es una aplicación elegida como Buena Práctica por el Consorcio W3C [7] en la que se utiliza una ontología⁷ para organizar la información RDF (*Resource Description*

Framework). Esta ontología define tres tipos de entidades que constituyen los elementos abstractos de la aplicación:

1. Recursos turísticos.
2. Perfiles de usuario.
3. Configuración de la ruta.

La estructura conceptual de nuestra ontología está basada en la ontología DOLCE. Los POI's (puntos de interés) y los eventos están definidos en términos de sus características intrínsecas: localización, estilo artístico, etc. Los perfiles de usuario contienen información sobre sus preferencias y datos de su viaje. El concepto central del vocabulario intermedio es "interés". Las preferencias de los visitantes se traducen a un grupo de "intereses", y los POI's y los eventos pueden atraer a gente con determinados "intereses". Esta traducción es capturada en forma de reglas de producción, que se ejecutan utilizando Jena.

- **Buscador Semántico de Trámites:** La Web Municipal presenta en la actualidad una solución que permite elevar la calidad de descripción, publicación y recuperación de la información del Ayuntamiento en Casa basado en la tecnología de la Web Semántica. Esta solución se basa en la construcción de una ontología que representa el conocimiento sobre el dominio de la administración local y permite que las peticiones de los ciudadanos sean interpretadas y entendidas por el ordenador. Evidentemente hay que seguir trabajando en mejorar las ontologías que describen los trámites del ayuntamiento con el objetivo, siempre, de facilitar el acceso a los mismos y mejorar la usabilidad de estos servicios.

- **Uso de Microformatos:** desde el servicio de la agenda se acaba de implantar el uso del microformato hCalendar (gracias al cual herramientas automáticas pueden detectar los actos de la ciudad), y en toda la Web el microformato hCard.

- **Addentra.** Gracias al uso intensivo de fuentes RSS ofrecemos al ciudadano la posibilidad de seleccionar los contenidos que son de su interés mostrándoselos de forma unificada en esta aplicación.

Han sido muy importantes en el desarrollo de estos servicios los convenios establecidos entre el Ayuntamiento de Zaragoza y diferentes instituciones: la Universidad de Zaragoza, Universidad Pompeu Fabra, Consorcio W3C, Fundación CTIC y El Laboratorio Aragonés de Usabilidad (Gobierno de Aragón).

Nos gustaría terminar este artículo destacando el gran esfuerzo que los profesionales de la Oficina Web Municipal y la Dirección General de Ciencia y Tecnología han realizado estos últimos años, con el objetivo de que la Web pueda ser visitada y utilizada de forma

satisfactoria por el mayor número posible de personas, independientemente de las limitaciones personales que tengan y de aquellas que se deriven de su entorno.

Nuestro objetivo se resume en una Web para todos, una Web accesible e independiente de dispositivos, como queda recogido en los compromisos de la **Carta de Servicios de la Web** aprobada en Pleno Municipal. Quisiéramos también destacar la colaboración activa en este proyecto de los diferentes departamentos y servicios del Ayuntamiento de Zaragoza.

6. Reconocimientos obtenidos

2001-2007: - En las Radiografías que anualmente ha realizado CiberPaís (Suplemento Tecnológico elaborado por el Diario El País) sobre las Webs Municipales españolas, la Web del Ayuntamiento de Zaragoza ha obtenido en dos años el 1er. puesto, en otros dos años el 2º puesto, y nunca ha bajado del 4º puesto.

2004: La Web Municipal recibe el premio "emobility" en SIMO TCI por su proyecto "Zaragoza Ciudad Móvil". Dentro de este proyecto destacamos el servicio de envío de niveles de polinización de la ciudad de Zaragoza al ciudadano a través de SMS.

2006: Zaragoza.es obtiene el premio TAW a la Web pública más accesible de la Administración Local, y en 2007 obtiene uno de los primeros certificados oficiales de accesibilidad otorgados por AENOR (*Asociación Española de Normalización y Certificación*).

2007: Artículo sobre las webs municipales realizado por la Organización de Usuarios y Consumidores (OCU) en su publicación Dinero y Derechos en 2007 [8]. Recoge una prueba práctica en 20 webs municipales para comprobar los servicios que ofrecen al ciudadano. En este informe el Portal de Medio Ambiente recibe la calificación de muy buena en información y usabilidad colocándose en un segundo puesto en la valoración global.

2008: En el artículo elaborado por la consultora Biko 2 y publicado en la revista "Computer hoy, nº 265":

- La lista de evaluaciones sería ésta: Barcelona (9,67 puntos); Zaragoza (9,59), Madrid (9,55) Bilbao (7,93), y Valencia (con 7,32).

- Atendiendo a la usabilidad, la Web del Ayuntamiento de Zaragoza se sitúa en primer lugar con una puntuación de 9,81 sobre los 10 puntos.

- Entre los elementos de www.zaragoza.es más valorados destacan los servicios de Blogs, la Personalización con la herramienta Addentra, Zaragoza Ciudad Móvil (para servicios a través de móviles), Servicio de Quejas y Sugerencias e Información medioambiental, calificada como excelente y exhaustiva.

Referencias

- [1] Javier Nogueras-Iso, Miguel Ángel Latre-Abadía, Pedro R. Muro-Medrano, F. Javier Zarazaga-Soria. Building eGovernment services over Spatial Data Infrastructures. *Lecture Notes in Computer Science (LNCS) 2004*, vol. 3183, pp. 387-391.
- [2] Lars Bernard, Alessandro Annoni, Ioannis Kanellopoulos, Michel Millot, Javier Nogueras-Iso, Joanna Nowak, Katalin Toth. What Technology does INSPIRE need? - Development and Research Requirements. *Proceedings of AGILE 2005: 8th Conference on Geographic Information Science. 2005*, pp. 333-340.
- [3] M^a Jesús Fernández, P. Álvarez, F.J. López, P.R. Muro. IDEZar: un ejemplo de implantación de una IDE en la administración local. *Actas de las IX Jornadas sobre Tecnologías de la Información para la Modernización de las Administraciones Públicas* (Tecnimap 2006). Sevilla, España, 30-may al 2-jun, 2006.
- [4] W3C. Web Content Accessibility Guidelines 1.0. <<http://www.w3.org/TR/WAI-WEBCONTENT/>>.
- [5] Ignacio Marín, Rodrigo García, Luis Miguel González, Julio Alberto Argüello, María Jesús Fernández, Víctor Morlán. Adaptación a dispositivos móviles de portales web diseñados para navegadores de escritorio. *Actas de las Jornadas sobre Web Móvil (MWeb'07- CEDI 2007)*. Zaragoza, España. 12 y 13 de septiembre de 2007.
- [6] W3C. Mobile Web Best Practices 1.0. <<http://www.w3.org/TR/mobile-bp/>>.
- [7] W3C. Semantic Web Case Studies and Use Cases. <<http://www.w3.org/2001/sw/sweo/public/UseCases/>>.
- [8] OCU. Encuesta sobre las Web Municipales. *Dineros y Derecho, julio de 2007, n. 7*. Organización de Consumidores y Usuarios (OCU). <<http://www.ocu.org/map/src/319544.htm>>.

Notas

- ¹ Ley 11/2007, de 22 de junio, de Acceso Electrónico de los Ciudadanos a los Servicios Públicos, <<http://www.boe.es/boe/dias/2007/06/23/pdfs/A27150-27166.pdf>>.
- ² Organización para la Cooperación y el Desarrollo Económicos (OCDE 2004), cap. 1.
- ³ Utilizando el sistema de gestión de blogs Blogia <<http://www.blogia.com/>>.
- ⁴ Carta de Servicios www.zaragoza.es. Aprobada en Pleno Municipal del Ayuntamiento de Zaragoza 2004.
- ⁵ Eurocities, 2005.
- ⁶ Aplicación desarrollada por la Oficina Web del Ayuntamiento con la colaboración del Dpt. I+D+i de la Fundación CTIC.
- ⁷ <<http://www.zaragoza.es/ciudad/turismo/es/visitar/ontologia.htm>>.

Bibliografía adicional

- **J.M. Alonso**. La Administración Electrónica y la Web ¿una historia de amor?. Comunicación presentada en *Día W3C en España 2007*. Madrid, del 23 de mayo de 2007.
- **P. Alvarez, M.J. Fernández**. IDEZar: un ejemplo de implantación de una IDE Local. Aplicación en la Atención al Ciudadano. Innovación y Conocimiento en la Sociedad Digital. *Internet Global Congress 2006*. Barcelona 29 de mayo-1 de junio de 2006.
- **J.M. Angos Ullate, Isaac Aguaron Bolea, M.J. Fernández Ruiz**. Information Systems of the European Public Administration on the Internet. *ONLINE INFORMATION 96 20th International Online Information Meeting* (3-5 diciembre 1996, Londres).
- **J.M. Angos Ullate, Isaac Aguaron Bolea, M.J. Fernández Ruiz**. Las administraciones Públicas Europeas en Internet. *Information World en español*, vol. 6, n^o 7-8, julio-agosto de 1997.
- **M.A. Amutio Gómez (MAP)**. Administración electrónica, estándares e interoperabilidad. Comunicación presentada en *Día W3C en España 2007*. Madrid, del 23 de mayo de 2007.
- **R. Borge**. La participación electrónica: estado de la cuestión y aproximación a una clasificación. *IDP revista d'Internet, Dret i Política, 2005, Vol. 1*.
- **H. Bouwman et al.** *Information and Communication Technology in Organizations*. Londres: Sage Publications, 2005. ISBN-10: 1412900905.
- **P.F. Brown**. (Presidente del Grupo de Interés sobre Administración electrónica, CEN). ¿Ha perdido la administración electrónica el rumbo? Comunicación presentada en *European W3C Symposium on eGovernment 2007*. Gijón, del 1 al 2 de febrero de 2007.
- **M. Castells**. Proyecto Internet Catalunya UOC. <http://www.uoc.edu/prensa/reportajes/proyecto_internet_cataluna.html> [2007].
- **Ayuntamiento de Zaragoza**. Carta de Servicios. <www.zaragoza.es/cartadeservicios/> [2007].
- **Cerrillo**. *L'Administració electrònica*. Barcelona: UOC, 2006. ISBN-10: 8497884167.
- **Ciberpaís**. I Radiografía de las webs municipales. *El País, 2001*.
- **Ciberpaís**. VII Radiografía de las webs municipales. *El País, 2007, 19 de julio*.
- **Commission of the European Communities**. The role of E-Government for Europe's future. Bruselas. *COM (2003), 567*.
- **W3C**. European W3C Symposium on eGovernment. Gijón, del 1 al 2 de febrero del 2007. <<http://www.w3c.es/Eventos/2007/eGov/>> [2007].
- **Unión Europea**. *e-europe 2005*. Plan para el desarrollo de la Sociedad de la Información en la Unión Europea.
- **Emergia**. *Los portales de las Administraciones públicas en España siguen sin ser accesibles en el Año Europeo de las Personas con Discapacidad*. abril de 2003. <<http://www.emergia.net/investigacion/articulos/20030428.asp>> [2007].
- **Eurocities**. *The European Charter of Rights for Citizens in the Knowledge Society*. <www.eurocities.org/uploads/load.php?file=know_charter_eRights_final-SMupdf> [2007].
- **M.J. Fernández**. Zaragoza Ciudad Móvil: servicios a través de Telefonía Móvil. *4ª Jornada de Telefónica Móviles España*. "El Móvil: Facilitador de la integración social, organizada por la Cátedra de Telefónica Móviles de España. UPM, Madrid, 17 de noviembre de 2004.
- **M.J. Fernández, C. García Fernández**. Política

de Accesibilidad para el Sitio Web del Ayuntamiento de Zaragoza. *TECNIMAP IX Jornadas sobre Tecnologías de la Información para la Modernización de las Administraciones Públicas*, Sevilla, junio de 2006.

- **J. Juris**. Networked Social Movements: Global Movements for Global Future. En: M.Castells (ed.). *The Network Society: A Cross-Cultural Perspective*. Cheltenham: Edward Elgar. Comunicación presentada en el *VIII Internet Global Congress. Barcelona*, del 29 de mayo al 1 de junio de 2006.
- **Gobierno de Aragón**. Laboratorio Aragonés de Usabilidad. <<http://www.laboratoriousabilidad.com/>> [2007].
- **Marín et al.** Adaptación a Dispositivos móviles de portales web diseñados para navegadores de escritorio. *I Jornadas sobre Web Móvil, MWEB 2007 (W3C)* dentro de CEDI 2007 II Congreso Español de Informática organizadas por el Departamento de Informática e Ingeniería de Sistemas de la Universidad de Zaragoza, Zaragoza, del 11 al 14 de septiembre de 2007.
- **P.F. Márquez**. (Director General para la Modernización de la Administración, Ministerio de Administraciones Públicas -MAP-). Estado actual de la Administración Electrónica. Comunicación presentada en *European W3C Symposium on eGovernment 2007*. Gijón, del 1 al 2 de febrero de 2007.
- **OCDE**. *L'administration électronique: un impératif*. París: OCDE, 2004-2007.
- **J.A. Salvador, J.M. Angós, M.J. Fernández**. Criterios para evaluar la calidad de las fuentes de información en Internet. En: *Scire: representación y organización del conocimiento, 1999, vol. 5, n. 2*, pp. 97-113.
- **J.A. Salvador, J.M. Angós, M.J. Fernández**. Organization of the information about health resources on the Internet. *7th. International ISKO Conference*, Universidad de Granada. 10-13 Julio de 2002. pp. 198-204.
- **E. Turban, J. Lee, D. King, H.M. Chung**. *Electronic Commerce: A Managerial Perspective*. Nueva York: Prentice Hall, 2000.

Javier Fabra Caro, Pedro Álvarez Pérez-Aradros, José Ángel Bañares Bañares, Joaquín Ezpeleta Mateo
Instituto de Investigación en Ingeniería de Aragón, Departamento de Informática e Ingeniería de Sistemas, Centro Politécnico Superior, Universidad de Zaragoza

<{jfabra, alvaper, banares, ezpeleta}@unizar.es>

Integración en DENEb de componentes para la conectividad dinámica de los procesos Web. Aplicación a escenarios de gestión de emergencias basados en Sensor Web

1. Introducción

Actualmente, las organizaciones especifican *a priori* los protocolos de interacción que regulan la integración de sus procesos de negocio. Esta estrategia garantiza el éxito de la colaboración global si los procesos de cada organización individual han sido posteriormente construidos en base a las restricciones del protocolo.

Por desgracia, en otros escenarios las organizaciones implementan sus procesos de negocio de manera individual. En este caso, cada proceso establece de manera local sus propias restricciones de interacción. Por tanto, para garantizar la integración de un conjunto de procesos que requieren colaborar es común tener que recurrir a la utilización de técnicas de mediación que garanticen la compatibilidad de sus restricciones.

Evidentemente, ambas estrategias tienen un claro carácter estático desde el punto de vista de las interacciones. Los procesos de negocio tienen programadas como parte de su implementación, entre otras restricciones de interacción, qué mensajes intercambian y en qué orden envían o reciben los mensajes. Por tanto, están implementados para colaborar en base a un protocolo concreto.

Sin embargo, los requisitos a nivel de integración de la siguiente generación de procesos de negocio (y aplicaciones Web, en general) exigen formas mucho más flexibles de integración [1][2]. Por ejemplo, los procesos deberían ser capaces de decidir en tiempo de ejecución con quién colaboran y cómo colaboran, adaptándose dinámicamente a las nuevas políticas y estrategias de negocio y respondiendo de manera eficiente a los cambios y eventos inesperados [3]. Este requisito choca frontalmente con las soluciones actuales, dado que los procesos tendrán que colaborar conforme a distintos protocolos durante su ejecución y, probablemente, conforme a protocolos que *a priori* desconozcan. Un ejemplo de dominio de problema donde esta flexibilidad es necesaria es la gestión de emergencias.

Las tecnologías semánticas y determinadas aproximaciones de diseño de procesos, como

Este artículo fue seleccionado para su publicación en *Novática* entre las ponencias presentadas a las IV Jornadas Científico-Técnicas en Servicios Web y SOA (JSWEB2008) celebradas en Sevilla y de las que ATI fue entidad colaboradora.

Resumen: *las futuras plataformas de ejecución de procesos de negocio deberán proporcionar un alto grado de flexibilidad en la integración de sus procesos. Éstos deberían ser capaces de colaborar en base a protocolos de interacción consensuados y no consensuados, e incluso de configurar en tiempo de ejecución el estilo de interacción, la tecnología de comunicación y el formato de codificación de los mensajes intercambiados como parte de sus interacciones. En este artículo se presenta una extensión de la plataforma DENEb (a platform for the Development and Execution of WEB processes) para el desarrollo y ejecución de procesos de negocio. Esta extensión proporciona el soporte necesario para satisfacer los requisitos a nivel de integración previamente descritos, convirtiendo a DENEb en una plataforma adaptable, flexible y muy adecuada para entornos de procesos complejos y dinámicos. La aplicabilidad de las soluciones propuestas ha sido probada en un escenario de prevención de riesgos de incendio basado en las directrices de la iniciativa Sensor Web.*

Palabras clave: *cooperación dinámica entre procesos, integración de procesos de negocio, protocolos de interacción, Redes de Petri.*

por ejemplo el enfoque conversacional, constituyen unos prometedores primeros pasos hacia la consecución de la flexibilidad requerida a nivel de integración [4][5]. No obstante, nuevos modelos de procesos de negocio y, sobre todo, entornos operativos que faciliten el desarrollo y la ejecución de esta nueva generación de procesos son aún necesarios. Un ejemplo de esta siguiente generación de entornos es DENEb (a platform for the *Development and Execution of WEB processes*), cuya descripción e implementación publicamos en un trabajo anterior en *Novática* [6].

DENEb está basada en el paradigma de las *redes-en-redes* y, en versiones anteriores, los procesos colaboraban en base a protocolos de interacción previamente consensuados.

En este trabajo presentamos un mecanismo para ir un paso más allá en el incremento de la flexibilidad, permitiendo a los procesos DENEb colaborar utilizando protocolos de interacción, estilos de comunicación y mecanismos no consensuados que son conocidos en tiempo de ejecución [7][8]. Esto supone un mayor nivel de flexibilidad con respecto a las actuales plataformas de ejecución de procesos y convierte a DENEb en una plataforma adaptable, flexible y muy adecuada para entornos de procesos complejos y dinámicos.

2. DENEb

En artículos anteriores [6][7] mostramos la arquitectura e implementación de DENEb, una plataforma para el desarrollo y la ejecución de procesos Web basada en el paradigma de *redes-en-redes*.

La idea intuitiva de DENEb es sencilla. De acuerdo al enfoque conversacional, la lógica de negocio de un proceso (*workflow*) se implementa como una red objeto y cada uno de los roles que ejecuta para conversar con otros procesos también se implementan como redes objeto. Estas redes objeto colaboran e intercambian información entre ellas a través de canales.

Por otro lado, el entorno operativo que ejecuta los procesos (los *workflows* y sus conversaciones) se implementa como la red sistema. Cabe destacar que también podrían utilizarse los estándares actuales de servicios Web, tanto en la descripción de los *workflows* (BPEL4WS -*Business Process Execution Language for Web Services*- o BPMN -*Business Process Modeling Notation*-, por ejemplo), como en la de los roles (WSDL -*Web Services Choreography Description Language*-, WSCI -*Web Service Choreography Interface*-), ya que actualmente DENEb soporta la traducción de varios de estos estándares a redes objeto.

A nivel de especificación, el paradigma de las *redes-en-redes* permite abordar los requisitos dinámicos de los procesos Web y ofrecer suficiente capacidad expresiva para modelar *workflows* y protocolos de coordinación complejos. Además, gracias a la herramienta Renew, el formalismo es directamente ejecutable, lo que permite el salto directo desde el modelado hasta la implementación, sin transformaciones intermedias.

sado en los roles) de los protocolos de transporte (HTTP, SMTP, RPC, etc.), formatos de codificación y del estilo de interacción (REST, SOAP, etc.) utilizados para el envío de los mensajes. Este *broker* consta de dos componentes: un repositorio asíncrono de mensajes, basado en RLinda o DRLinda, y un conjunto de mediadores que son ejecutados por su correspondiente motor [6][7][8].

exterior a los procesos ejecutados en DENEb, son recibidos por determinados mediadores, transformados a tuplas FIPA y escritos en el repositorio para que la conversación destinataria los recupere.

Funcionalmente, estos mediadores podrían realizar tareas más complejas, por ejemplo, la mediación semántica de los datos intercambiados en los mensajes o el filtrado de información.

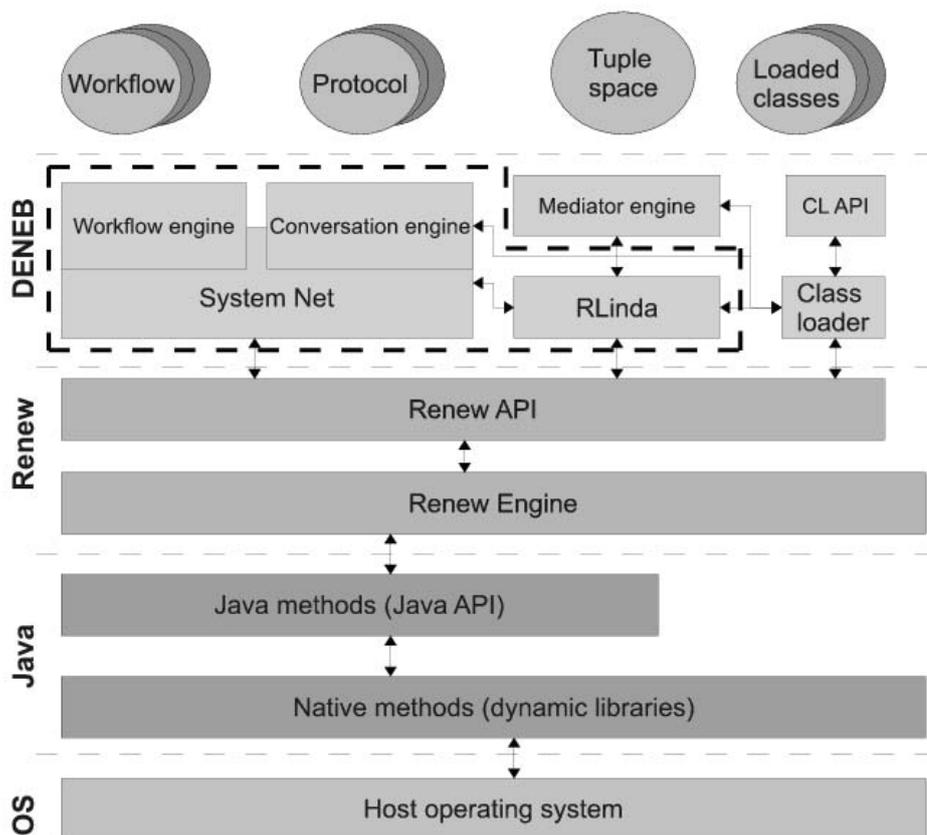


Figura 1. Arquitectura de alto nivel de DENEb.

La **figura 1** muestra la arquitectura de alto nivel de DENEb. La plataforma se ejecuta sobre la herramienta Renew y la máquina virtual de Java.

Desde un punto de vista arquitectural, DENEb consta de tres componentes clave:

- Componente de composición.
- Componente de conversación.
- Broker de mensajes.

El componente de composición gestiona el ciclo de vida de los procesos y ejecuta su lógica de negocio por medio de un intérprete de *workflows*.

Las conversaciones que tienen lugar entre los procesos son ejecutadas por el componente de conversación. Por tanto, este componente es responsable de la gestión de los roles en ejecución.

Por último, el *broker* de mensajes separa la lógica del intercambio de los mensajes (expres-

Los mediadores son los componentes (implementados como redes o como clases Java) responsables de conectar a DENEb con el mundo exterior. Gracias a ellos, los procesos DENEb pueden interactuar con procesos (o aplicaciones, en general) ejecutados en otras plataformas y/o entornos heterogéneos.

Internamente, las conversaciones envían y reciben mensajes expresados en un formato interno dependiente de la plataforma, más concretamente como tuplas FIPA SC00026H (FIPA Request Interaction Protocol Specification <<http://www.fipa.org/specs/fipa00026/SC00026H.pdf>>). Estos mensajes son temporalmente almacenados en el repositorio de mensajes. Los mediadores leen del repositorio los mensajes pendientes de enviar a otros procesos y, utilizando protocolos de comunicación, formatos de codificación y estilos de interacción concretos, los envían a su correspondiente destinatario.

Del mismo modo, mensajes enviados desde el

3. Diseño e integración del componente de conectividad dinámica

En el artículo publicado en *Novática* [6], además del diseño e implementación de la plataforma DENEb y del ciclo de vida de los procesos Web, se analizaron los requisitos a satisfacer para dar soporte a la interpretación de *workflows* y protocolos y toma de decisiones en base a políticas de negocio, lo que proporcionaba la capacidad de configuración en tiempo de ejecución a los procesos Web. Dicha capacidad permitía hablar de aspectos dinámicos, y su implementación se llevó a cabo mediante un componente de carga dinámica de redes y un conjunto de procesos de gestión que se ejecutaban sobre la propia plataforma.

En este trabajo se aborda un nivel más de flexibilidad en los procesos Web dinámicos, integrando en DENEb un nuevo componente capaz de recibir en tiempo de ejecución no sólo nuevos protocolos, sino también mediadores. Utilizando la funcionalidad de este componente, un proceso no tiene que conocer de antemano todos los protocolos que regulan sus conversaciones, sino que podrá intercambiar protocolos con otros procesos en tiempo de ejecución. Esto permite, por ejemplo, que un proceso pueda decidir durante su ejecución con quién colabora o incluso negociar las condiciones que regulan cómo llevar a cabo la colaboración en función de su historial previo de ejecución.

Por otro lado, la posibilidad de incorporar nuevos mediadores permite enriquecer las tecnologías de comunicación, los formatos y los mecanismos de mediación de datos (incluidos los semánticos) y los estilos de interacción soportados por la plataforma.

Estas capacidades que potencian la conectividad de los procesos Web a nivel de protocolo y a nivel de mensajes intercambiados, respectivamente, han sido declaradas como requisitos clave en las futuras plataformas de desarrollo y ejecución de servicios y procesos Web [1][2].

La parte superior derecha de la **figura 1** muestra el *cargador de componentes* (*component loader*, CL), responsable de la funcionalidad previamente descrita. La interacción desde el exterior (por ejemplo, desde una herramienta de administración)

con el componente cargador se realiza a través de su interfaz (CL API).

Por otro lado, la interacción entre el componente y el núcleo de la plataforma DENEb (por ejemplo, para registrar los nuevos protocolos o mediadores) tiene lugar a través de la API de la herramienta Renew.

En primer lugar, el *parser* de PNML comprueba la corrección de la descripción suministrada y, si es posible, genera a partir de ella una red ejecutable por Renew equivalente. Esta red está implementada como una clase Java. Por tanto, utilizando la API de reflexión de Java (*Java API Reflection*) y la biblioteca *Byte Code Engineering Library* (BCEL), los *bytecodes* de

El Centro de Investigación e Información Ambiental de Galicia (*Consellería de Medio Ambiente y Desenvolvemento Sostible* de la Xunta de Galicia <<http://medioambiente.xunta.es/ciiamb.jsp>>) tiene una red de estaciones meteorológicas que monitorizan permanentemente una colección de parámetros atmosféricos y ambientales en distintos puntos del territorio de la comunidad autónoma [9]. La información obtenida por estas estaciones puede ser de interés para predecir ciertos riesgos naturales, por ejemplo, el riesgo de incendios. Si la vegetación de una zona es densa y una estación próxima determina que la temperatura es elevada, la humedad es baja y el viento fuerte, el riesgo de incendio en esa zona y la velocidad de propagación en caso de suceder serían potencialmente altos (este ejemplo es una simplificación de los modelos de análisis que son utilizados para detectar los niveles de riesgo).

El objetivo es construir una aplicación para el centro de gestión de emergencias que notifique a un operador las zonas con un riesgo de incendio potencialmente alto.

Desde el punto de vista del operador, para simplificar su labor, se desea que las alarmas sean representadas sobre un mapa digitalizado de la comunidad autónoma. También es deseable que el operador pueda interactuar con el mapa y acceder a la información disponible sobre el posible riesgo notificado.

Desde el punto de vista de la implementación de la aplicación, se desea que ésta sea accesible vía Web y que su lógica de negocio interna sea implementada en base a procesos de negocio. Estos procesos para realizar su labor necesitarán poder cooperar de forma remota con las distintas estaciones, por ejemplo, para configurar sus planes de monitorización y acceder a las observaciones del territorio obtenidas.

Toda la infraestructura intermedia entre los procesos de negocio de la aplicación y las estaciones físicas deberá estar basada en el paradigma SOA (*Service-Oriented Architecture*), los estándares y tecnologías de servicios Web y los estándares y recomendaciones de la iniciativa *Sensor Web* de OGC (*Open Geospatial Consortium*).

4.1. Diseño arquitectural de la solución
La figura 3 muestra la arquitectura basada en Sensor Web que ha sido diseñada en el marco de este trabajo para el desarrollo de aplicaciones dedicadas a la gestión de emergencias y la prevención de riesgos naturales.

Esta arquitectura consta de tres niveles:

- Nivel de servicios.
- Nivel de procesos de negocio.
- Nivel de aplicación.

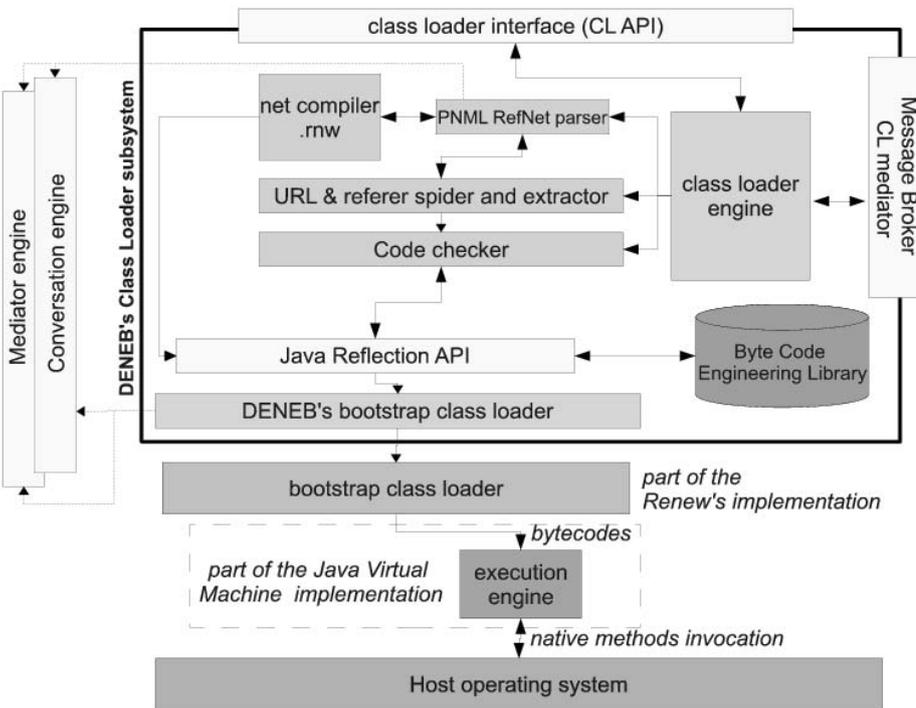


Figura 2. Arquitectura del componente de carga dinámica en DENEb.

La figura 2 presenta el diseño detallado del cargador de componentes. Desde el punto de vista de los protocolos, un proceso externo puede enviar directamente un protocolo descrito en el lenguaje PNML ISO/IEC 15909 (*Petri Net Markup Language*, un lenguaje de descripción de redes de Petri basado en XML) o puede enviar la URL donde está accesible el nuevo protocolo.

La utilización del estándar PNML para el intercambio de roles entre procesos resuelve los aspectos de interoperabilidad relacionados con la descripción de los protocolos.

En cualquiera de los dos casos, el proceso externo envía un *mensaje de control* conteniendo el protocolo o su URL que es almacenado en el repositorio de mensajes de DENEb. El cargador de componentes está monitorizando la escritura de este tipo de mensajes de control. Cuando detecta un nuevo mensaje, lo recupera del repositorio, extrae el protocolo (o lo descarga de la URL correspondiente) y, finalmente, ejecuta una serie de acciones internas para registrar el correspondiente protocolo en tiempo de ejecución.

A continuación se describen brevemente estas acciones:

Java son optimizados y posteriormente cargados en el núcleo de DENEb utilizando la API de Renew. Nótese que la implementación descrita es un prototipo que se asume se ejecutará en un entorno de confianza y, por tanto, no se darán comportamientos maliciosos. Un desarrollo futuro contará con los mecanismos de seguridad adecuados.

Por otro lado, la carga de un mediador es similar a la de un rol de un protocolo. Un mediador puede implementarse directamente como una clase Java o puede implementarse como una red. En este último caso, la red será traducida a la clase Java que la implementa utilizando el mismo procedimiento que para los protocolos. En cualquier caso, la clase resultante que implementa el mediador será traducida a sus *bytecodes* optimizados y registrada a través de la API de Renew en el motor de mediadores. También debemos resaltar que se ha desarrollado un verificador de código básico para un análisis preliminar del comportamiento de los mediadores implementados directamente en código Java.

4. Caso de estudio: integración de servicios Sensor Web en procesos de negocio en base a protocolos no consensuados

El nivel de servicios está formado por el núcleo de servicios Sensor Web y un conjunto de servicios externos accesibles a través de Internet. Los servicios Sensor Web permiten conocer qué sensores están disponibles (más concretamente, el servicio *Sensor Type Catalog*, STC), configurar planes de monitorización en las estaciones remotas (*Sensor Planning Services*, SPS) y acceder a las observaciones de las estaciones (*Sensor Collection Service*, SCS). Un servicio de notificación es el responsable de la coordinación entre estos servicios. Por otro lado, servicios externos de tipo GIS (*Geographic Information Systems*) también integran este nivel inferior. Entre estos servicios caben destacar servidores de mapas (*Google Maps*, por ejemplo) y distintos servidores especializados en proveer información medioambiental en diferentes formatos (*Web Feature Server*, WFS, y *Web Coverage Server*, WCS). Las interfaces de todos los servicios de este nivel están basadas en las especificaciones estándar propuestas por el consorcio OGC y son accesibles a través de los protocolos de transporte (HTTP, SOAP, etc.) y formatos de codificación de datos (XML) típicos de las soluciones Web.

modelo de riesgos y, finalmente, informar al operador a través de la aplicación de notificación en caso de que fuese necesario. Este proceso coopera con los servicios del nivel inferior (reutilizables por cualquier proceso de negocio que requiera de su funcionalidad) en base a los protocolos de interacción publicados como parte de su interfaz. Estos protocolos son conocidos en tiempo de diseño, excepto los protocolos que facilitan a los procesos la interacción con las estaciones para la configuración de sus parámetros de observación. Además, estos protocolos pueden ser distintos para cada estación concreta. El servicio SPS es responsable de facilitar a los procesos el protocolo concreto de la estación con la que desean colaborar.

El nivel de procesos de negocio podría haber sido implementado utilizando el estándar BPEL y alguna de las herramientas e intérpretes basados en este estándar. No obstante, la naturaleza estática de BPEL y la necesidad de integrar en tiempo de ejecución protocolos desconocidos en tiempo de diseño (los protocolos para interactuar con las estaciones) hizo aconsejable la búsqueda de alterna-

el centro de emergencias (por ejemplo, sistemas de información complejos, aplicaciones Web ejecutadas en navegadores, aplicaciones móviles, etc.). En general, estas aplicaciones integran en su lógica interna la funcionalidad ofrecida por los procesos del nivel inferior. Esta integración también se realiza vía Web.

4.2. Implementación del proceso de negocio

Dentro del marco arquitectural descrito anteriormente, en esta sección se presenta un proceso DENEb que explota la conectividad dinámica de la plataforma para la integración de servicios en base a protocolos conocidos en tiempo de ejecución. Más concretamente, se describe la parte correspondiente al acceso a la información de los sensores de una estación dentro de un proceso encargado de la monitorización de las observaciones. Este proceso se encuentra ejecutándose en el nivel de negocio, y persigue la integración de los servicios del nivel inferior.

Tal y como se describió previamente, los protocolos que facilitan a los procesos la interacción con las estaciones para la configuración de los parámetros de las mismas no son conocidos por los procesos, sino únicamente por el SPS (*Sensor Planning Services*). Estos procesos pueden ser heterogéneos, por lo que nuestro proceso de monitorización desconoce *a priori* los detalles del protocolo requerido para la interacción.

La figura 4 muestra cómo el acceso a una estación se realizará en dos pasos. Primero, el proceso de monitorización solicita el envío por parte del SPS del protocolo de interacción de la estación *i*. El SPS envía dicho protocolo, que es procesado por el cargador de componentes según la forma de envío y el tipo de implementación (figura 2), tras lo cual el proceso tiene acceso a la red correspondiente a dicho protocolo.

Así pues, sólo resta poner en ejecución el protocolo de interacción con los datos de configuración del proceso de monitorización. Estos datos van descritos en el estándar propuesto por el SWE (*Sensor Web Enablement WG*), y la interacción con la estación se realiza mediante invocaciones REST (*Representational State Transfer*).

Sin embargo, tanto la especificación de estas llamadas como el encapsulamiento de los datos de configuración en dichas llamadas son propios de cada sensor y, por tanto, desconocidos *a priori* para el proceso de monitorización.

Desde el punto de vista de la implementación, tanto el *workflow* del proceso de monitorización como el protocolo de interacción (una vez procesado por el cargador de componentes) se implementan utili-

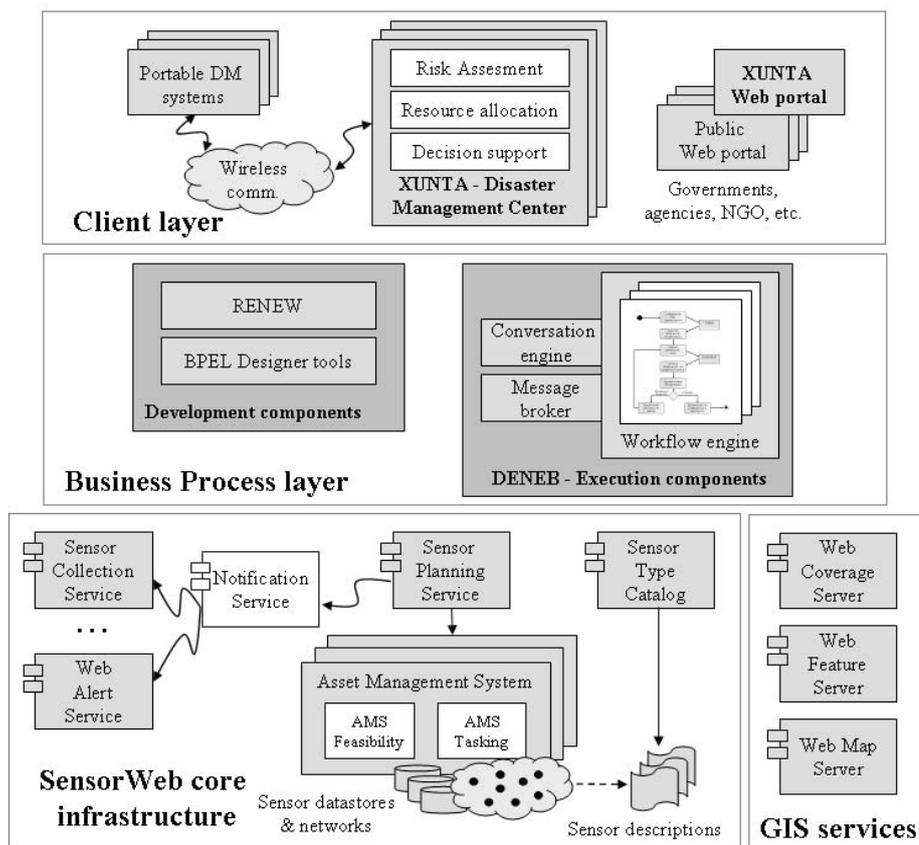


Figura 3. Arquitectura de tres niveles basada en Sensor Web.

En el nivel intermedio de la arquitectura están los procesos de negocio que implementan parte de la lógica de las aplicaciones finales. En nuestro caso concreto, uno de los procesos de negocio implementados es responsable de acceder a las observaciones de las estaciones, analizar si existen posible alarmas en base a un

tivas. En este caso concreto, DENEb ha sido utilizado como entorno de ejecución de procesos y las redes de Petri como lenguaje de implementación de estos procesos.

Por último, el nivel de aplicación está formado por las aplicaciones finales integradas en

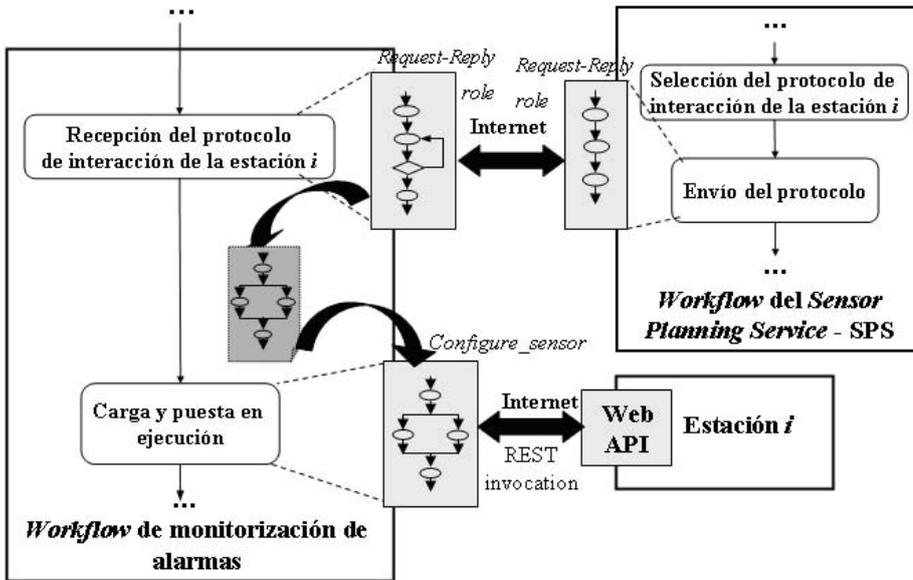


Figura 4. Interacciones en la adquisición del protocolo de acceso a la estación *i*.

zando redes de Petri, lo que produce un modelo ejecutable directamente mediante la utilización de la plataforma DENEb.

La figura 5 muestra los fragmentos de las redes correspondientes a la carga y ejecución del protocolo recibido, cuyo esquema se muestra en la parte derecha de la figura. La parte izquierda de la figura representa la implementación del proceso de puesta en marcha del protocolo recibido en el workflow del proceso de monitorización.

Obsérvese que, tras el proceso de adquisición dinámica llevado a cabo con el SPS, el nombre del rol recibido se almacena en la variable

Station_i_Role, que se utiliza para instanciar el protocolo recibido mediante el disparo de la transición t205 (que corresponde con el disparo sincronizado con la red sistema de DENEb [6]). El disparo sincronizado de las transiciones t206 y t111 a través de la red sistema pone en ejecución la instancia del protocolo recibido para el acceso a la estación *i*.

Una vez en marcha, el workflow y el protocolo intercambian información, como los datos de configuración y la operación a realizar en base al estándar SWE (*Sensor Web Enablement*), mediante el disparo de la transición t207, que implementa el canal *wf-roleInt(...)*.

En ciertas situaciones, este intercambio de información podría llevarse a cabo más de una vez en función de la existencia de subestaciones asociadas a la estación principal, hasta que la transición t210 se sensibilice.

Una vez enviada la información de configuración, el proceso de monitorización espera a que el protocolo de interacción termine mediante el disparo sincronizado de las transiciones t210 y t117, respectivamente. De esta forma, el proceso de monitorización ha conseguido acceder a la estación utilizando su protocolo de interacción específico y desconocido a priori.

La utilización de la plataforma DENEb en el diseño e implementación del SWE abre nuevos retos, como la posibilidad de añadir al sistema nuevas redes de sensores heterogéneas de forma *plug-and-play*, la utilización de protocolos de interacción diferentes con la misma estación o, incluso, el mismo protocolo con varias estaciones diferentes, en el caso de que lo permitiesen.

5. Conclusiones

En este trabajo se ha mostrado un enfoque que mejora los aspectos de interoperabilidad entre procesos y satisface los requisitos de integración de los entornos interorganizacionales dinámicos y evolutivos. Esta propuesta se basa en la descripción adecuada de los servicios y de los protocolos de interacción, separando la lógica de negocio de la lógica de interacción.

La plataforma DENEb aporta a los desarrolladores y usuarios finales una infraestructura ejecutable y una serie de abstracciones para implementar procesos Web y los

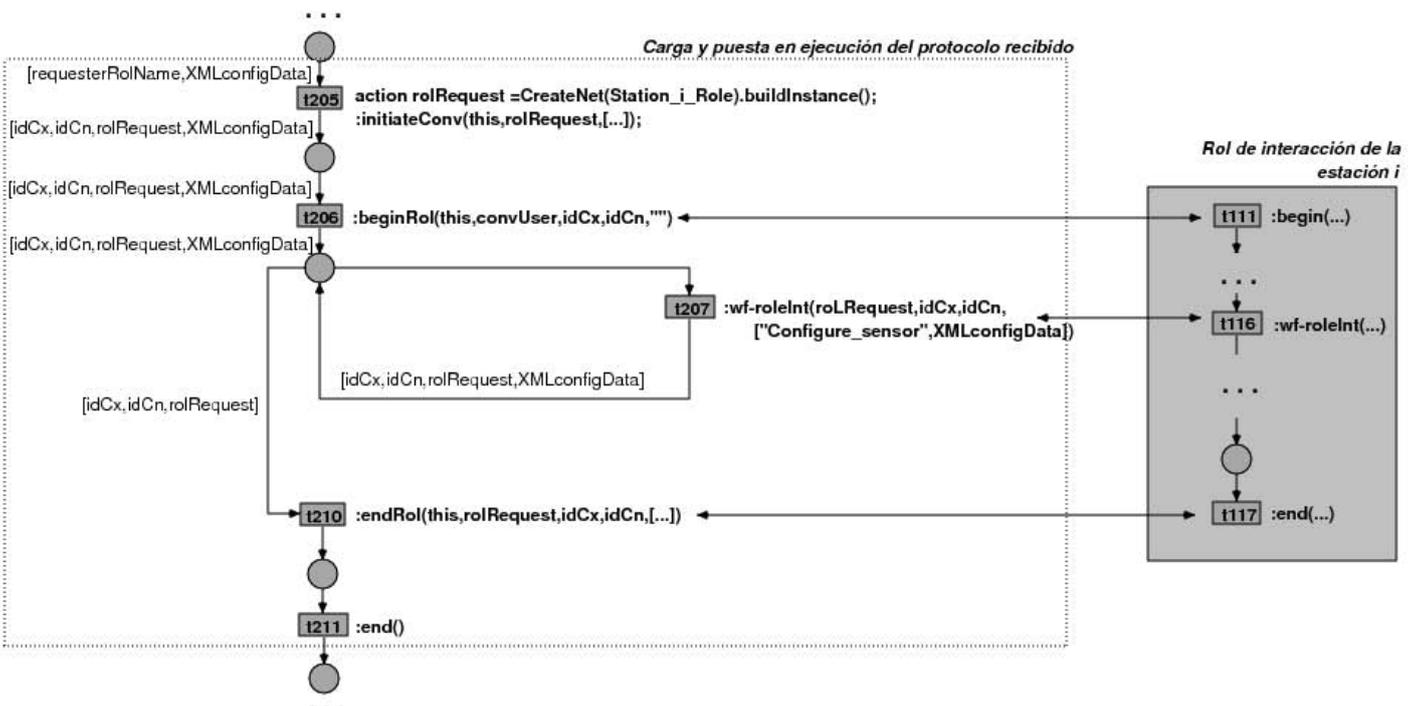


Figura 5. Workflow del proceso de monitorización correspondiente a la carga y puesta en ejecución del protocolo de interacción con la estación *i*.

aspectos relacionados de composición y orquestación dinámicos. DENEb implementa los *workflows* y los protocolos desde la perspectiva del paradigma de las *redes-en-redes* y facilita un mayor nivel de flexibilidad para la integración de servicios.

La flexibilidad de la plataforma DENEb se ha mejorado y extendido mediante el desarrollo de una componente de carga dinámica que no sólo facilita el trabajo con protocolos preestablecidos, sino que también permite que los procesos interactúen utilizando protocolos acordados en tiempo de ejecución.

Esta última característica abre nuevas posibilidades a DENEb, fusionando los aspectos relacionados con la integración de servicios que aporta el enfoque conversacional y la habilidad para llevar a cabo la composición dinámica *ad-hoc* de servicios, tanto para procesos complejos como para protocolos de interacción.

Agradecimientos

Este trabajo ha sido financiado por el proyecto TIN2006-13301 del Ministerio de Educación y Ciencia.

Referencias

- [1] M.P. Papazoglou, P. Traverso, S. Dustdar, F. Leymann. Service-oriented computing: State of the art and research challenges. *Computer* 40, págs. 38-45, 2007.
- [2] Commission of the European Communities. *3S Green Paper on Software and Service Architectures, Infrastructures and Engineering*. Technical report, 2007.
- [3] R. Schmidt. Web services based architectures to support dynamic interorganizational business processes. *International Conference on Web Services - ICWS-Europe'03. Volumen 2853 de LNCS*, Springer, págs. 123-136, 2003.
- [4] J.E. Hanson, P. Nandi, D.W. Levine. Conversation-enabled web services for agents and e-business. *2002 International Conference on Internet Computing - IC'02*, págs. 791-796, 2002.
- [5] L. Ardissono, G. Petrone, M. Segnan. Enabling flexible interaction with web services. En *Extending Web Service Technologies. The use of Multi-Agent approaches*. Springer Verlag, págs. 187-208, Serie: Multiagent Systems, Artificial Societies, and Simulated Organizations, Vol. 13. L. Cavedon, Z. Maamar, D. Martin, B. Benatallah (eds.) 2005. < <http://www.di.unito.it/%7Egiovanna/papers/wsBook.pdf> >.

- [6] J. Fabra, P. Álvarez, J.A. Bañares, J. Ezpeleta. DENEb: Una plataforma para el desarrollo y ejecución de procesos Web dinámicos. *Novática*, num. 192 (marzo-abril 2008), págs. 51-55.
- [7] J. Fabra, P. Álvarez, J.A. Bañares, J. Ezpeleta. Runtime Protocol Binding: Flexible Service Integration By Means Of Flexible Service Interactions. *2008 International Conference on Services Computing - SCC 2008, Lecture Notes in Computer Science*, IEEE Computer Society Press, págs. 291-298, 2008.
- [8] J. Fabra, P. Álvarez, J.A. Bañares, J. Ezpeleta. Integración en DENEb de componentes para la conectividad dinámica de los procesos Web. Aplicación a escenarios de gestión de emergencias basados en Sensor Web. *IV Jornadas Científico-Técnicas en Servicios Web y SOA - JSWEB'08*, págs. 42-55, 2008.
- [9] Consellería de Medio Ambiente y Desenvolvemento Sostible de la Xunta de Galicia. Red de estaciones meteorológicas del Centro de Investigación e Información Ambiental de Galicia, < <http://www.meteogalicia.es/> >.

JORNADAS
CIENTÍFICO
TÉCNICAS
EN SERVICIOS
WEB Y SOA

Madrid, 30 de septiembre a 1 de octubre

JSWEB'09

Patrocinadores

Oro:



Plata:

INTERSYSTEMS



Objetivo

Esta quinta edición de las jornadas persigue consolidar el éxito de las JSWEB de ediciones anteriores, estableciéndose como un punto de referencia de profesionales, empresas e investigadores interesados en el uso y la adopción de servicios. Si bien las jornadas comenzaron centrándose en tecnología de Servicios Web, en la actualidad su ámbito se ha extendido al de la Ciencia e Ingeniería de Servicios, incluyendo también aspectos de tecnologías y plataformas (Servicios Web, Servicios Web Semánticos, Servicios Grid, etc.) así como Arquitecturas Orientadas a Servicios (SOA).

Fecha de recepción de contribuciones: 15 de junio de 2009

Contacto: jsweb09@jsweb.es - Web: <http://www.jsweb.es>

Organizado por:



Universidad
Rey Juan Carlos

Colaboradores:



Joan Baiget Solé

Socio Senior de ATI; Profesor de Teoría y Gestión del Conocimiento en la Universidad Autónoma de Barcelona

<jbaiget@uoc.edu>

1. Introducción

Hace unos años publiqué el artículo "Escenarios del conocimiento: conocimiento orgánico e inorgánico"¹, en donde se presentaba una aproximación del conocimiento en relación al individuo. Durante estos años he ido explicando a mis alumnos de la Universidad Autónoma de Barcelona (UAB)², y en diversas ponencias³, la equivalencia de dichos ámbitos de conocimiento, respecto a la organización, sin haber llegado nunca a publicarlo formalmente, por parecerme obvio y por no contener el grado de detalle que la práctica requiere. Pero tras observar el permanente deambular de la Gestión del Conocimiento en la órbita de las organizaciones, he pensado que un punto de vista más, tal vez podría ayudar a consolidar, al menos en la parte teórica, el entendimiento de esta incipiente disciplina. Este artículo pretende presentar unos macro-ámbitos del conocimiento en la organización, según el esquema previo del conocimiento en el individuo.

2. Antecedentes: Conocimiento e individuo

De manera muy sintética podríamos decir que, como individuos, conocemos el mundo real y lo representamos (o proyectamos) en un mundo simbólico (libro, foto, poesía, pintura...), en un documento (en definitiva) que es su soporte, siendo la palabra el soporte más efímero que usamos, y los soportes digitales, los más novedosos.

De cómo se relaciona el sujeto individual con uno y otro mundo (real/simbólico), en relación al conocimiento, emergen los siguientes 'escenarios': Experimentar, Pensar, Enseñar, Aprender y Aplicar, según el esquema de la figura 1.

De manera que, en sentido amplio, 'experimentamos' el mundo real, 'aprendemos' del

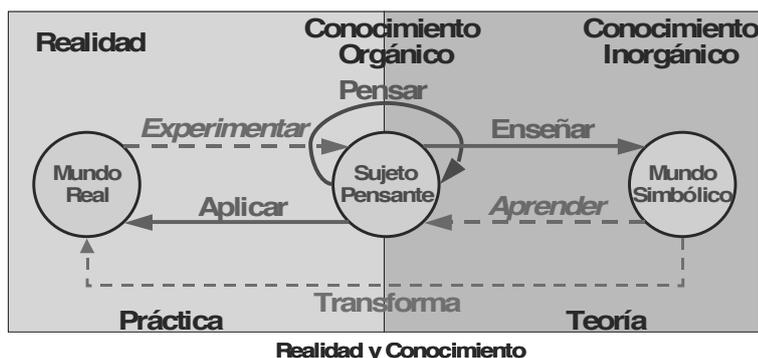


Figura 1. Escenarios del conocimiento (fuente: Joan Baiget).

Escenarios del conocimiento en la organización

Resumen: la práctica de la Gestión del Conocimiento en las organizaciones puede abordarse desde una perspectiva de ámbitos, siguiendo la misma lógica que los ámbitos del conocimiento para el individuo: es decir, la relación de ambos (individuo / organización) con el mundo real y con el mundo simbólico que representa a éste: datos, información, o documentación, de manera genérica. En el individuo estos ámbitos, en sentido amplio, serían: Experimentar, Aprender, Pensar, Enseñar y Aplicar. Y siguiendo la misma lógica, pero para la Organización, estos ámbitos serían: la Investigación, la Innovación, el Desarrollo, la Inteligencia Competitiva (Vigilancia) y la Producción.

Palabras clave: desarrollo, Gestión del Conocimiento, innovación, Inteligencia Competitiva, investigación, producción.

mundo simbólico, 'pensamos' (input/output en el propio individuo), 'enseñamos' (en sentido de externalizar) trasladando nuestro conocimiento al mundo simbólico, y lo 'aplicamos' de vuelta al mundo real, dejando rastros de ello. De estos ámbitos se sigue un simple modelo (ACT) de cómo se produce la **Adquisición**, la **Creación** y la **Transmisión** del

han existido siempre, se han sucedido desde la noche de los tiempos, de manera que este esquema no requiere pagar tributo a ninguna situación actual o coyuntural. No obstante, lo que sí es cierto, es que el mundo ha cambiado, y mucho. Por ello ha variado la forma en que suceden las cosas dentro de los escenarios, pero no han cambiado los escenarios en sí mismos.

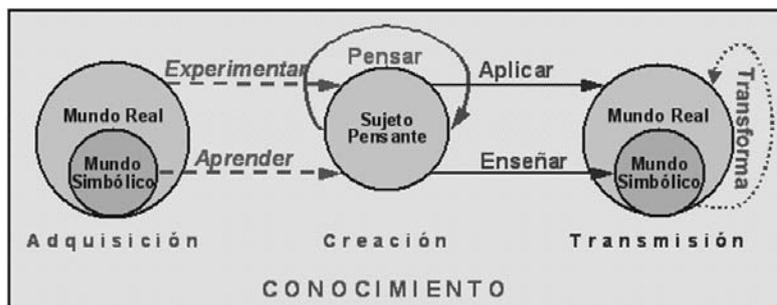


Figura 2. Modelo ACT (fuente: Joan Baiget).

conocimiento en el individuo (ver figura 2). Según esta lógica que proponemos, la 'Adquisición' de conocimiento (input) en el Individuo se realiza 'experimentando' y 'aprendiendo'. La 'Creación' se realiza 'pensando' (input y output), y la 'Transmisión', 'enseñando' y 'aplicando'.

Estos 'escenarios' señalados son 'naturales',

De estos cambios, los más importantes en este contexto, son los experimentados en el mundo simbólico, ya que es ahí donde delegamos nuestros conocimientos *compartidos* sobre el mundo real. Vamos a comentar muy sucintamente estos cambios.

3. Mundo simbólico, tecnología y revolución digital

Durante siglos, el exponente más emblemático del mundo simbólico fue el libro. El libro ha sido la principal tipología de documento en la que se ha confiado el conocimiento existente (obviando la palabra, no por poco importante, sino por efímera). Y las bibliotecas, han sido los almacenes por excelencia de ese conocimiento.

Pero el mundo simbólico ha cambiado de una manera radical en los últimos años: se ha digitalizado, de la mano de la revolución tecnológica. Y esta revolución tecnológica ha permitido la proliferación de conocimientos contenidos en soportes digitales, distribuidos

y accesibles a lo largo de todo el planeta.

De manera que el mundo simbólico ya es (en gran medida) digital, y se gestiona con nuevas tecnologías. Nuevas tecnologías que han permitido convertir, por ejemplo, la comunicación asíncrona en síncrona, ya que hoy es posible realizar una videoconferencia a través de nuestro ordenador personal conectado a Internet, en lugar de enviar una carta que tardaría días, e incluso semanas, en llegar a su destinatario.

Esta comunicación virtual en el mundo, que ha coincidido también en paralelo con una mejora de las comunicaciones físicas (avión, trenes de alta velocidad, ferrys, autopistas...), ha contribuido a la *globalización*, y ello ha redundado en una aceleración y masificación de los procesos de adquisición, generación y transmisión de conocimiento. De manera que en estas circunstancias, se hace inaplazable e inevitable la disciplina de saber gestionar un conocimiento que es cada vez mayor, cada vez más cambiante, y cada vez más crítico. En una organización, esto es una función que debe realizarse en todo su conjunto, y no de manera individualizada.

4. Del individuo a la organización

En el esquema propuesto hasta ahora, los distintos escenarios descritos, han sido posibles porque el individuo es un ser vivo, capaz de percibir, razonar y actuar. Y por ello, es capaz de 'relacionarse' con el mundo real, el mundo simbólico y consigo mismo como sujeto, los tres elementos de la ontología propuesta.

Pero no pasa lo mismo con la organización, que tiene una existencia virtual, y como ser, estrictamente hablando, es inerte. Por lo tanto, la organización tiene la necesidad de crear una estructura para interactuar con el mundo real y con el mundo simbólico.

En los departamentos creados con esa finalidad encontraremos, inevitablemente formalizados, objetivos (por qué), personas (quién), procesos (qué), procedimientos (cómo) y, en buena medida, tecnologías (con qué). El 'donde' y el 'cuando' son las coordenadas espacio-temporales.

Sin querer hacer una aproximación restrictiva de esa relación con el mundo real, y el mundo simbólico, proponemos unos ámbitos o 'escenarios principales' asociados al contexto del conocimiento en la organización. Estos ámbitos serían la **Investigación**, la **Innovación**, el **Desarrollo**, la **Vigilancia** o **Inteligencia Competitiva** y la **Producción** (ver figura 3).

Es evidente que cualquier persona, en cualquier departamento, interactuará tanto con el mundo real como con el mundo simbólico. Pero es a alto nivel, que debemos considerar cuales son los escenarios principales que relacionan la organización con cada elemento

propuesto (mundo real/simbólico), pensando en la organización como un 'ente' global.

5. Conocimiento del individuo y conocimiento de la organización

Antes de entrar en detalle en los escenarios del conocimiento en la organización anteriormente enunciados, conviene hacer una primera consideración acerca de una diferencia substancial entre el esquema presentado para la organización y su precedente, el esquema del sujeto pensante.



Figura 3. Escenarios del conocimiento en la organización (fuente: Joan Baiget).

En ambos modelos hablamos de un 'mundo real' común, como aquello que existe: el único sitio donde está la realidad. Y hablamos de un 'mundo simbólico', también común, donde depositamos el conocimiento acerca de ese mundo real.

Lo que cambia entre ambos modelos es que en el caso del sujeto pensante, existe también un 'conocimiento' en el propio sujeto, en su cerebro, accesible exclusivamente a él. De manera que podemos decir que **existe una sola realidad y su conocimiento está representado en dos planos distintos: en la individualidad de cada sujeto, y en un mundo simbólico, para su uso en común.**

Pero cuando sustituimos en el esquema 'sujeto pensante' por 'organización', ésta última no tiene un 'cerebro propio', de manera que el 'conocimiento de la Organización', está distribuido entre parcelas de la mente de sus miembros (que deben usarlo coordinadamente) y en una parte del mundo simbólico que la organización gestiona internamente. De aquí podríamos pasar a hablar de conocimiento organizacional, conocimiento propio (*firm specific knowledge*), conocimiento interno vs. externo, etc. Pero no es éste el foco del artículo, y sólo nos interesa ahora esta distinción como reflexión previa al comentario de los diversos escenarios del conocimiento en la organización.

6. Escenarios y Gestión del Conocimiento

Una segunda consideración a tener en cuenta, para ambos esquemas (sujeto/organización), es que los escenarios descritos son una extrapolación de los *inputs* y *outputs* entre sus

elementos, pero en la práctica pueden sucederse a la vez, mezclados o combinados. En el caso del sujeto, si pensamos en un aprendiz de pastelero al lado de un experto, éste está 'experimentando' el mundo real, está 'aprendiendo' al escuchar a su maestro lo que le enseña, y está 'pensando' mientras 'aplica' sus incipientes conocimientos, e incluso es posible también que sea capaz, a la vez, de 'enseñar' algo a su maestro. De manera que todos los escenarios se producen conjuntamente. Pero si queremos desarrollar una política en relación al conocimiento, tendremos que

considerarlos de manera individualizada para fomentar cada uno de los ámbitos.

De igual manera, en los ámbitos detallados para la organización, puede producirse esa mezcla y esos solapamientos, pero la conciencia activa de esos flujos principales nos puede facilitar la política y las prácticas para hacer una Gestión del Conocimiento adecuada en cada uno de los escenarios.

Hay que resaltar, por si no fuese obvio, que una política de Gestión del Conocimiento en una organización, en caso de tomar como referencia el modelo aquí presentado, debería aplicarse a los escenarios descritos tanto del sujeto como de la organización.

Así, en una organización es recomendable plantearse todos esos escenarios: ¿Cómo hago que mis empleados ganen experiencia?, ¿Cómo les facilito que piensen y sean creativos?, ¿Cómo lo hacemos para que nos enseñen sus conocimientos y contribuyan a explicitarlos en documentos?, ¿Cómo es la política de formación, como aprenden?, ¿Cómo aplican sus conocimientos?.

De igual modo, a nivel de organización: ¿Cómo investigamos?, ¿Cómo fomentamos la innovación?, ¿Cómo desarrollamos aportando conocimiento al mundo?, ¿Cómo aprendemos, como organización, del mundo simbólico? y ¿Cómo producimos?. Todos esos son escenarios del conocimiento (Adquirir, Crear, Transmitir).

A continuación comentamos 'escenarios principales' propuestos para la organización (ver figura 4)

6.1. Escenario de Investigación

De la misma manera que el individuo 'experimenta' el mundo por sus sentidos, y su conocimiento le permite sobrevivir en éste, las organizaciones deben 'ir' a experimentar el mundo, para entenderlo y así también avanzar y sobrevivir en él.

Investigamos, por ejemplo, la naturaleza de los materiales, para producir nuevas aleaciones de metales. Investigamos en química para tratar los alimentos y hacerlos perdurables en sus envases. Investigamos en microbiología para curar enfermedades, etc. Investigamos, en definitiva para 'conocer', para saber cómo es el mundo real en donde nos encontramos, cómo es la vida, la existencia, y perdurar en este medio como organización, satisfaciendo unos objetivos perseguidos, o utilizando esos conocimientos para poder crear nuevas realidades.

De la misma manera que buscamos entender como es la 'inmutable intimidad de la materia' (objetos del mundo real), también en sentido amplio, podemos investigar cómo son algunos aspectos de la 'cambiante intimidad del individuo' (sujetos del mundo real), cuando nuestro objetivo como organización así lo

Vigilancia). Debemos ser capaces de conocer el entorno, qué sucede en nuestro sector, en la competencia, etc. Y, en un mundo globalizado, una buena forma de hacerlo es acudir al mundo simbólico, donde extensamente está representado el mundo real, y de allí obtener todo el conocimiento que necesitamos. Actualmente existe un enorme desarrollo en este ámbito, y el uso de robots de búsqueda en ese mundo digitalizado ayuda de manera importante a la consecución de esos objetivos, aunque, lógicamente, la función de vigilancia no se circunscribe únicamente a ello.

Vigilar requiere, hoy en día, una permanente mirada digital al mundo simbólico.

6.3. Escenario de la Innovación

Que una empresa innove, significa que hace algo nuevo, lo idea, o lo mejora. Podemos ver la innovación como el equivalente al pensar del individuo, cuando esto le aporta nuevos conocimientos o realidades. De la misma manera que el pensamiento en el individuo se apoya en la experiencia y el aprendizaje, más allá de éste, la innovación en la organización puede apoyarse en la investigación y la vigilancia (Inteligencia Competitiva), *inputs* de co-

output suyo y un *input* en el mundo simbólico, es decir, algo sale de él, de su privacidad, e ingresa en un espacio para compartir.

Análogamente, podemos identificar como un ámbito importante en la organización, el desarrollo, es decir, cuando una organización después de investigar, vigilar (por Inteligencia Competitiva) e innovar (en cuanto a ideas), 'enseña' o tangibiliza los resultados de su conocimiento (los transmite), poniéndolos en el mundo simbólico, como por ejemplo el caso de generar un 'prototipo', que puede acabar con la constitución de una patente. Hemos formalizado en documentos, el conocimiento procedente de la organización, aunque éste restase como confidencial por un tiempo. Al 'desarrollar', transmitimos conocimiento.

6.5. Escenario de la Producción

Finalmente, cuando producimos nuestros productos o servicios, aplicamos nuestros conocimientos al mundo real, y esos cambios en el mundo real, son un rastro de los conocimientos que hemos desarrollado a nivel social. Un motor de un coche es una parte de la realidad que lleva implícito el conocimiento del hombre sobre aleaciones de metales, sobre combustión de carburantes, sobre elementos de electricidad y mecánica, etc.

Cualquier piedra con la que nos tropecemos por un camino, no deja de ser un 'documento' en el que podríamos leer aspectos del pasado. Pero una piedra uniforme no tiene el mismo valor para nosotros que una piedra de sillaría perteneciente a una muralla romana. Ésta última nos aporta implícitamente el conocimiento de la época respecto a la construcción, a la cultura, etc., nos aporta cómo el conocimiento humano transformó la realidad natural.

La producción transforma el mundo gracias a los conocimientos de la organización y deja en él el rastro de dichos conocimientos.

7. A modo de resumen

Una sola realidad: el mundo real. Dos ámbitos de conocimiento de esa realidad: el sujeto pensante (privado) y el mundo simbólico (público).

Individuo y organización se relacionan con esos ámbitos. De ahí salen los distintos escenarios. Una política de Gestión del Conocimiento puede tomar esos escenarios como punto de reflexión y sistematización de sus políticas, que desembocarán en muy diversas prácticas.



Figura 4. Modelo ACT en la organización (fuente: Joan Baiget).

requiera. De manera que si fabricamos camisas, tan importante como la investigación acerca de determinadas fibras o acerca de procesos de fabricación, puede ser la investigación acerca de determinados gustos de la gente, que no dejen de ser parte de ese mundo real.

La investigación se realiza cuando la organización mira hacia la realidad que la rodea, para descubrirla y obtener así un *input* de conocimiento que le sea útil.

6.2. Escenario de la Inteligencia Competitiva (Vigilancia)

Aprender (ámbito en el sujeto), en sentido estricto, es sólo posible en el individuo. Pero la organización debe estar pendiente también de todo el conocimiento que existe en el mundo simbólico, para poder aprovecharlo. Evidentemente este acceso al mundo simbólico va a estar presente en muchos ámbitos (como la formación por ejemplo), pero si existe un dominio de creciente importancia hoy en día, este es el de la Inteligencia Competitiva (o

conocimiento del mundo real y simbólico respectivamente).

Sin pensamiento individual, no habría progreso. Sin innovación organizacional tampoco. El uno precede al otro y ambos transforman la realidad del mundo de manera encadenada. Pero a efectos de una política de Gestión del Conocimiento, debería considerarse prestar atención separada a ambos: por un lado la estimulación del pensar en todos y cada uno de los empleados de la organización; por otro, la formalización organizativa de mecanismos específicamente dirigidos a la innovación de los productos o servicios.

6.4. Escenario del Desarrollo

Cuando el individuo, después de haber experimentado, aprendido y pensado, quiere compartir su conocimiento, lo 'enseña' sacándolo de la intimidad de su pensamiento para enviarlo al mundo simbólico en forma de palabra o documento. Ese mostrar su conocimiento puede ser más novedoso o menos, pero consiste en un

Notas

¹ Joan Baiget Sole. Escenarios del conocimiento: conocimiento orgánico e inorgánico. *Novática* n.º 175, mayo-junio 2005, pp 36-37.

² Asignatura Teoría y Gestión del Conocimiento.

³ Master Latinoamericano de Gestión del Conocimiento, Buenos Aires 2006, 2007, 2008.

¹Eugenio Verbo García,
²Ismael Caballero Muñoz-Reja,
²Ricardo Pérez del Castillo,
²Coral Calero Muñoz,
²Mario Piattini Velthuis

¹Indra Software Labs, Departamento de I+D de Indra Software Labs, Ciudad Real; ²Grupo de Investigación ALARCOS, Instituto de Tecnologías y Sistemas de la Información, Ciudad Real

<emverbo@indra.es>, {Ismael.Caballero, Ricardo.PdelCastillo, Coral.Calero, Mario.Piattini}@uclm.es

1. Introducción

Levis et al. en [26] mencionan ejemplos de escenarios en los que datos con niveles de calidad inadecuados originan problemas que afectan negativamente a los Sistemas de Información (SI) y, por tanto, al rendimiento organizacional. Las causas más comunes de esos niveles inadecuados son una serie de obstáculos a lo largo del ciclo de vida de los datos en el SI como los descritos por Strong et al. en [33]. Por su parte, los problemas producidos por esta falta de calidad pueden ser clasificados a diferentes niveles según su naturaleza: técnicos (como los referidos a la implementación de almacenes de datos [29]), organizacional (tales como pérdida de clientes [31], grandes pérdidas financieras [13][27] o incluso insatisfacción de los trabajadores [12][33]) y legales (como la violación del artículo cuarto del Apartado II de la LOPD de 1999).

Con el objetivo de minimizar el impacto negativo de estos problemas en el desarrollo de sus actividades, es fundamental que las organizaciones sean capaces de evaluar si el nivel de "calidad de datos" de sus datos es el adecuado. Esto implica por un lado la definición de medidas de Calidad de los Datos (CD) sobre ellos y por otro el establecimiento de unos rangos de aceptación válidos para los valores obtenidos en esas medidas. De esta forma, y aplicando las técnicas y herramientas clásicas de calidad será posible localizar de forma más eficiente los datos con niveles de CD inadecuados y sus causas.

Sin embargo, debido a la propia naturaleza de los datos, puede ser bastante difícil realizar una definición adecuada de las medidas. Una aproximación que puede hacer más sencillo el proceso de definición de medidas, es la propuesta por el Programa TDQM (*Total Data Quality Management Program*) del MIT (*Massachusetts Institute of Technology*) en [32][37], en la que se aporta una visión desde las teorías clásicas de calidad, en la que los datos pueden ser considerados como la materia prima de un proceso de producción en el que la información es el producto resultante. Teniendo en cuenta esta analogía de información con los clásicos productos manufactu-

Una metodología basada en ISO/IEC 15939 para la elaboración de planes de medición de calidad de datos

Este artículo obtuvo el "Premio al mejor artículo" en las XIII Jornadas de Ingeniería del Software y Bases de Datos (JISBD2008) que tuvieron lugar en Gijón durante el mes de Octubre de 2008. Fue publicado también en las actas de las Jornadas (pp.253-263. ISBN: 978-84-612-5820-8).

Resumen: hoy en día, los datos juegan un papel fundamental en las organizaciones y la gestión de su calidad se está convirtiendo en una actividad imprescindible. Como parte de dicha gestión, y en aras de obtener medidas útiles, las organizaciones necesitan realizar planes de medición de Calidad de los Datos (CD). Estos planes de medición deben hacerse teniendo en cuenta la propia naturaleza de los datos así como los factores organizacionales que afectan al uso de los mismos. Dado que no existen muchos trabajos en la bibliografía con este objetivo, este artículo presenta una metodología, MEPLAMECAL, para elaborar planes de medición de CD. MEPLAMECAL está basada en ISO/IEC 15939 que, a pesar de ser un estándar de calidad de software, consideramos que puede ser aplicado en este contexto debido a las similitudes existentes entre software y datos. La metodología propuesta está compuesta por dos actividades: (1) establecer y mantener el compromiso de la organización en el proceso de medición, y (2) la elaboración del plan de medición. Estas dos actividades se estructuran en tareas para las que se proponen productos de entrada y salida, así como técnicas y herramientas a utilizar, muchas de ellas tomadas de la Ingeniería del Software.

Palabras clave: aprendizaje organizativo, conocimiento, conocimiento colectivo, conocimiento explícito, conocimiento tácito, creación de conocimiento, management, organización, transmisión de conocimiento.

rados se habilita la aplicación de los principios clásicos de gestión de la calidad a la CD como algunos autores, de forma implícita o explícita, habían sugerido ([3][6][26]).

La bibliografía consultada describe algunas metodologías genéricas de evaluación de CD como la propuesta por Lee et al. en [25] (donde no se describen los procedimientos de medición como parte de la definición de las medidas), y otras particulares centradas en los problemas específicos de los escenarios donde se requería la medición como por ejemplo la propuesta por Al-Hakim en [1] para entornos médicos, donde habitualmente las medidas son realizadas *ad hoc*, sin una planificación previa y casi siempre sin tener en cuenta algunos aspectos importantes relativos tanto a la disponibilidad y disposición de los recursos organizacionales para el proceso de medición como a la naturaleza específica de los datos que afectan al propio proceso de medición. Desafortunadamente, esta suele ser una carencia común en los trabajos encontrados.

No obstante, algunos trabajos como los propuestos en [4][17][19][28] analizan varios de los aspectos mencionados, y tratan de generalizar y priorizar sus conclusiones. Sin embargo, y a pesar de haber obtenido un conjunto de conceptos con bastantes elementos comunes, se da la circunstancia de que normalmente han usado diferentes términos para el mismo concepto, lo que dificulta la

puesta en práctica de los resultados que se han obtenido en escenarios distintos de los suyos. Para tratar de paliar esta inconsistencia en la nomenclatura, Caballero et al. en [7] han analizado los términos de los autores más referenciados y han propuesto una terminología unificada de acuerdo con el Modelo de Información de Medición de Software propuesto en ISO/IEC 15939 [23] y la han ampliado con los aspectos específicos de CD. El resultado es un **Modelo de Información de Medición de Calidad de Datos** (*Data Quality Measurement Information Model, DQMIM*).

Un valor añadido de ISO/IEC 15939 es que proporciona una metodología para la definición de planes de medición de software observando aspectos organizacionales. Esta metodología podría ser adaptada y extendida con los aspectos específicos de CD a los que antes se hacía referencia para cubrir el hueco existente en el campo de CD. La principal contribución de este artículo al campo de la CD es **MEPLAMECAL**, una metodología basada en la propuesta por ISO/IEC 15939 para la elaboración de planes de medición de CD.

La definición de la metodología implica no sólo la identificación de las actividades y tareas correspondientes, sino también la identificación de los productos de entrada y salida para cada una de ellas. Además, se identifican, a partir de las utilizadas en el campo de

Ingeniería del Software, algunas técnicas y herramientas útiles para transformar las entradas en salidas. Pensamos que esta estructura facilitará la incorporación de los aspectos de calidad de datos a las buenas prácticas habituales de calidad del software.

El resto del artículo se estructura de la siguiente manera: la **sección 2** presenta un breve repaso al estado del arte de la medición de CD, incluyendo una sinopsis del DQMIM. La **sección 3** describe la metodología en sí misma. La **sección 4** perfila algunas conclusiones y el trabajo futuro; por último, los agradecimientos y referencias se pueden encontrar al final del artículo.

2. Factores que influyen en la medición de la calidad de datos

Hay un gran número de definiciones para el concepto de CD [2]. De cualquier forma, la mayoría de los autores coinciden en que un dato es de calidad cuando es válido para el propósito para el que un usuario de esos datos quiere utilizarlo ("*adecuación al uso*" [2][34]). De aquí en adelante, para ganar generalidad en las explicaciones, se ampliará el término "*usuario*" mediante el término "*implicado*", refiriéndose a *cualquier agente (persona, sistema o proceso) involucrado en el uso de los datos* [37].

La definición anterior de calidad, basada en la adecuación al uso, tiene dos componentes importantes: percepción multidimensional de la calidad y dependencia de contexto. En torno a estos dos componentes, los implicados deben tener en cuenta los factores a los que se hacía referencia en la introducción de cara a la definición de las medidas de CD.

En cualquier caso, los implicados deben primero identificar las razones por las que quieren medir el nivel de CD de los datos usados por alguno de los procesos de la organización. De acuerdo al DQMIM, se puede decir que los implicados podrían querer satisfacer determinadas *Necesidades de Información*.

Una de las estrategias más usadas para tratar el estudio de la percepción multidimensional de la CD es descomponerla en características más pequeñas, igual que ISO/IEC 9126 [22] hace para el software. Estas características se denominan Dimensiones de CD [2][24] aunque DQMIM propone el término "*concepto medible*" en su lugar.

La identificación y definición de estos conceptos medibles de CD es todavía un reto para la comunidad de CD debido a su dependencia del contexto. Una definición adecuada permitiría a los investigadores y expertos entender (y, por tanto, medir) la CD, evitando la ambigüedad en las definiciones, términos y conceptos utilizados. Algunos trabajos, como [2][24][36] son de obligada referencia

para consultar el significado de los conceptos medibles de CD más usados. Para cada escenario en el que se necesita medir la CD de los datos usados, se deben elegir los conceptos medibles más adecuados. El conjunto de conceptos medibles de CD escogidos es conocido como *modelo de CD*. Existen numerosos ejemplos en las referencias bibliográficas de modelos de CD para distintos entornos: salud y asistencia sanitaria [1], militar [5], sistemas de apoyo a la decisión [18], o web [9][14], por citar algunos. Es importante reseñar que ISO está trabajando actualmente en el estándar ISO/IEC 25012 [21], una parte de la familia de estándares SQUARE que propondrá un modelo de CD para SI.

De cualquier forma, y a la espera de que el estándar esté terminado, la clasificación propuesta por Diane M. Strong et al. [34] es la más utilizada (ver **tabla 1**).

Categoría	Conceptos medibles de CD	Descripción
CD intrínseca	Precisión (<i>Accuracy</i>), Objetividad (<i>Objectivity</i>), Credibilidad (<i>Believability</i>), Reputación (<i>Reputation</i>)	Calidad que tienen los datos por sí mismos
CD de accesibilidad	Accesibilidad (<i>Accessibility</i>), Seguridad de Acceso (<i>Access security</i>)	Suministran significado sobre la facilidad de acceso a los datos
CD contextual	Relevancia (<i>Relevancy</i>), Valor añadido (<i>Value-Added</i>), Oportunidad (<i>Timeliness</i>), Completitud (<i>Completeness</i>), Cantidad de datos (<i>Amount of data</i>)	Tratan con el uso de los datos en un contexto
CD representacional	Interpretabilidad (<i>Interpretability</i>), Facilidad de comprensión (<i>Ease of understanding</i>), Representación concisa (<i>Concise Representation</i>), Representación consistente (<i>Consistent representation</i>)	Características de la representación de los datos que los hacen útiles

Tabla 1. Modelo genérico de CD según Diane M. Strong et al. [34].

Cualquier implicado, sea cual sea su rol, necesitará determinar el grado de bondad de un dato con respecto a los conceptos medibles más adecuados para la tarea que se esté realizando. Esta medida dependerá del uso previsto para los datos y de la naturaleza de los conceptos medibles de CD, que determinan el método o la función de medición [23]. Las típicas medidas de CD tienen una escala de ratio con valores suministrados por la función de medición como, por ejemplo, la fórmula propuesta por [2][24] y que se muestra en la **ecuación 1**.

En la fórmula citada, *UnidadesDeDatos* se refiere a las instancias de los atributos medibles. Además hay dos medidas base:

■ *NúmeroTotalDeUnidadesDeDatos*, cuyos valores pueden ser calculados de forma objetiva usando como método de medición el conteo de las instancias de los atributos medibles de las entidades cuya CD está siendo evaluada.

■ *NúmeroDeUnidadesDeDatosQueNoSatisfacenUnCriterio*, cuyos valores también pueden ser calculados de una forma objetiva contando el número de "falsos" obtenido después de comprobar si las unidades de datos satisfacen el criterio.

En algunas ocasiones, para asignar valores a la medición de un concepto medible y poder aplicar el criterio, es preciso completar mediante metadatos el significado de una unidad de datos de acuerdo a ese concepto medible. Para tomar una decisión, se necesita además de un valor para el metadato, una regla que usando los metadatos describa el criterio

de aceptación como, por ejemplo, la pertenencia del valor del metadato a un dominio determinado. [28] identifica como fuentes de valores para metadatos a los propios implicados (se considera que normalmente proporcionaría un valor subjetivo), al proceso de producción de la información o incluso al mismo almacén de datos.

Para hacer repetibles los procesos de medición, es necesario que el valor del metadato quede adjuntado al dato que completa en el modelo de datos del SI. En [38], se propone una solución para el modelo relacional mediante el etiquetado de los datos como si fueran atributos relacionales. [7] propone un esquema XML llamado DQXSD que permite añadir etiquetas a ficheros XML, [8] incluso

$$CD_{Medida} = 1 - \frac{\text{NúmeroDeUnidadesDeDatosQueNoSatisfacenUnCriterio}}{\text{NúmeroTotalDeUnidadesDeDatos}}$$

Ecuación 1. Fórmula de CD propuesta por C. Batini et al. [2] y Y.W. Lee et al. [24].

propone el uso de Tecnologías Semánticas para realizar este adjuntado, habilitando así a las aplicaciones Web el procesamiento de los aspectos referidos a la medición de CD.

Otro factor importante que debe ser tenido en cuenta, es la posibilidad de limitar el número de unidades de datos en las que el criterio va a ser aplicado [12] de cara a minimizar el esfuerzo computacional invertido en la medición y que consumiría recursos no dedicados al procesamiento de los datos. Así, para mejorar el rendimiento del SI puede ser necesario seleccionar un conjunto representativo de unidades de datos. El número de unidades de datos y su distribución dependerá de la naturaleza de las necesidades de información.

Para no interferir en el proceso de medición, es

necesario fijar en el ciclo de vida cuál es el mejor momento o el mejor intervalo de tiempo para ejecutar el proceso de medición sobre los datos [31].

Como se ha visto a lo largo de este apartado, gestionar todos estos factores puede resultar complicado. Por esto, comprendemos la necesidad de alguna guía para la definición de planes de medición de CD que tenga en cuenta estos factores y su influencia en el proceso de planificación de la medida. A continuación se presenta MEPLAMECAL.

3. Metodología MEPLAMECAL

El objetivo de esta sección es describir de forma resumida los puntos más interesantes de MEPLAMECAL. La metodología se compone de dos actividades, descritas más adelante, divididas a su vez en tareas. Para cada

una de las tareas propuestas, se enumeran tanto los principales productos de entrada como los productos de salida esperados. También se propone el uso de técnicas y herramientas (procedentes en su mayoría del campo de la Ingeniería del Software) para obtener los productos de salida, aunque la elección depende de las preferencias de cada organización. Con el objetivo de facilitar el uso de la metodología, también se identifican los implicados que deberían participar en la ejecución de cada tarea. Las actividades se etiquetan con COM y con EPM, y para las tareas se añade un número que indica el orden en el que deberían ser realizadas. En la **figura 1** se muestra un diagrama que resume las actividades y tareas de la metodología, junto con su orden de ejecución y los implicados en cada una de ellas.

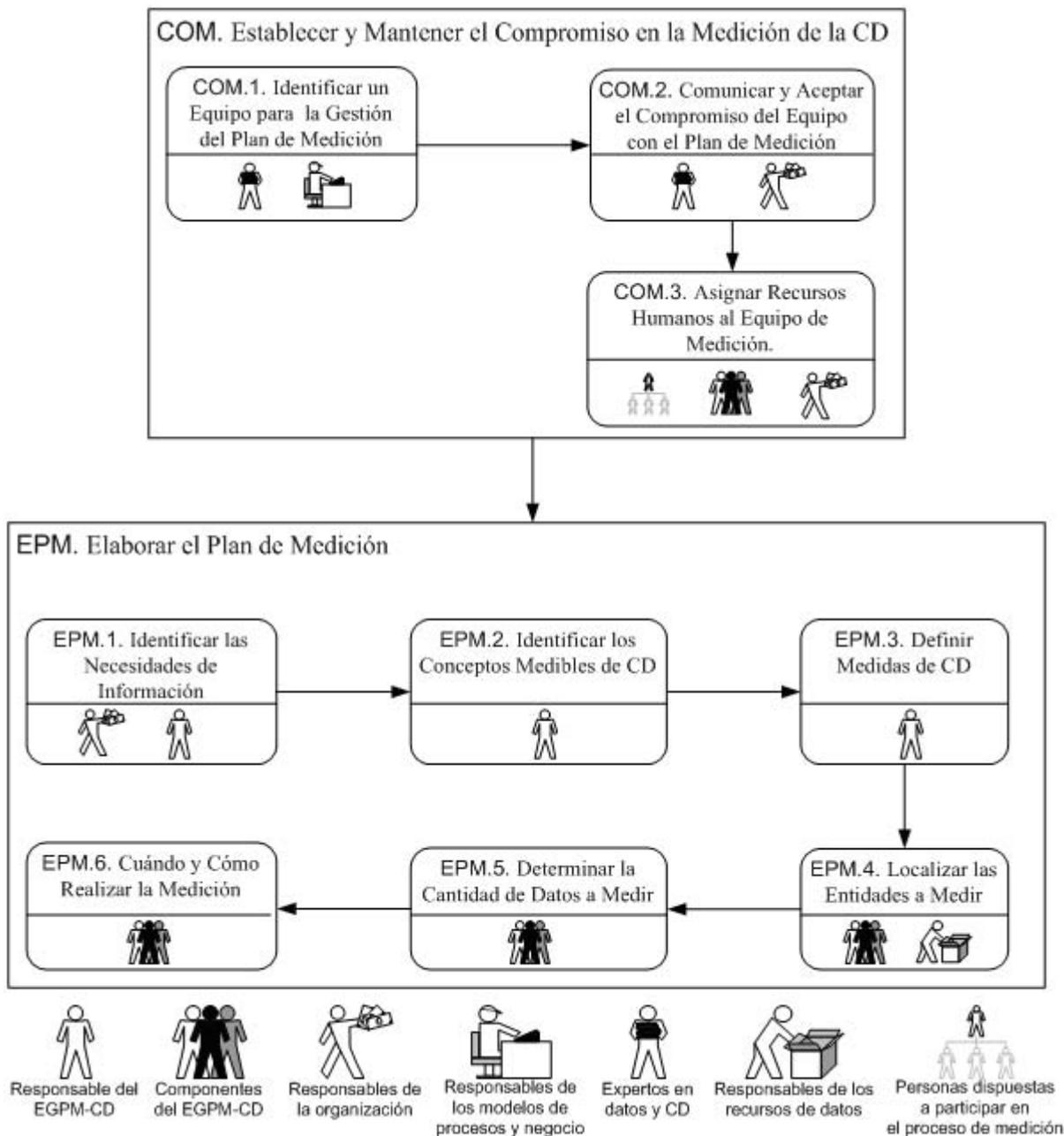


Figura 1. Diagrama de la metodología MEPLAMECAL.

3.1. COM. Establecer y Mantener el Compromiso en la Medición de la CD

Los expertos en calidad coinciden en la necesidad de delegar la responsabilidad de la gestión de la calidad a un grupo de gente comprometida de la organización [11]. Para dar soporte a esta necesidad, uno de los objetivos de la metodología es esbozar un equipo multidisciplinar de trabajadores, que puede responsabilizarse de la realización del plan de medición de la CD. Para satisfacer el objetivo propuesto en esta actividad, se deben ejecutar las siguientes tareas.

3.1.1. COM.1. Identificar un Equipo para la Gestión del Plan de Medición

Este equipo será el responsable de coordinar los esfuerzos y recursos disponibles de la organización de cara a la planificación de la medición de la CD. El **Equipo de Gestión del Plan de Medición de la CD (EGPM-CD)** debería estar compuesto por roles con responsabilidad directa sobre los datos y su integridad, de forma que se pueda determinar quién está usando los datos y para qué propósito, con el fin de valorar la naturaleza y alcance de cualquier deficiencia que pueda existir, así como evaluar el impacto que los problemas relacionados con la CD puedan ocasionar. En consecuencia, si se decide externalizar el proceso de medición, el EGPM-CD no debería estar formado exclusivamente por personas externas a la organización sino que sería necesaria la cooperación entre componentes externos e internos para la elaboración del plan.

Productos	Entrada	- Lista de personas que gestionan o trabajan con los datos de la organización.
	Salida	- Propuesta de Equipo de Gestión del Plan de medición (EGPM-CD).
Herramientas y técnicas		- Entrevistas - Sesiones de trabajo.
Implicados		- Expertos en datos y CD. - Responsables de los modelos de procesos y negocio de la organización.

Tabla 2. Artefactos para COM.1.

Dicho equipo debe ser multidisciplinar, cohesionado, sus miembros deben complementarse en conocimientos y cualidades, y tener capacidad de crítica a su propio trabajo. Además deben disponer de un entorno operativo donde se identifique un método de trabajo con formas efectivas de comunicación, así como contar con un catálogo de técnicas y herramientas útiles para cada tarea. La **tabla 2** nos muestra los principales artefactos para esta actividad.

3.1.2. COM.2. Comunicar y Aceptar el Compromiso del Equipo con el Plan de Medición

Una vez identificado el EGPM-CD, se debe llevar a cabo una segunda tarea de asignación de personas a los roles componentes del EGPM-CD. Sería interesante llevar a cabo entrevistas y

sesiones de trabajo con los candidatos para comprobar su disposición, adecuación y, llegado el caso, su compromiso en la elaboración del plan de medición. La **tabla 3** nos da más detalles acerca de los artefactos de esta actividad.

Productos	Entrada	- Propuesta de Equipo de Gestión del Plan de Medición (EGPM-CD). - Organigrama de la organización.
	Salida	- Compromiso de los componentes de la organización para participar en la elaboración del plan de medición. - Documentos de carga de trabajo.
Técnicas y herramientas		- Entrevistas. - Sesiones de trabajo. - Compromiso.
Implicados		- Expertos en datos y CD. - Responsables de la organización.

Tabla 3. Artefactos para COM.2.

3.1.3. COM.3. Asignar Recursos Humanos al Equipo de Medición

Tras conseguir el compromiso de los componentes del EGPM-CD propuesto, hay que estudiar su disponibilidad y restricciones temporales. Para ello, se deberían realizar entrevistas con los responsables de la organización para identificar posibles solapamientos con otros proyectos de la organización que pudieran malograr la ejecución del plan de medición. Algunas herramientas útiles para conseguir este objetivo son los diagramas de Gantt

o PERT. Por otra parte, se deberían llevar a cabo sesiones de trabajo y entrevistas para detectar y clasificar habilidades individuales de los componentes del EGPM-CD y, dependiendo de ellas, asignar a cada participante las tareas en las que sería más útil. La **tabla 4** nos muestra los artefactos correspondientes.

3.2. EPM. Elaborar el Plan de Medición

El principal objetivo de esta segunda actividad es esbozar el Plan de Medición de CD teniendo en cuenta todos los factores mencionados en la **sección 2**. Para documentar el plan se usarán los términos proporcionados por la terminología DQMIM. Como resultado se obtendrá un documento con el plan de medición de CD. Para conseguir este objetivo, se propone la realización de las siguientes tareas.

3.2.1. EPM.1. Identificar las Necesidades de Información

Para cada uno de los escenarios donde sea necesaria la medición, es esencial recoger y comprender los requisitos de calidad que

deben satisfacer los datos y comprobar que el funcionamiento actual del SI satisface dichos requisitos [17]. Si se encontraran no-conformidades (niveles inadecuados de CD), el EGPM-CD debe profundizar en las causas que las provocaron para arreglarlas o, al menos, mitigar sus efectos.

La clasificación propuesta en [29] recoge causas comunes relacionadas con los SI que pueden ser utilizadas como guía en la identificación de dichas no-conformidades. Puesto que las necesidades de información no se refieren únicamente a aspectos técnicos sino también a organizativos y de gestión, se deberían tener en cuenta las relaciones de negocio existentes entre los principales usuarios del SI. La **tabla 5** nos muestra los artefactos para esta tarea.

3.2.2. EPM.2. Identificar los Conceptos Medibles de CD

Los conceptos medibles de CD son criterios racionales que representan los requisitos de usuario para juzgar la CD. El EGPM-CD se debe encargar de seleccionar aquellos conceptos medibles que mejor satisfagan las necesidades de información.

Como base para la realización de esta tarea se pueden tomar de la bibliografía aquellos modelos de CD que mejor se adapten al escenario de medición o incluso, si estuvieran disponibles en la organización, aquellos elaborados a partir de experiencias previas. Sobre estos modelos se aplicarán técnicas como tormentas de ideas, sesiones de trabajo, entrevistas o el método Delphi. También se podría utilizar la metodología de Franch y Carvalho descrita en [15] pero adaptada a nuestro campo de calidad de datos.

De todos modos, en nuestra metodología se propone continuar la estrategia propuesta por Strong et al. en [33]. En ella se identifican

los obstáculos de CD más comunes y los conceptos medibles relativos a cada uno de ellos. De esta forma, relacionando los obstáculos que afectan a las necesidades de información definidas anteriormente, se pueden obtener los conceptos medibles relativos a cada necesidad de información.

han sido identificados, es el momento de definir las medidas propiamente dichas, correspondientes a los atributos medibles de las entidades que contienen datos. Tres de las metodologías genéricas más importantes para definir medidas son IEEE 1061, ISO/IEC 15939 y "Goal-Question-Metric" (GQM).

Productos	Entrada	- Componentes del EGPM-CD.
	Salida	- Composición definitiva del EGPM-CD.
Técnicas y herramientas		- Sesiones de trabajo. - Entrevistas. - Herramientas de planificación temporal como los diagramas de Gantt o PERT. - Estimaciones. - Análisis de coste/beneficio.
Implicados		- Responsables de la organización. - Componentes del EGPM-CD. - Personas dispuestas a participar en el proceso de medición.

Tabla 4. Artefactos para COM.3.

Productos	Entrada	- Especificación de requisitos de CD. - Informes de actividad de los SI de la organización.
	Salida	- Lista de necesidades de información.
Técnicas y herramientas		- Entrevistas. - Sesiones de trabajo.
Implicados		- Responsables de EGPM-CD. - Responsables de la organización.

Tabla 5. Artefactos para EPM.1.

Productos	Entrada	- Necesidades de información. - Especificación de requisitos de CD. - Catálogo de conceptos medibles de CD.
	Salida	- Lista de conceptos medibles de CD relevantes. - Lista de entidades a medir.
Técnicas y herramientas		- Sesiones de trabajo. - GQM. - Catalogación. - Lista de obstáculos de Strong et al. en [32]
Implicados		- Responsables del EGPM-CD.

Tabla 6. Artefactos para EPM.2.

Finalmente, hay tener en cuenta que puede haber dependencias entre los conceptos medibles de CD que forman el modelo, tal y como [10][16] han analizado, siendo necesario modelar cómo afectan al proceso de medición. En la **tabla 6**, se resumen los principales artefactos para esta tarea.

3.2.3. EPM.3. Definir Medidas de CD

Una vez que los conceptos medibles de CD

de acuerdo al modelo propuesto en DQMIM, las medidas de CD pueden ser de uno de los siguientes tipos: medidas base, medidas derivadas o indicadores. Es importante resaltar que la definición de un plan de medición puede provocar cambios tanto en el modelo de procesos o modelo de datos del SI que contiene los datos cuya calidad se pretende medir. Un ejemplo de este caso es la inserción de metadatos tal y como se describió en la **sección 2**.

Para cada concepto medible de CD, se debe especificar un procedimiento de medición dependiendo del tipo de medida. Esto lleva aparejado la identificación de una unidad de medida, de una escala y de los correspondientes elementos para completar la definición de la medida. Por ejemplo, para un indicador se debe añadir un criterio de decisión y un modelo de análisis.

Dada la complejidad de la situación, es importante remarcar que el EGPM-CD debe ser multidisciplinar y con habilidades suficientes como para afrontar esta actividad, que puede ser considerada como una de las más importantes de la metodología presentada. La **tabla 7** nos muestra los principales artefactos para esta tarea.

3.2.4. EPM.4. Localizar las Entidades a Medir

Las entidades a medir se pueden encontrar en almacenes de datos de distinta naturaleza como, por ejemplo, bases de datos relacionales o semi-estructuradas, ficheros de acceso secuencial, documentos XML u hojas de cálculo.

Para planificar la medición, es preciso localizar la ubicación de las entidades cuyo nivel de CD va a ser medido. Estas entidades tienen atributos medibles que deben ser inspeccionados a la hora de trazar el plan de medición. Algunos ejemplos de entidades pueden ser esquemas de datos, valores de datos, dominios de datos, reglas de negocio, o interfaces de usuario [7].

Una de las principales técnicas para localizar estas entidades son las sesiones de trabajo con los responsables de los recursos de datos de la organización. En la **tabla 8**, se resumen los principales artefactos para esta tarea.

3.2.5. EPM.5. Determinar la Cantidad de Datos a Medir

Dependiendo del propósito de la medición o la necesidad de no sacrificar el rendimiento del SI, puede resultar necesario delimitar el número de entidades que se deben tener en cuenta para la medición. En tal caso, se debe extraer una muestra estadísticamente representativa de todo el conjunto de entidades y luego extrapolar los resultados.

Los parámetros de la muestra (ratio aceptable de datos no válidos, tamaño de la muestra, tipo de muestreo, máximo y mínimo valor para la aceptación o rechazo) pueden ser calculados de acuerdo a estándares como ISO 2859 o UNE 66020, siempre y cuando las entidades sobre las que se va a medir verifiquen las condiciones y limitaciones impuestas por estos estándares. En la **tabla 9** se pueden ver los artefactos para esta tarea.

3.2.6. EPM.6. Cuándo y Cómo Realizar la Medición

La medición puede no ser una actividad puntual sino que su realización requiera un tiem-

Productos	Entrada	- Necesidades de información. - Conceptos medibles relevantes. - Modelo de datos y procesos.
	Salida	- Lista de medidas a aplicar para cada concepto medible.
Técnicas y herramientas		- GQM.
Implicados		- Responsables del EGPM-CD.

Tabla 7. Artefactos para EPM.3.

po, o simplemente puede ocurrir que alguien esté interesado en estudiar la evolución temporal del nivel de CD de una entidad [31]. También es fundamental considerar la asignación temporal que se hizo durante la tarea COM.3 para evitar colisiones temporales entre los componentes del EGPM-CD que desarrollen otras labores dentro de la organización aparte de su intervención en la planificación de la medición.

4. Conclusiones y trabajo futuro

La medición es una actividad clave en cualquier iniciativa de gestión de la calidad. Dado que las organizaciones han empezado a darse cuenta de que algunos de los problemas que sufren son debidos a niveles inadecuados en la calidad de los datos, han comenzado a dedicar esfuerzos y recursos orientados a gestionar la calidad de los datos usados en sus procesos de negocio.

investigación en fundamentos sólidos, como los estándares existentes para la medición y la calidad del software. Hemos elegido uno de estos estándares, ISO/IEC 15939, como la base para la metodología MEPLAMECAL, presentada en este trabajo. Nuestra propuesta pretende llenar un vacío existente y además complementar a los trabajos de evaluación de CD ya existentes en la bibliografía.

Entendemos que la diferencia fundamental entre medición y evaluación consiste en el enfoque específico de ésta última de determinar la validez y utilidad de los datos dentro de un contexto; mientras que la medición está sólo encaminada a la obtención de valores que se utilizarán en la evaluación, sin dar soporte a juicio alguno. A pesar de que pueda parecer que esta simplificación resta relevancia a nuestra propuesta, no existe en la literatura ninguna iniciativa similar con la suficiente generalidad, y que además permita tener en cuenta las características especiales de los datos así como los aspectos inherentes a la medición de su calidad dentro del contexto organizacional.

La principal contribución de este artículo no es la metodología en sí, sino la ventaja de poder utilizarla como guía eficiente que tiene en cuenta los aspectos anteriormente citados en la planificación de la medición de la CD de las entidades organizacionales que contienen los datos que intervienen en los procesos de negocio.

MEPLAMECAL consiste en dos actividades con sus correspondientes tareas. Para facilitar el uso de la metodología, se han identificado los artefactos para cada una de estas tareas. Dado que la metodología ha sido adaptada a partir del estándar ISO/IEC 15939, muy conocido en el campo de la medición de software, cualquier profesional familiarizado podría aplicar fácilmente la metodología en su propio contexto e ir introduciendo en su catálogo de buenas prácticas los conceptos de calidad de datos.

Productos	Entrada	- Lista de entidades a medir. - Modelo de datos de la organización.
	Salida	- Lista de la ubicación de los repositorios de datos a medir.
Técnicas y herramientas		- Inspecciones. - Sesiones de trabajo.
Implicados		- Componentes del EGPM-CD. - Responsables de los recursos de datos.

Tabla 8. Artefactos para EPM.4.

Por estas razones, es necesario realizar una planificación temporal de la medición de modo que se puedan obtener los valores más significativos, y no se entorpezca el trabajo realizado por los implicados. Algunas herramientas que se pueden utilizar para la planificación temporal son los diagramas de Gantt o incluso, si se pretendiese localizar en determinados puntos del proceso de negocio, se podría utilizar alguna notación de modelado de procesos de negocio, como BPMN [30] o IPMAP, notación específica de calidad de datos desarrollada por Shankaranarayan et al. [32].

Una idea muy interesante sería automatizar los procedimientos de medición de cara a ahorrar recursos necesarios en la ejecución del plan de medición y obtener resultados más fiables.

Para poder interpretar adecuadamente los resultados, es importante describir apropiadamente la forma en que los resultados de la medición serán comunicados como parte del Plan de Medición [20][35].

En la **tabla 10** se recogen los principales artefactos para esta tarea

En este sentido, los investigadores en el campo de la CD están comprometidos con el desarrollo de artefactos que ayuden a las organizaciones a medir su CD. Como el campo de la CD es todavía muy joven, necesita apoyarse en otros más consolidados, como el de la Ingeniería del Software, de modo que se puedan basar los esfuerzos de

Productos	Entrada	- Necesidades de información. - Repositorios de datos a medir. - Coste computacional necesario para medir la colección de datos completa. - Relación esfuerzo-coste para la medición de la colección de datos completa.
	Salida	- Estudio de viabilidad de la medición de la colección de datos completa. - Cantidad de datos a medir para cada entidad.
Técnicas y herramientas		- UNE-EN-ISO66020 / ISO 2859.
Implicados		- Componentes del EGPM-CD.

Tabla 9. Artefactos para EPM.5.

Productos	Entrada	- Lista de medidas a aplicar. - Lista de entidades a medir. - Cantidad de datos a medir para cada entidad. - Planificación temporal de la actividad de la organización. - Ciclo de vida de los datos.
	Salida	- Plan de medición.
Técnicas y herramientas		- Diagramas de Gantt. - BPMN, IPMAP, Diagramas de actividad UML. - Sesiones de trabajo
Implicados		- Componentes del EGPM-CD.

Tabla 10. Artefactos para EPM.6.

Finalmente, es importante resaltar que para conseguir que el proceso de medición sea lo más efectivo posible es necesario que no se vea como una actividad puntual o un proceso aislado, sino como un proceso más que se integra dentro de la actividad diaria de la organización.

Somos conscientes de la importancia de obtener medidas de CD repetibles. En consecuencia, nuestra línea de trabajo actual se centra en dos aspectos: por un lado, validar la metodología mediante su aplicación a casos prácticos, con el objetivo de identificar las necesidades de información que podrían aceptar medidas con procedimientos de medición automatizables; y por otro lado, estamos desarrollando un conjunto de herramientas para la automatización del propio proceso de elaboración de Planes de Medición de CD mediante el uso de Tecnologías Semánticas.

Agradecimientos

Esta investigación es parte de los proyectos ES-FINGE (TIN2006-15175-C05-05), CALIPSO (TIN 2005-24055-E), ambos apoyados por el Ministerio de Educación y Ciencia, y HERMES (TSI-020100-2008-155) apoyado por el Ministerio de Industria, Turismo y Comercio.

Referencias

[1] **L. Al-Hakim.** Procedure for Mapping Information Flow: A case of Surgery Management Process. *Information Quality Management: Theory and Applications*, L. Al-Hakim, Editor. 2007, Idea Group Publishing: Hershey, PA, USA. p. 168-188.

[2] **C. Batini, M. Scannapieco.** Data Quality: Concepts, Methodologies and Techniques. *Data-Centric Systems and Applications*. 2006, Berlin: Springer-Verlag, Berlin, Heidelberg.

[3] **M. Bobrowski, M. Marrá, D. Yankelevich.** A Software Engineering View of Data Quality. *Second International Software Quality in Europe*. 1998. Brussels, Belgium.

[4] **M.S.E. Burgess, W.A. Gray, N.J. Fiddian.** Quality Measures and the Information Consumer. *Ninth International Conference on Information Quality (ICIQ'04)*. 2004. MIT, Cambridge, MA, USA.

[5] **T. Burzynski.** Establishing the Environment for Implementation of a Data Quality Management Culture in the Military Health System. *Third International Conference on Information Quality (ICIQ'98)*. 1998. MIT, Cambridge, MA, USA.

[6] **I. Caballero, Ó. Gómez, M. Piattini.** Getting Better Information Quality by Assessing and Improving Information Quality Management. *Ninth International Conference on Information Quality (ICIQ'04)*. 2004. MIT, Cambridge, MA, USA.

[7] **I. Caballero et al.** A Data Quality Measurement Information Model based on ISO/IEC 15939. *12th International Conference on Information Quality*. 2007. MIT, Cambridge, MA.

[8] **I. Caballero et al.** DQRDFS: Towards a Semantic Web Enhanced with Data Quality. *Web Information Systems and Technologies*. 2008. Funchal, Madeira, Portugal.

[9] **A. Caro et al.** A proposal for a set of attributes relevant for Web Portal Data Quality. *Software Quality Journal*. 2008.

[10] **F. DeAmicis, D. Barone, C. Batini.** An Analytical Framework to analyze Dependencies among data Quality Dimensions. *11th International Conference on Information Quality (ICIQ'06)*. 2006. MIT, Cambridge, MA, USA.

[11] **W.E. Deming.** *Out of Crisis*. 1986, Cambridge: MA: MIT Center for Advanced Engineering Study.

[12] **L.P. English.** Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits. 1999, New York, NY, USA: Wiley & Sons. ISBN-10: 0471253839.

[13] **M. Eppler, M. Helfert.** A Classification and Analysis of Data Quality Costs. *Ninth International Conference on Information Quality (ICIQ'04)*. 2004. MIT, Cambridge, MA, USA.

[14] **M. Eppler, P. Muenzenmayer.** Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology. *Proceeding of the Seventh International Conference on Information Quality*. 2002.

[15] **X. Franch, J.P. Carvallo.** Using Quality Models in Software Package Selection. *IEEE Software*, 2003. 20(1): p. 34-41.

[16] **M. Ge, M. Helfert.** A Review of Information Quality Research. *12th International Conference on Information Quality*. 2007. MIT, Cambridge, MA, USA.

[17] **M. Gebauer, P. Caspers, N. Weigel.** Reproducible Measurement of Data Quality Field. *Tenth International Conference on Information Quality (ICIQ'05)*. 2005. MIT, Cambridge, MA, USA.

[18] **M. Gendron, M.J. D'Onofrio.** Formulation of a Decision Support Model Using Quality Attributes. *Seventh International Conference on Information Quality (ICIQ'02)*. 2002. MIT, Cambridge, MA, USA.

[19] **M. Gustavsson.** Information Quality Measurement. *11th International Conference on Information Quality (ICIQ'06)*. 2006. MIT, Cambridge, MA, USA.

[20] **W.S. Humphrey.** *Managing the Software*

Process. The SEI Series in Software Engineering. 1989, Pittsburgh, PA Addison-Wesley. ISBN-10: 0201180952.

[21] **ISO/IEC.** *ISO-25012, ISO/IEC 25012: Software Engineering - Software Quality Requirements and Evaluation (SQuaRE) - Data Quality Model (Draft)*. 2006.

[22] **ISO/IEC.** *ISO/IEC 9126. Software Engineering-Product Quality. Parts 1 to 4*. 2001: International Organization for Standardization/International Electrotechnical Commission.

[23] **ISO/IEC.** *ISO/IEC 15939. Information Technology - Software Measurement Process*. 2000.

[24] **Y.W. Lee et al.** *Journey to Data Quality*. 2006, Cambridge, MA, USA: Massachusetts Institute of Technology.

[25] **Y.W. Lee et al.** *AIMQ: A Methodology for Information Quality Assessment. Information and Management*, 2002. 40(2): p. 133-146.

[26] **M. Levis, M. Helfert, M. Brady.** Information Quality Management: Review of an Evolving Research Area. *12th International Conference on Information Quality (ICIQ'07)*. 2007. MIT, Cambridge, MA, USA.

[27] **D. Loshin.** Enterprises Knowledge Management: The Data Quality Approach. 2001, San Francisco, CA, USA: Morgan Kaufman. ISBN-10: 0124558402.

[28] **F. Naumann, C. Rolker.** Assessment Methods for Information Quality Criteria. *Fifth International Conference on Information Quality (ICIQ'2000)*. 2000. MIT, Cambridge, MA, USA.

[29] **P. Oliveira et al.** A Taxonomy of Data Quality Problems. *Second International Workshop on Data and Information Quality (in conjunction with CAISE'05)*. 2005. Porto, Portugal.

[30] **OMG.** *Business Process Model and Notation 2*. 2008, Object Management Group.

[31] **T. Redman.** *Data Quality: The field guide*. 2000, Boston: Digital Press. ISBN-10: 1555582516.

[32] **G. Shankaranarayan, R.Y. Wang, M. Ziad.** IP-MAP: Representing the Manufacture of an Information Product. *Fifth International Conference on Information Quality (ICIQ'2000)*. 2000. MIT, Cambridge, MA, USA.

[33] **D.M. Strong, Y. Lee, R. Wang.** Ten Potholes in the Road to Information Quality. *IEEE Computer*, 1997: p. 38-46.

[34] **D.M. Strong, Y.W. Lee, R.Y. Wang.** Data Quality in Context. *Communications of the ACM*, 1997. 40(5): p. 103-110.

[35] **J.J. Ukko, J. Karhu, H. Rantanen.** How to communicate measurement information successfully in small and medium-sized enterprises: a regression model. *International Journal of Information Quality*, 2007. 1(1): p. 41-59.

[36] **R.Y. Wang.** *Information Quality (Advances in Management Information Systems)*. M.E. Sharpe: Saddle River, NJ, April 2005. ISBN-10: 0765611333.

[37] **R.Y. Wang.** A Product Perspective on Total Data Quality Management. *Communications of the ACM*, 1998. 41(2): p. 58-65.

[38] **R.Y. Wang, M. Reddy, H. Kon.** Towards quality data: An attribute-based approach. *Journal of Decision Support Systems*, 1995. 13(3-4): p. 349-372.

Notas

¹ <<http://web.mit.edu/tdqm/>>.

² <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44344>.

Martin Wolpers

Fraunhofer Institut für Angewandte Informationstechnologie (FIT), Alemania

<martin.wolpers@fit.fraunhofer.de>

Metadatos de atención contextualizada en entornos de aprendizaje

1. Introducción

Hoy en día hay información en abundancia. Las nuevas tecnologías como los mash-ups¹ de la Web 2.0 aceleran todavía más el número de fuentes de información y los modos de interaccionar con ellas. Los estudiantes pierden grandes cantidades de tiempo en gestionar continuamente la información digital que bien es encontrada (*information pull*) o bien proporcionada (*information push*) a través de medios electrónicos, por ejemplo, vía Internet. En consecuencia, el aprendizaje se ve severamente obstaculizado por las continuas distracciones en la gestión del contenido proporcionado lo que incrementa la carga cognitiva de los estudiantes más allá de lo estrictamente necesario.

Hacen falta nuevos enfoques para ayudar a los estudiantes, por ejemplo individualizando y personalizando y por lo tanto mejorando la experiencia para cada usuario. Para facilitar dichos enfoques, la información detallada sobre el estudiante debe ser capturada más allá de perfiles de usuario comunes, los así llamados metadatos de atención contextualizada (CAM, *Contextualized Attention Metadata*). Una fuente de CAM son, por ejemplo, las observaciones acerca de la manipulación de información por parte del usuario dentro de entornos de conocimiento y trabajo digital.

Este artículo describe a través de tres escenarios cómo pueden utilizarse metadatos de atención contextualizada en diferentes aplicaciones que tratan con el aprendizaje. La aplicación de CAM ejemplifica cómo el uso de CAM en entornos de aprendizaje puede mejorar el reenfoque del estudiante en la tarea de aprendizaje en lugar de en la gestión de la información.

La **sección 2** describe cómo el aprendizaje y la atención están relacionados, mientras que la **sección 3** refleja la importancia de la relación entre contexto y atención. En la **sección 4** se da un repaso al estado del arte y se esbozan los retos identificados. La **sección 5** ejemplifica cómo enfrentarse a los retos en determinados proyectos. El resumen de la **sección 6** concluye el artículo.

2. Aprendizaje y atención

El aprendizaje en este contexto se ve como dependiente de dos mecanismos distintos de procesamiento, a saber: el suministro de información y los procesos de aprendizaje. El suministro de información es controlado

Este artículo fue publicado en su versión original inglesa en **UPGRADE** (vol. IX, issue no. 3, June 2008), no habiendo sido incluido en la monografía del número 193 de **Novática** (mayo-junio de 2008) por razones de espacio.

Traducción: Ramón López Gadea (Grupo de Trabajo de Lengua e Informática)

Resumen: este artículo presenta la noción de metadatos de atención contextualizada (*contextualized attention metadata, CAM*) en entornos de aprendizaje. CAM describe observaciones acerca de la manipulación de información digital en relación al contexto en el que tuvieron lugar las respectivas actividades. Se ejemplifica el uso de CAM en tres escenarios: (1) utilizando CAM para apoyar el proceso de aprendizaje de empleados en la ejecución de procesos de negocio ágiles, (2) enriqueciendo la descripción del proceso de aprendizaje con CAM y (3) identificando patrones de uso de recursos de aprendizaje en el ámbito de la arquitectura con CAM. CAM ayuda a individualizar la experiencia de aprendizaje proporcionando información detallada sobre la manera que tiene el alumno de tratar con la información digital que puede, por ejemplo, ser utilizada para ajustar la provisión de información a las necesidades de los estudiantes ayudándoles a focalizarse en las actividades de aprendizaje en lugar de en la gestión de la información.

Palabras clave: atención, aprendizaje potenciado por la tecnología, contexto, metadatos de atención contextualizada.

Autor

Martin Wolpers obtuvo un doctorado en Ingeniería Eléctrica e Informática en la Leibniz University Hannover. Lidera el grupo "Contexto y Atención para Entornos de Aprendizaje Personalizados" en el FIT ICON (Departamento de investigación en información contextual del *Fraunhofer Institut für Angewandte Informationstechnologie*) que estudia la identificación de tendencias y objetivos de usuario desde flujos de metadatos de atención contextualizada. Algunas de sus más importantes responsabilidades en proyectos de investigación son el cargo de dirección de proyecto del FP6 EU/IST TEL NoE PROLEARN <http://www.fit.fraunhofer.de/profil/bereiche/icon_en.html> y el liderazgo del equipo de implantación técnica del proyecto EC eContent+ MACE). Su investigación se centra en cómo usar metadatos para mejorar los escenarios del aprendizaje potenciado por la tecnología. Más en detalle, su trabajo se focaliza en los metadatos de atención contextualizada y en la representación del conocimiento en educación. Sus intereses en investigación se extienden al modelado conceptual, bases de datos y extracción de la información.

básicamente por la información que llega al sistema cognitivo humano. Al estudiante se le pide que procese la información de una manera significativa.

Así, en un entorno de aprendizaje dado, el proceso de aprendizaje recae sobre la propia persona en base a la información y conocimientos que ya posee. Por ejemplo, lo que un estudiante ya sabe, sus expectativas, su estado mental, sus experiencias previas, su motivación, etc. Toda esa información ha sido bastante descuidada hasta ahora en el desarrollo de tecnologías de aprendizaje.

Para tratar la información entrante, los humanos utilizan filtros para seleccionar y procesar sólo la que es relevante y descartar la no deseada. El resultado de esta selección es visible en la atención que un estudiante presta a la información.

Consecuentemente, ser capaces de observar y

medir la atención permite un mejor ajuste del suministro de información según las necesidades del estudiante. Por ejemplo, observar su atención permitirá a los sistemas de aprendizaje ayudar a sus usuarios en el tratamiento de las grandes cantidades de informaciones recibidas, priorizándolas y restringiéndolas a aquellas necesarias para la tarea en curso.

3. Contexto y aprendizaje

La atención del estudiante es el resultado de uno de los principales procesos de selección de los humanos para reducir la cantidad de información a lo que se necesita para el proceso de aprendizaje. Un filtro posterior para el suministro de información es el entorno de conocimiento usado por el estudiante.

De esta forma, para ayudar al estudiante en su proceso de adquisición de la información, es necesario explicitar y capturar el contexto en el que opera. Por ejemplo, los sistemas de gestión del aprendizaje desarrollados hasta

ahora (Clix, Blackboard, Moodle...) están altamente especializados en ayudar al usuario en tareas específicas (por ejemplo, adquirir una determinada habilidad o competencia) en entornos específicos.

Consecuentemente, tales sistemas de gestión del aprendizaje proporcionan excelentes fuentes de información contextual en el entorno de aprendizaje del usuario, por ejemplo a qué curso está asistiendo el alumno, qué recursos de conocimiento lee, cuál le queda por ver, qué comunicación se fomenta con los otros estudiantes y profesores, etc.

A pesar de que todo esto ya produce información de mucha utilidad sobre el contexto del estudiante, hay más disponible: el estudiante normalmente opera en un entorno técnico mucho más amplio mientras aprende, por ejemplo, utilizando herramientas internas de no-aprendizaje y medios de comunicación. Capturar tal información extiende significativamente el contexto del estudiante.

Mediante la combinación de la información del contexto con las observaciones acerca de la atención del usuario (en términos de su actividad), los metadatos de atención contextualizada describen cómo un estudiante trata la información específica en contextos específicos. Los datos sobre la manipulación de la información se capturan de observaciones de actividades controladas por el usuario como la manipulación de la información en el ordenador y su localización, y también de observaciones fisiológicas como el movimiento del ojo, conductancia dérmica (relacionada científicamente con los estados de ánimo) y comportamientos en el movimiento del ratón (gestos del ratón).

La idea principal es utilizar las observaciones y el contexto del estudiante para facilitar su aplicación y ayuda, independiente de la tarea, en las experiencias de aprendizaje individualizadas. Esto incluye una definición sólida de contexto y atención contextualizada.

4. Estado del arte en metadatos de atención contextualizada

El actual estado del arte en modelado de usuario se basa en que se utilizan sistemas de minería de datos para extraer datos de naturaleza relativamente simple (por ejemplo, secuencias de elementos que son todos del mismo tipo), y en que los modelos de extracción se orientan a relacionar usuarios con estereotipos predefinidos [1].

Esta situación deriva naturalmente del estado del arte en minería de datos: los métodos de minería de datos más recientes disponibles (1) están limitados a analizar datos con una estructura relativamente simple, y (2) trabajan en un contexto fijo: para clasificación, las clases están predefinidas; para agrupamien-

to, la medida de similitud está predefinida.

Existen numerosas enfoques en la captura de observaciones sobre la atención del usuario; por ejemplo, dentro de los proyectos europeos Nepomuk, Aposdle, Gnowsis así como otros [2][3][4][5][6]. Estos enfoques más recientes intentan observar y seguir al usuario dentro de entornos específicos y restringidos, por ejemplo tiendas on-line o sistemas de gestión del conocimiento. Hacen uso de una definición a priori de posibles estereotipos de usuario, actividades y contenido.

Consecuentemente, los modelos que capturan información contextualizada se limitan en términos de metadatos a los contextos, los usuarios y los contenidos. Permiten la correlación de la información capturada en sus contextos específicos. Para salvar estas deficiencias, y para ampliar el ámbito y la validez de los modelos de usuario, surgieron nuevos enfoques que permiten la continua actualización y el intercambio de modelos de usuarios predefinidos necesarios para ajustarse a sus entornos fijos (ver por ejemplo [7]). Por consiguiente, son habitualmente difíciles de adaptar a cambios en el entorno de los usuarios, a nueva información disponible, etc. Las investigaciones actuales se centran en establecer interoperabilidad entre sistemas adaptativos, por ejemplo dentro del proyecto EU/IST GRAPPLE (Generic Responsive Adaptive Personalized Learning Environment)².

Los datos sobre la experiencia del usuario con contenidos digitales (metadatos de uso o atención) describen los gustos del usuario, sus aversiones, lo que lee, publica y escucha siempre que presta atención a contenidos digitales. Hasta el momento, esta rica fuente de información no ha sido explorada en profundidad.

Los dispositivos que registran servicios tales como Apache o servicios de registro de bibliotecas digitales, siguen la pista de la información a la que el usuario accede, qué errores comete mientras usa los sistemas, o desde qué países accede. Estos datos proporcionan información limitada sobre la intención del usuario.

Además, se proponen diferentes formatos para representar los metadatos de atención, por ejemplo AttentionXML o el Attention Profiling Markup Language (APML)³. Todos los formatos están dirigidos a aplicaciones especiales así como a capturar sólo metadatos de atención específicos dificultando la interoperabilidad entre aplicaciones y sistemas.

A pesar del hecho de que los metadatos de atención son altamente personalizados, los enfoques emergentes más recientes no proporcionan medios convenientes para asegurar la privacidad y la seguridad de los datos.

La investigación en este área indica que una aproximación como K-Anonymity [8] producirá resultados más seguros para el almacenamiento de tales datos, al mismo tiempo que transmisiones cifradas y seguras, por ejemplo vía SSL o servicios web seguros (propuesto por ejemplo por el grupo OASIS) garantizan la privacidad y seguridad para el posible intercambio de metadatos de atención. Por ejemplo, los sistemas de seguimiento de metadatos de atención relacionados con navegación web almacenan la información de forma que otras personas puede acceder a ella sin las medidas de seguridad apropiadas (por ejemplo AttentionTrust) mientras que el propietario ni siquiera puede modificar o borrar los datos.

CAM también proporciona la posibilidad de capturar información de los contenidos que maneja el usuario. Algunos nuevos enfoques por parte de la comunidad de la Web Semántica proporcionan ya los primeros resultados en este área, focalizándose principalmente en clasificar la información utilizando ontologías existentes. Combinadas con información proveniente de puntos de vista más orientados al uso, por ejemplo a qué horas fue utilizado un contenido, cuándo y en qué actividades fue usado, cómo fue clasificado (estructuras de carpetas de archivos y carpetas de correo electrónico), etc. Tales descripciones de contenido forman una excelente fuente para describir tanto el contexto como la atención del estudiante.

Hasta el momento, no se han formulado por completo los fundamentos que describan la recopilación y procesamiento de metadatos de atención contextualizada (CAM). Existen modelos que hacen uso de ciertos aspectos de los CAM, pero ninguno en conjunto se orienta a capturarlos todos para proporcionar una visión integral del usuario y sus actividades. Wolpers et. al [9] describen un primer enfoque de trabajo sobre esquemas de metadatos de atención contextualizada.

5. Aplicación de CAM en escenarios del mundo real

Esta sección describe tres ejemplos incipientes de la aplicación de CAM en escenarios de aprendizaje.

El primero trata del uso de observaciones sobre la ejecución de procesos de negocio, las conclusiones derivadas de ellas y las acciones apropiadas para mejorar la eficacia y eficiencia de la ejecución del proceso de negocio. El enfoque esbozado proporciona una medida apropiada para detectar necesidades de aprendizaje que surgen a través de pequeños cambios funcionales o de procedimiento en los flujos de trabajo. Recomendamos consultar la referencia [10] para una discusión más extensa; aquí proporcionaremos simplemente un extracto resumido.

El segundo ejemplo describe el uso de CAM

para el enriquecimiento de metadatos de recursos de aprendizaje.

El tercer ejemplo describe brevemente cómo se empleará el uso de metadatos en el proyecto MACE para facilitar la clasificación de recursos de aprendizaje arquitectónicos.

Los ejemplos dan una perspectiva del estado actual respecto al empleo de CAM en escenarios relacionados con la educación.

5.1. CAM en la ejecución de procesos de negocio

Las infraestructuras de gestión de procesos de negocio, aprendizaje y conocimiento funcionan bastante bien dentro de su contexto específico. A pesar de ello, tales infraestructuras apenas son capaces de presentar cualquier información precisa sobre el uso del conocimiento en ciertas condiciones de trabajo y necesidades de conocimiento, o la identificación de lagunas en las competencias y en la detección de los objetos de aprendizaje apropiados para satisfacer las necesidades.

De esta manera, la cualificación de los empleados así como la mejora de sus competencias constituyen una importante precondition para una efectiva y eficiente ejecución de procesos de negocio, la realización de la gestión del cambio y su capacidad para anticipar mejor y más rápidamente relaciones causa-efecto de los procesos y cambios en el mercado.

Los metadatos de atención contextualizada pueden ser útiles para dirigir la identificación y provisión de contenidos de aprendizaje adecuados (formales, y especialmente informales) dentro de entornos débilmente estructurados y flujos de trabajo ágiles.

Un proceso de negocio proporciona la información de contexto necesaria para identificar las necesidades de conocimiento y aporta recursos de aprendizaje apropiados respetando los objetivos de aprendizaje de negocio e individuales. Las funciones de los procesos de negocio predefinen la información relevante y los sistemas de información de soporte así como las referencias de habilidades/competencias necesarias para operar en los contextos de negocio dados.

Los objetivos de conocimiento pueden ser obtenidos y las consecuentes acciones formativas puestas en marcha para asegurar (todavía más) el desarrollo de las competencias profesionales de los empleados y sus habilidades en la ejecución de procesos basados en esas condiciones explícitas.

Observaciones recientes indican que la mayor parte de los cambios en los flujos de trabajo suceden gradualmente. Normalmente, las funciones que participan en procesos de negocio no son completamente alteradas sino, en

cambio, modificadas para adecuarse a los nuevos flujos de trabajo. Tales modificaciones conducen a actividades que cambian ligeramente y que por tanto se describen como funciones derivadas.

Analistas de negocio como los del Aberdeen Group afirman que los principales problemas de los cambios de procedimiento o intrafuncionales se refieren a la (falta de) información del empleado implicado con respecto al suministro de acciones formativas rediseñadas.

No obstante, no pueden tomarse medidas formativas adecuadas porque, hasta ese momento, no existen medidas para señalar las causas exactas de mal funcionamiento del flujo de trabajo. En su lugar, a los empleados se les proporciona una formación bastante amplia para afrontar la función derivada completa. El entrenamiento no es eficiente y probablemente tampoco efectivo porque el empleado ya conoce la mayor parte de las actividades relevantes para la función derivada.

La necesidad de formación real de los empleados ha de ser incorporada en las actividades que cambian durante la ejecución de la función derivada teniendo en cuenta al mismo tiempo su experiencia y conocimientos.

Hasta el momento no existe ninguna herramienta informática para identificar o derivar las necesidades de formación de un modo contextualizado. La comparativa de las observaciones con la definición (del tipo) de la función describe la diferencia entre las actividades reales y las requeridas. Consecuentemente, esta brecha define las necesidades de conocimiento del individuo teniendo en cuenta además sus conocimientos y experiencia. Por eso, la combinación de las observaciones acerca de los empleados con el conocimiento tácito sobre el flujo de trabajo y la estructura de la función permite la deducción de las necesidades de aprendizaje.

5.2. Enriquecimiento de metadatos de recurso de aprendizaje basado en CAM

En este ejemplo, la combinación de varios tipos de metadatos se emplea para enriquecer los metadatos de recursos de aprendizaje de cara a permitir una provisión de recursos de aprendizaje más ajustada e individualizada y, por lo tanto, a mejorar la experiencia de aprendizaje.

El ejemplo se basa en la implementación de una prueba de concepto llevada a cabo en colaboración con el *Centre for Artificial Intelligence* alemán (DFKI, *Deutsches Forschungszentrum für Künstliche Intelligenz*), la *Open University* del Reino Unido, la asociación europea *Ariadne Foundation for the European Knowledge Pool* y el *Fraunhofer Institute for Applied Information*

Technology (FIT, *Fraunhofer Institut für Angewandte Informationstechnologie*) de Alemania. La prueba de concepto combina la plataforma social ALOE con el repositorio de recursos de aprendizaje Ariadne y el repositorio FlashMeeting.

El repositorio Ariadne utiliza "metadatos para objetos de aprendizaje" (LOM, *Learning Object Metadata*) generados por expertos para describir los recursos de conocimiento. Específicamente, los campos de metadatos que representan valores semánticos de interés, por ejemplo, título, descripción y palabras clave, así como valores que representan información educacional como la dificultad o el tipo de interactividad.

El repositorio FlashMeeting proporciona descripciones acerca de las grabaciones de conferencias en FlashMeeting, incluyendo el título de la grabación, una descripción, lista de participantes, palabras clave, registro de la conversación, etc. Las descripciones son creadas parcialmente por el organizador de la grabación y parcialmente capturadas de manera automática durante la sesión.

El sistema ALOE [11] es una plataforma social de compartición de recursos que sigue los principios de diseño de la Web 2.0. Permite a los usuarios compartir y organizar listas de marcadores favoritos y recursos digitales de tipo arbitrario. Además, pueden asociarse conjuntos arbitrarios de metadatos con cada recurso. De este modo, ALOE se constituye en un recurso orientado socialmente y en un colector de metadatos.

En este ejemplo, un usuario encuentra un recurso de conocimiento específico del repositorio Ariadne y lo añade a su carpeta personal de recursos en el sistema ALOE. El usuario describe el recurso mediante las respectivas etiquetas. A través del uso de las etiquetas generadas por el usuario y los metadatos del recurso, automáticamente se encuentra y se sugiere al usuario una grabación apropiada de FlashMeeting. Después de visualizar la grabación, el usuario guarda su marcador de FlashMeeting y añade más etiquetas.

Este ejemplo describe cómo los metadatos de recursos de conocimiento pueden ser enriquecidos con valiosas descripciones no disponibles con anterioridad.

Por un lado, el usuario describe los recursos con etiquetas que se añaden como palabras clave en su conjunto de metadatos. Si un determinado número de usuarios emplea las mismas palabras claves para un recurso, aplicando algunos algoritmos simples de agrupación y cuantificación, tales palabras claves pueden agruparse y valorarse para asegurar una medida apropiada de su relevancia, lo que permite mejores clasificaciones.

Por otro lado, el usuario almacena los recursos en su carpeta de marcadores. Este hecho implica que los recursos están relacionados unos con otros. Si un número determinado de usuarios asocian los mismos recursos de conocimiento de la misma forma, las respectivas relaciones son también añadidas al conjunto de metadatos de recursos de conocimiento y así se proporcionan las bases para nuevas medidas de relevancia.

Mediante el uso de CAM, las actividades del usuario sirven para realimentar los repositorios de Ariadne y FlashMeeting para mejorar las respectivas descripciones de metadatos. Además, a través de las observaciones de los tres sistemas, CAM permite derivar conclusiones sobre patrones de comportamiento y, posteriormente, sobre tendencias mientras trata con recursos de aprendizaje incluyendo su proceso de creación dentro de los repositorios.

5.3. Uso de patrones de recursos de aprendizaje arquitectónicos

El proyecto MACE⁴ (*Metadata for Architectural Contents in Europe*) financiado dentro del programa eContent+ de la UE, se orienta a mejorar la educación arquitectónica en Europa proporcionando acceso a objetos de aprendizaje arquitectónicos utilizando nuevos enfoques y tecnologías avanzadas [12].

Los recursos de aprendizaje se hacen accesibles a través de una infraestructura de repositorios de conocimiento federados. Mediante la correlación de varios tipos de contenido, uso y metadatos sociales y contextuales, el proyecto proporciona acceso a contenido arquitectónico de alta calidad a través de múltiples perspectivas e itinerarios de navegación que guían de manera efectiva la multiplicación de experiencias para el estudiante. El proyecto MACE creará un marco sostenible para conectar e intercambiar conocimiento arquitectónico en Europa.

Un punto importante del proyecto MACE es la utilización de metadatos de atención contextualizada para métricas avanzadas y, posteriormente, nuevos enfoques de clasificación de recursos de aprendizaje arquitectónicos. Basadas en el trabajo de Ochoa y Duval [6], se desarrollan métricas para clasificar recursos de conocimiento basados en su uso dentro de la arquitectura MACE, así como de los respectivos repositorios.

Por ejemplo, una sencilla métrica se basa en el número de visionados de recursos de conocimiento. Otra, más avanzada, establece una correlación entre las fechas de uso de unos recursos de conocimiento con otros y así genera un cronograma del uso por temática y recursos de conocimiento relevantes. El cronograma es el primer paso hacia la identificación de posibles patrones y tendencias de uso relevante.

6. Resumen

La observación del alumno en su tratamiento de información digital proporciona una medida de cómo invierte su atención en según qué información. Enriqueciendo estas observaciones con información sobre el contexto se consiguen metadatos de atención contextualizada. Usando CAM, los elementos del proceso de aprendizaje del usuario pueden ser mucho mejor ajustados a sus necesidades tal y como se observa en los escenarios de ejemplo mostrados. Por ejemplo, CAM puede ayudar a controlar la provisión de información dentro del proceso de aprendizaje y de esta manera focalizar al estudiante en el conocimiento en lugar de en la gestión de la información.

Los primeros experimentos dan indicios de que CAM facilita la detección de tendencias y tareas de aprendizaje no descubiertas hasta el momento. La identificación estaría basada en la correlación de observaciones sobre cómo un usuario manipula la información digital en general. Ser capaz de describir tendencias y tareas en términos de CAM mejorará significativamente la experiencia de aprendizaje de los usuarios permitiendo herramientas que directamente materialicen las necesidades del estudiante en lugar de supuestas necesidades tal y como se hace en la mayor parte de los entornos de aprendizaje de hoy en día.

El grupo de investigación "Contexto y Atención en Experiencias de Aprendizaje Personal" (CAPLE, *Context and Attention in Personalized Learning Environments*) del Fraunhofer Institute for Applied Information Technology (FHG FIT) afronta los retos de investigación descritos en un buen número de proyectos. El CAPLE se orienta a mejorar la experiencia de aprendizaje individual combinando conocimiento sobre el usuario, su contexto y recursos de aprendizaje relevantes con enfoques avanzados para un aprendizaje duradero. Se focaliza en la individualización más bien que en la reciente personalización masiva.

Referencias

- [1] N. Henze, W. Nejd. Logically Characterizing Adaptive Educational Hypermedia Systems. *Proceedings of the Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems, Session at WWW2003 conference*, May 2003, Budapest, Hungary.
- [2] H. Holz, O. Rostanin, A. Dengel, A., Suzuki, T., Maeda, K., Kanasaki. Task-based Process Know-how Reuse and Proactive Information Delivery in TaskNavigator. *Proceedings of ACM Conference on Information and Knowledge Management (CIKM) 2006*, Washington, USA, November 2006.
- [3] S. Braun, A. Schmidt. Don't annoy the Informal Teacher: Context-Aware Mediation of communication for Workplace Learning. *Proceedings of the 6th International Conference on Knowledge Management*, Graz, Austria, September 2006.

[4] J. Garofalakis, T. Giannakoudi, E. Sakkopoulos. Semantic Web Site Usage Analysis: The ORGAN System. *Proceedings of the World Wide Web 2006 Workshop Logging Traces of Web Activity: The Mechanics of Data Collection*, Edinburgh, UK, May 2006.

[5] C. Roda, J. Thomas. Attention Aware Systems: Theories, Application and Research Agenda. *Journal on Computers in Human Behaviour*, 22 (2006).

[6] X. Ochoa, E. Duval. Use of contextualized attention metadata for ranking and recommending learning objects. *Proceedings of the 1st international Workshop on Contextualized Attention Metadata: Collecting, Managing and Exploiting of Rich Usage information, CAMA '06*. (2006) ACM, New York, NY, 9-16.

[7] S. Sosnovsky, P. Dolog, N. Henze, P. Brusilovsky, W. Nejd. Translation of Overlay Models of Student Knowledge for Relative Domains Based on Domain Ontology Mapping. *AIED 2007: 13th International Conference on Artificial Intelligence in Education*. LA, US, July 2007, IOS Press 2007.

[8] L. Sweeney. K-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 2002.

[9] M. Wolpers, J. Najjar, K. Verbert, E. Duval. Tracking Actual Usage: the Attention Metadata Approach. *International Journal Educational Technology and Society, Special Issue on «Advanced Technologies for Life-Long Learning»*, 2007.

[10] M. Wolpers, G. Martin, E. Duval. CAM in Process Execution. *Proceedings of the 7th International Conference on Knowledge Management (I-Know), Special track on Integrating Working and Learning (IWL'07)*, Journal of Universal Computer Science, Graz, Austria, September 2007.

[11] M. Memmel, R. Schirru. Sharing Digital Resources and Metadata for Open and Flexible Knowledge Management Systems, en K. Tochtermann, H. Maurer (editors), *Proceedings of the 7th International Conference on Knowledge Management (IKNOW)*, Graz, Austria, Journal of Universal Computer Science, 2007.

[12] M. Stefaner, E. Dalla Vecchia, M. Condotta, M. Wolpers, M. Specht, S. Apelt, E. Duval. MACE – enriching architectural learning objects for experience multiplication. *Proceedings of the 2nd European Conference on Technology Enhanced Learning (EC-TEL 2007)*, Crete, Greece, September 2007, in *Creating New Learning Experiences on a Global Scale*, LNCS, Vol. 4753, ISBN: 978-3-540-75194-6.

Notas

- ¹ Una aplicación web híbrida (*mashup* o *remezcla*), es un sitio web o aplicación web que usa contenido de otras aplicaciones Web para crear un nuevo contenido completo, consumiendo servicios directamente siempre a través de protocolo http. El contenido de un mashup normalmente proviene de sitios web de terceros a través de una interfaz pública o usando un API. Otros métodos que constituyen el origen de sus datos incluyen: sindicadores web (RSS o Atom), *Screen scraping*, etc. <[http://es.wikipedia.org/wiki/Mashup_\(aplicaci%C3%B3n_web_h%C3%ADbrida\)](http://es.wikipedia.org/wiki/Mashup_(aplicaci%C3%B3n_web_h%C3%ADbrida))>.
- ² <<http://www.grapple-project.org/>>.
- ³ El APML es un formato basado en XML para recoger los intereses y las aversiones de las personas < <http://en.wikipedia.org/wiki/APML>>.
- ⁴ <<http://www.mace-project.eu/>>.

Las habituales referencias que desde 1999 nos ofrecen los coordinadores de las Secciones Técnicas de nuestra revista pueden consultarse en <http://www.ati.es/novatica/lecturas.html>.

Sección Técnica "Acceso y recuperación de información" (José María Gómez Hidalgo, Manuel J. Maña López)

Tema: popularidad de las Redes Sociales en España

En los últimos años se ha popularizado el concepto de *Acceso a la Información*, referido al proceso a través del cual los usuarios usan las tecnologías de la información para la búsqueda, comprensión y organización de la información [1]. Si algo ha cambiado con el advenimiento de la Web 2.0, inicialmente representada por los blogs o bitácoras, y por los sitios para compartir fotos y vídeos, y ahora por la Redes Sociales generalistas en su pleno apogeo, es la capacidad de participación de los usuarios en la generación y difusión de contenidos. Las Redes Sociales se han convertido en una de las herramientas por excelencia de organización de información, y por tanto de Acceso a la Información, para los usuarios de Internet.

Después de los muchos y continuos lamentos sobre la brecha digital y la falta de penetración de la banda ancha en España, cabe destacar que se registran casi 18 millones de usuarios de Internet en España, de acuerdo al estudio de comScore [2] en diciembre de 2008. Más aún, un 73% de ellos visitaron las Redes Sociales generalistas en diciembre de 2008, superando en porcentaje a países como Alemania, Francia u Holanda. Quizá este fenómeno comunicativo es el responsable del crecimiento de usuarios de Internet de un 20% en tan solo el año 2008.

De acuerdo a dicho estudio de comScore, la Red Social más popular en España es Tuenti, que cuenta con más de 5 millones y medio de visitantes, seguida por Facebook, que ha experimentado un espectacular crecimiento en el 2008, superando los 4,3 millones de visitantes. Otras redes populares son Fotolog, Hi5 o Metroflog.

[1] **Marti Hearst.** *What is Information Access?*. Lecture Slides from the Course "Current Topics in Information Access". <http://www2.sims.berkeley.edu/courses/is296a-3/f98/lectures/intro/index.htm>.

[2] **comScore.** *Tuenti Most Popular Social Networking Site in Spain.* comScore Press Release, 25/02/2009, <http://www.comscore.com/press/release.asp?press=2733>.

Tema: conferencia

El *Third International PAN Workshop and First Competition on Plagiarism Detection* es un taller que se celebra como evento satélite de la conferencia anual de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN) [1], y se dedica a la publicación de resultados de investigación sobre la detección del plagio, la identificación de autor, y la detección del uso malintencionado del software social. La mayoría de técnicas usadas en estos tipos de aplicaciones provienen del Procesamiento del Lenguaje Natural, la Recuperación de Información y el Aprendizaje Automático.

Como novedad, este taller incluye este año una competición centrada en la detección del plagio, con dos tareas:

- Análisis de plagio externo, consistente en la detección de los pasajes de una serie de documentos que pueden haber sido plagiados a partir de otra serie de documentos de origen.
- Análisis de plagio interno, que consiste en la detección de los pasajes de una serie de documentos que no han sido escritos por sus autores respectivos.

Para ambas competiciones se proporcionan una serie de documentos de prueba, previamente etiquetados, sobre los que entrenar o afinar los parámetros de los sistemas que compitan, o sobre los que evaluar y comparar la efectividad de los mismos. Las colecciones de entrenamiento se harán públicas el 21 de marzo, y las de evaluación el 21 de mayo. Los resultados deben entregarse dos semanas más tarde, y la

comparativa se hará pública una semana más tarde. Más información en el sitio Web de la conferencia [2].

[1] **SEPLN.** <http://www.sepln.org>.

[2] **PANWorkshop and Competition.** <http://www.webis.de/pan-09>.

Tema: noticia

La empresa Idée Inc. [1] ha desarrollado un conjunto de herramientas para la búsqueda y recuperación de imágenes. Piximilar [2] tiene como objetivo la búsqueda de imágenes en grandes colecciones sin la utilización de palabras clave o metadatos. Sus algoritmos analizan atributos de la imagen como el color, la forma, la textura, la luminosidad, la complejidad, los objetos y las regiones. Estos atributos conforman un perfil visual único que es almacenado en el índice visual del conjunto de imágenes. El software de Idée Inc. es capaz de calcular estos perfiles y de utilizar el índice para buscar y recuperar las imágenes. En colecciones extremadamente grandes, Piximilar puede ser utilizado en combinación con búsquedas basadas en palabras claves, lo que permite al usuario refinar la búsqueda.

Las búsquedas de imágenes pueden realizarse escogiendo uno o varios colores de una paleta de colores, usando Multicolour Search [3]. En este caso se recuperan imágenes de la colección que contienen dichos colores. También, mediante BYO Image Search [4], puede buscar imágenes existentes en la colección que son parecidas a una proporcionada por el usuario o recuperada usando Multicolour Search.

La otra herramienta, denominada TinEye [5], es un motor de búsqueda de imágenes en la Web. Este buscador permite al usuario proporcionar una imagen, obteniendo como respuesta las páginas Web que contiene dicha imagen.

Todas estas herramientas están disponibles y pueden ser probadas en el sitio Web de esta empresa.

[1] **Idée Inc.** <http://ideeinc.com/>.

[2] **Piximilar.** <http://ideeinc.com/products/piximilar/>.

[3] **Multicolour Search.** <http://labs.ideeinc.com/multicolour>.

[4] **BYO Image Search.** <http://labs.ideeinc.com/upload/>.

[5] **TinEye.** <http://ideeinc.com/products/tineye/>.

Sección Técnica "Auditoría SITIC" (Marina Touriño Troitiño, Manuel Palao García-Suelto)

Tema: CGEIT – "nueva" certificación profesional

El próximo día 13 de junio tendrá lugar, en más de 260 ciudades de todo el mundo (Barcelona, Madrid y Valencia entre ellas), el segundo examen semestral para optar a la certificación CGEIT¹ administrada por ISACA². Simultáneamente se administran también los exámenes para optar a las certificaciones CISA³ y CISM⁴, más conocidas y cubiertas anteriormente en estas "referencias autorizadas".

Los requisitos para la obtención y mantenimiento de las tres certificaciones siguen el esquema clásico: i) acreditación de experiencia en la función (parcialmente convalidable por estudios universitarios, certificaciones y experiencias afines); ii) superación del examen; iii) adhesión al Código de Ética Profesional de ISACA; y iv) [para la renovación anual] acreditar el cumplimiento del Programa de Formación Continuada.

La certificación CGEIT es "nueva" cronológicamente y también por su campo funcional.

Cronológicamente, porque fue creada en 2007 y durante 2008 sólo se concedieron unos centenares de certificaciones a "profesionales con alta experiencia profesional", por un procedimiento de convalidación del examen (*grandfathering*) inicial y transitorio, limitado a ese año. El primer examen tuvo lugar el pasado diciembre y las primeras certificaciones, siguiendo el esquema clásico, se están otorgando este año.

También es nueva por su campo funcional: el buen gobierno (la gobernanza) de los SITIC. CGEIT contempla 6 grandes dominios de práctica profesional:

- El marco de buen gobierno de los SITIC.
- La alineación estratégica de los planes y operaciones SITIC con los corporativos.
- La generación y entrega de valor en el ciclo de vida completo.
- La gestión de riesgos SITIC.
- La gestión de los recursos.
- La medida de su productividad y desempeño.

Todos ellos agrupan técnicas y habilidades clave, altamente demandadas en todos los países y empresas avanzados.

Encuestas y estudios muestran que de todos ellos hay un déficit global. Quizá más acusado en nuestro país, enfrentado ahora además a la necesidad de un cambio de modelo económico: del ladrillo (baja productividad, mano de obra de baja cualificación) a la sociedad de la información (alta productividad, mano de obra cualificada y continuamente reciclada).

"Los SITIC evolucionan continuamente para asegurar la competitividad, permitir acceder a los mercados globales y gestionar presiones externas, como la regulación. Para satisfacer la creciente demanda del mercado y reconocer el amplio abanico de profesionales, cuyos conocimientos y aplicación de los principios del buen gobierno SITIC resultan clave para la gestión de las fuerzas de la transición, ISACA ofrece una certificación para estos profesionales en evolución continua: CGEIT"⁵.

(Los exámenes CISA se ofrecen en unas docenas de idiomas, entre ellos el castellano; los CISM y CGEIT, por el momento sólo en inglés. Las inscripciones para los exámenes se cierran el 8 de abril).

¹ CGEIT - Certified in the Governance of Enterprise IT; Certificado en el Buen Gobierno de los SITIC Empresariales.

<http://www.isaca.org/Template.cfm?Section=CGEIT_Certification&Template=/TaggedPage/TaggedPageDisplay.cfm&TPLID=16&ContentID=36126>.

² <<http://www.isaca.org/>>.

³ CISA - Certified Information Systems Auditor; Auditor de Sistemas de Información Certificado. Desde su creación en 1978, 60.000 certificados. Acreditada por ANSI - American National Standards Institute; Instituto Norteamericano de Normalización y por ISO/IEC 17024:2003.

⁴ CISM - Certified Information Security Manager; Director / Responsable de Seguridad de la Información Certificado. Desde su creación en 2003, 10.000 certificados. Acreditada por ANSI - American National Standards Institute; Instituto Norteamericano de Normalización y por ISO/IEC 17024:2003.

⁵ Traducción libre de parte de "IT Governance Certification—Your Blueprint for Success"; 'Certificación en Buen Gobierno SITIC—Su Mapa hacia el Éxito'. (El subrayado es mío). <<http://www.isaca.org/AMTemplate.cfm?Section=CGEIT2&Template=/ContentManagement/ContentDisplay.cfm&ContentID=47300>>.

Sección Técnica "Derecho y Tecnologías" (Elena Davara Fernández de Marcos)

Tema: primera denuncia ante la AEPD por suplantación de personalidad en una red social

Tomando como base los estudios, informes y demás noticias publicadas hasta la fecha, cabe afirmar que 2008 será recordado como el año en el que las redes sociales se convirtieron en un fenómeno de masas y, lo que es más, expandido en todos los rincones de la tierra. Ante tal realidad, han sido muchos los que se han apresurado a señalar las múltiples ventajas de esta herramienta, entre las que se encuentran: la rapidez y facilidad de envío de mensajes entre los conocidos, así como la posibilidad de retomar contacto con personas de la infancia o de la etapa universitaria. No obstante, para que estas ventajas sean, de hecho, una realidad, es necesario partir de una premisa que, no en todas las ocasiones, se cumple, a saber: la veracidad de la identidad de quien envía el mensaje.

En este sentido, la Agencia Española de Protección de Datos (AEPD) acaba de recibir la primera denuncia por "suplantación de personalidad" a través de una red social. Es más, fruto de la denuncia presentada por un ciudadano al respecto, la Agencia ha comenzado una investigación, de doce meses de duración, para "verificar los niveles de riesgo" de estos espacios y, por tanto, en caso de encontrar indicios o fundamentos de incumplimiento de la normativa en protección de datos, la AEPD iniciaría un procedimiento sancionador a resolver en seis meses.

Como ya hemos mencionado, no cabe duda de la importancia y actualidad de las redes sociales y, por tanto, la AEPD no puede quedarse al margen en estas cuestiones y, precisamente en este sentido, ha mostrado en más de una ocasión una significativa preocupación por las amenazas y riesgos a la esfera más íntima de la persona y, muy especialmente, de los menores de edad.

Es así que destacamos, según informa la Agencia, que más del 40% de los usuarios disponen de un perfil accesible por todos los usuarios, con independencia de la relación que se mantenga con cada uno de ellos, incluso en caso de que sea un completo desconocido, además de las dificultades a las que han de hacer frente los usuarios cuando quieren dar de baja su perfil, todo ello debido en gran medida a la "falta de claridad" en las condiciones generales y política de uso de las redes sociales.

<<http://www.consumer.es/web/es/tecnologia/2009/02/15/183431.php>>.

Tema: la Unión Europea avanza en la reforma de la normativa comunitaria sobre telecomunicaciones

No cabe duda de que el uso y la total expansión de las Tecnologías de la Información y la Comunicación es una de las mayores preocupaciones a nivel europeo y es que, hoy en día, todos los sectores de la actividad comercial, personal y de ocio tienen, en mayor o menor medida, vinculación con las Nuevas Tecnologías y, muy especialmente, con las telecomunicaciones.

En este sentido, los Gobiernos de la Unión Europea (UE) han avanzado un escalón más en el camino de la regulación comunitaria sobre telecomunicaciones y, en concreto, en lo relacionado con la expansión de la banda ancha a todos los ciudadanos y la mejora en el uso del espectro radioeléctrico, entre otros. Y es que los Estados Miembros han aprobado una "posición común" sobre la base del acuerdo político del pasado mes de noviembre.

Ahondando en la letra del nuevo texto, las novedades vienen referidas a los siguientes aspectos: de un lado, una referencia a los riesgos a los que las compañías han de hacer frente a la hora de invertir en aras a la promoción y fomento de la innovación y el uso de nuevas infraestructuras. De otro, se hace hincapié en el modo en el que los Estados Miembros cumplirán las obligaciones contenidas en la nueva normativa en aras de garantizar un mercado transparente, proporcionado y no discriminatorio. Por último, destacar que los Estados Miembros adoptaron una posición común sobre la llamada "separación funcional". En concreto, esta propuesta supone exigir a un operador dominante que disocie la infraestructura de red de su negocio de servicios para mejorar la competencia en el mercado. No obstante, hay que tener presente en todo momento que aún son tres los textos normativos comunitarios relacionados con la modificación del marco regulador de las telecomunicaciones: el propuesto por el Consejo de la Unión, el de la Comisión Europea y el del Parlamento Europeo.

<<http://www.hoytecnologia.com/noticias/presenta-nueva-propuesta-reforma/83615>>.

Tema: la Unión Europea establece medidas para fomentar la sanidad electrónica

Cada día son más las implicaciones y aplicaciones que van surgiendo en relación al uso de las Nuevas Tecnologías en diversos sectores sociales: desde la educación, hasta el ocio pasando por la sanidad o el aspecto comercial. Y es precisamente en el campo de la sanidad donde ha tenido lugar una de las últimas innovaciones tecnológicas.

La nueva propuesta corre a cargo de la actual presidencia de la Unión Europea (UE), es decir, de la República Checa que, según acaba de informar, tiene intención de crear una plataforma de alto nivel político para coordinar la aplicación de las Tecnologías de la Información y las Comunicaciones al campo sanitario.

En concreto, la nueva herramienta bajo el nombre "E-Health" tiene como principal objetivo aplicar las TIC al sector sanitario, de manera que se produzca un significativo ahorro en tiempo y en personal al mismo tiempo que se mejore el tratamiento y seguimiento de pacientes con dificultades de movilidad. Es más, la nueva propuesta pretende que el sistema sanitario electrónico se expanda a través de las fronteras, integrando a todos los ciudadanos de la Unión, de forma que, en un futuro no muy lejano, se pueda hablar con propiedad de un servicio de atención sanitaria integral en todo el territorio de la Unión Europea.

En todo caso, se ha querido hacer hincapié en la dificultad de poner en marcha esta plataforma, debido principalmente a las diversas tecnologías de los Estados Miembros y a sus distintos grados de implantación y desarrollo. Por todo ello, y por cuanto esta iniciativa constituye, a día de hoy, una de las prioridades de la presidencia checa de la UE, se llevarán a cabo varias charlas, discusiones y debates para lograr integrar todas las preferencias, intereses y preocupaciones de todos los afectados, desde los pacientes hasta los hospitales, pasando por los farmacéuticos y personal sanitario, entre otros.
<<http://www.hoytecnologia.com/noticias/lanzara-plataforma-politica-para/98243>>.

Tema: Italia endurece la regulación comunitaria sobre juegos online

La Unión Europea ha dictado una serie de informes y decisiones relacionadas con diversos aspectos reguladores del sector de las Tecnologías de la Información y las Comunicaciones. En este sentido, una de las cuestiones que más controversia está suscitando para crear un marco comunitario regulador es la relacionada con las casas de apuestas y los casinos que ejercen su actividad telemáticamente. Y es que son muchas y muy dispares las regulaciones de los Estados Miembros que componen la Unión Europea en este sentido. En concreto, el Gobierno italiano, mediante la ley de transposición a su ordenamiento jurídico nacional de la Directiva sobre el juego en Internet, ha decidido incluir una serie de disposiciones que no hacen sino endurecer los requisitos de acceso y participación en los sitios web de apuestas *online*.

Por lo que se refiere a medidas concretas, este endurecimiento de los requisitos tiene su reflejo más fiel en la imposición de penas de cárcel a los responsables de aquellos negocios de apuestas *online* que ejerzan sin la pertinente concesión del Gobierno. Como no podía ser de otra manera, en la ley italiana se mencionan algunas obligaciones con las que habrán de cumplir los usuarios, siendo la más destacada que para la utilización de estos servicios tendrán que abrir una "cuenta de juego" en la página elegida, en la que introducirán el dinero que quieran jugar. Por último, destacar que la ley italiana en su intento por hacer una regulación exhaustiva y en la que queden incluidos todos los supuestos de apuestas *online*, no ha dudado en afirmar que, en caso de que la cuenta del usuario no sea utilizada durante un período de tres años, todo el contenido allí dispuesto pasará a manos del Erario público
<<http://www.elmundo.es/elmundo/2009/02/06/navegante/1233905920.html>>.

Sección Técnica "Entorno Digital Personal"
(Alonso Álvarez García, Diego Gachet Pérez)

Tema: libro

Jerome (J. F.) DiMarzio. *Android: A Programmers Guide*. McGraw-Hill, 2008. ISBN-13: 9780071599887

En julio del año 2008 la editorial McGraw-Hill sacó al mercado este texto de referencia para los desarrolladores de aplicaciones móviles sobre dispositivos que usan Android como plataforma de desarrollo
<<http://www.mhprofessional.com/product.php?isbn=0071599886&cat=112>>. Su autor es J.F. DiMarzio, un desarrollador con muchos años de experiencia en el tema.

Este libro ha sido pensado y diseñado como una primera aproximación hacia el desarrollo de aplicaciones para dispositivos móviles de código abierto. Android es la más nueva de las plataformas de desarrollo para dispositivos móviles, está basado en Java y se apoya en un *kernel* de Linux 2.6. El libro en cuestión hace una descripción profunda acerca de la arquitectura de Android para luego tratar el desarrollo de aplicaciones, concluyendo con una descripción acerca de la vinculación con herramientas de Google como Google Maps y Gtalk. Desde el punto de vista de la programación podemos decir que no es un libro para principiantes sino que es necesario un conocimiento relativamente profundo acerca del lenguaje de programación Java.

Recordemos que Android fue liberado por Google en noviembre de 2007, bajo los auspicios de la *Open Handset Alliance*, un grupo conformado por fabricantes de hardware y software entre los que podemos mencionar a Google, NTT DoCoMo, Sprint Nextel, y HTC, cuya meta es la creación de un entorno de desarrollo más abierto para las aplicaciones móviles <<http://www.openhandsetalliance.com>>.

Con el lanzamiento de Android, Google puso también a disposición de los desarrolladores una gran cantidad de herramientas, así como tutoriales para ayudarles con el nuevo sistema. Entre estos elementos se pueden mencionar: archivos de ayuda, el *kit* de desarrollo de software (SDK) e incluso una comunidad de desarrolladores, recursos que se puede encontrar en el sitio web de Google Android, <<http://code.google.com/android>>. Este sitio debería considerarse como un punto de partida que desde luego recomendamos visitar.

Sección Técnica "Gestión del Conocimiento"
(Joan Baiget Solé)

Tema: libro

Mario Pérez-Montoro Gutierrez. *Gestión del Conocimiento en las Organizaciones*. Ed. TREA. ISBN: 978-84-9704-376-2.

El libro de Mario Pérez-Montoro es una de las pocas obras acerca de la Gestión del Conocimiento (GC) que puede tener la consideración de 'libro de texto' para ser usado en un contexto introductorio al tema. Nos conduce hacia la Sociedad del Conocimiento a través del análisis del cambio tecnológico, la globalización, etc.

Dedica un primer capítulo a la fundamentación conceptual de la Gestión del Conocimiento y aborda una extensa lista de aspectos clave: Datos, Información, Conocimiento y sus tipos, tipologías existentes en las organizaciones, el conocimiento como proceso y como objeto, la espiral de conocimiento de Nonaka-Takeuchi, el aprendizaje, etc., con profusión de ejemplos y explicaciones. Un segundo capítulo nos introduce al "Capital Intelectual", su medición y la relación con la GC, para abordar en el capítulo posterior el mercado de la GC en las organizaciones, su flujo y su ciclo de vida, con un detallado recorrido por cada una de las fases de dicho ciclo. El penúltimo capítulo, sobre metodología de implantación de un programa de GC explica en las diferentes etapas de análisis, diseño e implantación, la auditoría y el mapa de conocimiento, la gestión de contenidos, las comunidades, etc., así como algunas soluciones tecnológicas actuales. Finalmente, en el último capítulo, se exponen algunos proyectos emblemáticos -casos prácticos- de Gestión del Conocimiento.

Una obra seria, rigurosa y sistemática, altamente recomendable para todos aquellos interesados en el tema.

Sección Técnica "Informática Gráfica"
(Roberto Vivó Hernando, Miguel Chover Sellés)

Tema: *Processing*

Processing es un proyecto de software abierto comenzado en 2001 por Ben Fry y Casey Reas durante su estancia en el Media Lab del MIT. Processing es un entorno de programación con su propio lenguaje orientado a la programación gráfica y estética, muy relacionado con las artes. Asimismo, fue utilizado originalmente, y aun hoy, para fines

docentes ya que permite un rápido prototipado de aplicaciones con resultado gráfico gracias a su lenguaje interpretado.

Processing es gratuito y constituye una alternativa a otros programas licenciados sobre todo para escuelas y estudiantes. Su facilidad de uso y su extensibilidad ha permitido una rápida evolución de la plataforma con contribuciones de la comunidad en librerías para visión artificial, visualización de datos, música, etc. En su portal web <<http://www.processing.org>> pueden visitarse multitud de proyectos con resultados visuales impresionantes.

El código de Processing se traduce a código Java con un preprocesador, por lo que es posible utilizar todas las plataformas y herramientas donde corre Java, por ejemplo, Eclipse. Existen bastantes restricciones de uso de la AWT de Java ya que interfiere con el modelo gráfico controlado por Processing.

Existen diversas publicaciones sobre Processing que iremos comentando en esta sección en futuros números. Para empezar a conocer los fundamentos del uso y programación es suficiente descargarlo del portal y seguir los ejemplos y tutoriales que allí se presentan diferenciados por niveles introductorio, intermedio y avanzado. En resumen, Processing es otra iniciativa más de éxito de la comunidad "open source" que se amplía cada día gracias a la desinteresada contribución de esos admirables desarrolladores anónimos.

Sección Técnica "Ingeniería del Software" (Javier Dolado Cosín, Luis Fernández Sanz)

Tema: libro

Luis Fernández Sanz, Pedro J. Lara Bercial. *Test Case Generation, UML, and Eclipse. Encouraging good practices to ensure software quality.* Dr. Dobbs, <<http://www.ddj.com/architect/211600996>>, Oct. 2008.

Eclipse, auspiciado por IBM, supone un entorno libre y flexible para el desarrollo en el que la aplicación de UML está perfectamente integrada con el código. La elaboración de *plugins* resulta una tarea abordable por cualquier profesional del software. En este trabajo se presenta la aplicación de las técnicas de generación automática de pruebas a partir de diagramas de actividad UML soportada por un *plugin* que facilita su aplicación práctica. En general, la revista Dr. Dobbs incluye mucha información práctica así como código y ejemplos de soporte.

Tema: repositorio de patrones

Portland Software Patterns Repository <<http://c2.com/cgi/wiki?PatternIndex>>.

Este es un repositorio de patrones para el desarrollo y diseño de software creado por Ward Cunningham para que sirva de referencia a los desarrolladores, con vistas a tener un lugar de consulta permanente si bien tiene una especial orientación a la XP (*eXtremme Programming*). Está acompañado por un wiki que fue el primero que se creó en la historia.

Sección Técnica "Interacción Persona-Computador" (Pedro M. Latorre Andrés, Francisco L. Gutiérrez Vela)

Tema: noticia

Los ámbitos académicos relacionados con la Informática no podían quedar fuera de la marea de opinión conducente al reconocimiento profesional de nuestra especialidad. Por este motivo se ha lanzado en este momento la constitución de la Sociedad Científica Informática de España (SCIE).

La SCIE, es una federación de asociaciones y sociedades científicas que nace con la vocación de participar en los procesos relacionados con la evolución de la Informática y la formación de los futuros profesionales. Entre otros cometidos se encargará de dar continuidad al CEDI (Congreso Español de Informática) organizando futuras ediciones del

mismo, y conceder los Premios Nacionales de Informática.

Inicialmente la SCIE está integrada por:

- Sociedad de Arquitectura y Tecnología de Computadores (SARTECO).
- Asociación Española para la Inteligencia Artificial (AEPIA). EUROGRAPHICS Sección Española.
- Sociedad de Ingeniería del Software y Tecnologías de Desarrollo de Software (SISTEDES).
- Asociación Interacción Persona Ordenador (AIPO).
- Sociedad Española para el procesamiento del lenguaje natural (SEPLN).

Tema: libro

Miguel Redondo, Crescencio Bravo, Manuel Ortega (Eds.). *Engineering the User Interface. From Research to Practice.* Springer 2009, XII, 277 páginas. ISBN: 978-1-84800-135-0.

Este libro contiene una selección de 16 artículos de alta calidad presentados al 7º Congreso Internacional sobre Interacción Persona - Ordenador Interacción'2006. El Congreso se celebró en Puertollano (Ciudad Real) del 13 al 17 de noviembre de 2006, organizado por el grupo de investigación CHICO de la Universidad de Castilla-La Mancha y la Asociación Interacción Persona-Ordenador (AIPO).

Los objetivos de INTERACCIÓN'06 se centraron en tres líneas: diseminar los resultados de la investigación y promover la discusión en Interacción Persona-Ordenador (IPO); poner en contacto grupos de investigación de todo el mundo, principalmente de lengua española, con los máximos especialistas internacionales; y finalmente enriquecer las relaciones entre el mundo empresarial y académico en los temas relacionados con la IPO.

El texto se articula de acuerdo a las principales líneas de investigación pura y aplicada que se presentaron al Congreso: Ergonomía y Factores Humanos, Utilizabilidad y Accesibilidad, Inteligencia Ambiental y Sistemas Sensibles al Contexto, Diseño centrado en el Usuario, Sistemas y Modelos para Trabajo Colaborativo, IPO en e-Learning, Diseño y Desarrollo de Interfaces de Usuario, Realidad Virtual y Aumentada, Interacción Multimodal, Dispositivos de Computación Ubicua e IPO para Personas con Necesidades Especiales.

Está dirigido a investigadores, desarrolladores y usuarios de sistemas interactivos y proporciona una idea bastante aproximada de las líneas de trabajo de los grupos investigadores fundamentalmente españoles e iberoamericanos. Su lenguaje resulta accesible a los no expertos, y puede ser un punto de partida idóneo para establecer contactos para futuras líneas de colaboración con empresas e instituciones.

El congreso INTERACCIÓN se celebra con periodicidad anual. La próxima edición se celebrará en Barcelona en septiembre de 2009.

Sección Técnica "Lenguajes Informáticos" (Óscar Belmonte Fernández, Andrés Marín López)

Tema: Cloud computing

Cloud computing es el nuevo término de atracción en el mundo de las Tecnologías de la Información. Cualquiera que sea la aplicación o servicio, si se le añade el término *cloud computing* se le añade una impronta que lo hace más atractivo. El término es tan nuevo que no existe, de momento, una definición ampliamente aceptada, aunque sí que existe consenso sobre las características que deberá contener la definición: Las aplicaciones *cloud computing* están orientadas al servicio, el proveedor mantiene una única aplicación a la que concurrentemente acceden los usuarios, pueden estar integradas en Internet o ejecutarse en el escritorio, los datos de los usuarios se almacenan en la nube.

Las ventajas para los clientes de estas aplicaciones son claras: se paga por el uso de servicio en términos de memoria consumida y

de transacciones realizadas, las actualizaciones de las aplicaciones son transparentes al usuario, la inversión requerida es mínima y la disponibilidad de las aplicaciones es independiente del lugar desde el que se realice el acceso. Algunos ejemplos de aplicaciones en *cloud computing* son: Amazon Elastic Compute Cloud, Google App Engine, Oracle Cloud Computing Center y Microsoft Cloud Computing Tools.

Sección Técnica "Lingüística computacional" (Xavier Gómez Guinovart, Manuel Palomar)

Tema: sistemas de respuesta a preguntas

Tomek Strzalkowski, Sanda Harabagiu (eds.). *Advances in Open Domain Question Answering*. Springer, Dordrecht, 2008, 566 páginas. ISBN 9978-1-4020-4744-2.

Los sistemas de respuesta a preguntas son aplicaciones informáticas de recuperación de la información capaces de contestar de manera concreta (y no con una lista de documentos más o menos relevantes) a preguntas expresadas en lenguaje humano. En los últimos años, se han producido importantes contribuciones al desarrollo de estos sistemas, al mismo tiempo que se ha incrementado el interés que despiertan, especialmente debido a su aplicación a las búsquedas textuales en la web.

Este libro editado por Strzalkowski y Harabagiu constituye un buen manual académico para iniciarse en esta materia. El volumen está formado por 18 capítulos en cuya redacción han participado 68 especialistas del ámbito internacional. Estas 18 contribuciones, de orientación eminentemente práctica, se agrupan en los seis capítulos que estructuran el libro: enfoques actuales de la investigación, procesamiento de las preguntas, recuperación de la información, extracción de las respuestas, evaluación de sistemas y perspectivas de la investigación. Se trata de un libro altamente recomendable para estudiantes avanzados, investigadores y desarrolladores en el campo de la recuperación de la información y del procesamiento del lenguaje natural. Para más información o adquisiciones, puede consultarse la web del editor en <<http://www.springer.com>>.

Sección técnica "Redes y Servicios Telemáticos" (José Luis Marzo Lázaro, Germán Santos Boada)

Tema: nuevas Arquitecturas para Internet

Internet ha sufrido un crecimiento exponencial desde su creación hace tan sólo 4 décadas. Dicho crecimiento está provocando diversos problemas en su arquitectura Internet y en especial en su escalabilidad. Uno de los que más preocupan es el futuro del sistema de enrutamiento global interdominio. En la actualidad el encaminamiento en Internet está en manos del protocolo BGP (*Border Gateway Protocol*). BGP se encarga de intercambiar información de encaminamiento entre dominios. Debido al crecimiento exponencial y al número siempre creciente de dominios y enlaces entre dominios las tablas de rutas de BGP en la DFZ (*Default Free Zone*) tienen alrededor de 300.000 entradas. Esto hace que administrarlas se haya convertido en una tarea muy compleja, además de estar provocando una reducción del rendimiento. Este hecho dificulta la posibilidad de ofrecer *Traffic Engineering* (TE) o *Multihoming*.

Recientemente tanto la IETF (*Internet Engineering Task Force*) y en especial el grupo de trabajo *Routing Research Group* de la IRTF (*Internet Research Task Force*) han empezado a explorar la posibilidad de diseñar una nueva arquitectura para Internet con el objetivo de solucionar los problemas mencionados anteriormente. En particular la aproximación que mayor consenso está alcanzando es separar los identificadores de los nodos de su posición topológica. En la actualidad las direcciones IP representan tanto la identidad del nodo (quién es) así como su posición topológica en Internet (dónde está conectado).

Esta separación ya había sido estudiada y analizada anteriormente y se considera como una de las principales características de una (posible) nueva arquitectura para Internet. Los principales beneficios que se esperan de dicha separación son una considerable reducción del

número de rutas presentes en la DFZ y la posibilidad de ofrecer TE. En especial, una funcionalidad que se espera obtener de dicha separación es la posibilidad de asignar direcciones IP que sean independientes del proveedor (*Provider Independent addresses*, PI). Estas direcciones se asignan a las redes clientes independientemente del proveedor de Internet contratado y permiten cambiarlo sin tener que reasignar direcciones a la red cliente. Esto hace que cambiar de proveedor no tenga un alto coste asociado para el cliente y se espera que aumente considerablemente la competencia entre los proveedores.

Del *Routing Research Group* de la IRTF han surgido multitud de nuevas arquitecturas basadas en la separación de la identidad y la localización. De entre estas propuestas la que está alcanzando mayor consenso es LISP (*Locator Identifier Separation Protocol*). LISP divide el sistema de direccionamiento y enrutamiento en dos funciones: *Routing Locators* (RLOCs) que describe dónde está conectado el nodo y *Endpoint Identifiers* (EIDs) que representan la identidad del nodo, en un espacio de direccionamiento único y separado. En LISP los RLOCs se asignan a los *border routers* (*tunnel router* en LISP) y por lo tanto, el núcleo (*core*) de Internet sólo es capaz de encaminar paquetes según su RLOC. Por otro lado los EIDs se asignan a los nodos y sólo tienen significado dentro de los dominios. LISP ha sido diseñado con dos grandes objetivos en mente, en primer lugar aumentar la escalabilidad del sistema global de encaminamiento inter-dominio y en segundo lugar ser un protocolo con un coste de despliegue muy bajo. En particular LISP no requiere modificar nada en las redes cliente.

El funcionamiento (resumido) de LISP es el siguiente: cuando un nodo origen quiere comunicarse con un nodo destino busca en el DNS su EID (dirección IP) y envía el paquete tal y como se hace en la arquitectura actual de Internet. Dicho paquete se encamina por el dominio hasta que llega al *tunnel router* del dominio. En ese momento el router debe conocer cuál es el RLOC del dominio destino del EID. Por ello LISP define una base de datos distribuida (*mapping system*) que relaciona EIDs con RLOCs. Así pues el router consulta dicha base de datos y obtiene el RLOC destino, encapsula el paquete y lo dirige al *tunnel router* del nodo destino. Una vez el paquete llega al destino se desencapsula y se encamina por el dominio hasta el nodo destino siguiendo el procedimiento estándar.

Las nuevas arquitecturas basadas en la separación de la identidad y la localización aún tienen aspectos por definir. Por ejemplo en LISP aún se debe analizar cómo se diseñará la base de datos distribuida que relaciona RLOCs con EIDs. Actualmente existen propuestas para implementarla usando el propio BGP o redes P2P. Al margen de estos aspectos es importante destacar que existe un cierto consenso en torno a la necesidad de una nueva arquitectura, tanto entre el mundo académico como en la industria. De hecho LISP está siendo apoyado por CISCO y ya dispone de implementaciones y despliegues experimentales. Más información en:

<http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_11-1/ipj_11-1.pdf>

Información proporcionada por **Albert Cabellos-Aparicio**, Dept. Arquitectura de Computadores de la Universitat Politècnica de Catalunya <<http://people.ac.upc.edu/acabello>>.

Sección técnica "Seguridad" (Javier Areitio Bertolín, Javier López Muñoz)

Tema: libros

L. Silverman. *Physical Security and Wireless Access Control Systems*. Auerbach Publications. ISBN 1420078151. 2009.

S.K. Chin, S.B. Older. *A Mathematical Introduction to Access Control: A Logical Approach*. Chapman and Hall/CRC. ISBN 1544888628. 2009.

M. Blyth. *Business Continuity Management: Building an Effective Incident Management Plan*. Wiley. ISBN 0470430346. 2009.

T.R. Walsh. *A Manager's Guide to Handling Information Security Incidents*. Auerbach Publications. ISBN 0849390532. 2009.

J.R. Vacca. *Computer and Information Security Handbook*. Morgan

Kaufmann. ISBN 0123743540. 2009.

S.F. Mjolsnes. *A Multidisciplinary Introduction to ICT Information Security.* Chapman and Hall/CRC. ISBN 1420085905. 2009.

W.J. Buchanan. *The Handbook of Data and Networks Security.* Springer. ISBN 0387286020. 2009.

R. Panko. *Corporate Computer and Network Security.* Prentice-Hall. ISBN 0131854755. 2009.

Tema: *congresos-conferencias-simposium*

ISC'2009 (Information Security Conference). Del 7 al 9 de septiembre 2009. Pisa, Italia.

ICISIE'2009 (The 2009 International Conference of Information Security and Internet Engineering). Del 1 al 3 de julio 2009. Londres, Reino Unido.

Securmática'2009. XX Congreso Español de Seguridad de la Información. Del 21 al 23 de abril 2009. Campo de las Naciones. Madrid.

ICISC'2009 (The 12th International Conference on Information Security and Cryptology). Del 2 al 4 de diciembre 2009. Seul, Corea.

International Conference on Security of Information and Networks. SINCONF. Del 6 al 10 de octubre 2009. Eastern Mediterranean University. Gazimagusa. TRNC. Chipre Norte.

Sección Técnica: "Tecnología de Objetos"
(Jesús García Molina, Gustavo Rossi)

Tema: *desarrollo de software dirigido por modelos con la plataforma Oslo de Microsoft*

A finales del pasado año Microsoft presentó su nueva plataforma "Oslo" para el Desarrollo de Software Dirigido por Modelos (DSDM). Y desde noviembre de 2008 está disponible el libro *"The Oslo Modeling Language"*, de **David Langworthy, Brad Lovering y Don Box** (Addison-Wesley, 2008) que presenta dicha plataforma. Se trata de un manual de referencia con muy poca información, realmente es una *draft specification*, que tendría más sentido como un documento público sin coste alguno, disponible en la web de Microsoft.

Esta plataforma se articula en torno al lenguaje de modelado "Oslo" (también conocido como "M"), de la misma forma que la plataforma de DSDM de Eclipse utiliza el lenguaje Ecore y OMG propone el lenguaje MOF. Por tanto, Oslo es la alternativa de Microsoft para aplicar las técnicas del metamodelado.

M es un potente lenguaje textual que combina las capacidades de un UML-textual para expresar la parte estructural de un modelo (o metamodelo) y de OCL para especificar propiedades (invariantes, precondiciones, etc.) y *queries* sobre modelos. Cuando se define el metamodelo de una sintaxis abstracta con Ecore o MOF es necesario usar OCL para expresar la semántica estática (restricciones), mientras que con M sólo se necesita un único lenguaje. En una lectura del libro mencionado se puede observar que M es un lenguaje muy elegante, legible, y con un sistema de tipos muy potente. Incluye el concepto de módulo como alternativa a los paquetes de UML.

MGrammar (Mg) es un lenguaje asociado a M que permite crear lenguajes específicos de dominio (DSL) textuales, tanto la estructura sintáctica del lenguaje como las estructuras de datos que se generarán cuando se procese (lo que se denomina proyección). Los datos que se producen conforman a un modelo M y pueden ser procesados mediante *queries* M.

Una definición Mg de un DSL está formada por un conjunto de reglas "syntax" y reglas "token" (identifican *tokens* del lenguaje). Cada regla establece la sintaxis y la salida (texto generado). Una regla "syntax" consta de dos partes: patrón y proyección que es opcional. Un patrón describe un conjunto de secuencias de caracteres que constituyen una entrada válida, y una proyección describe qué información es generada para esa entrada. Aunque M y Mg son lenguajes textuales, Oslo también incluye una herramienta visual para manipular modelos.

El libro publicado sobre Oslo está organizado en dos partes, la

primera es un manual de M y la segunda un manual de Mg. Su lectura requiere algún conocimiento de metamodelado y creación de DSL textuales para valorar la justa aportación de estos lenguajes.

Hasta ahora, se había asociado Microsoft al DSDM a través del enfoque de las factorías de software propuesto por **Jack Greenfield** en el excelente libro *"Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools"* (John Wiley, 2004), en el que se describe cómo aplicar las técnicas del DSDM para crear líneas de producto. Microsoft también ha desarrollado la herramienta "DSL Tools" para la creación de DSL visuales aplicando las técnicas del metamodelo pero con un lenguaje de metamodelado propio. Ahora habrá que ver cómo se alinean estos enfoques y herramientas con la plataforma Oslo y el lenguaje M.

En definitiva, Microsoft apuesta por el DSDM, lo cual confirma que se trata de una visión con gran potencial para mejorar la productividad del desarrollo de software, y lo hace a través de un lenguaje de metamodelado con muy buenas propiedades.

Sección Técnica: "TIC y Turismo"
(Andrés Aguayo Maldonado, Antonio Guevara Plaza)

Tema: *Congreso Internacional de Tecnologías Aplicadas al Turismo, ENTER 2009*

El pasado mes de enero de 2009, organizado por la *International Federation for Information Technology and Travel & Tourism* (IFITT), se celebró en Ámsterdam la 16ª edición del congreso internacional ENTER. Dicha conferencia reunió a expertos internacionales en todos los ámbitos de las tecnologías de la información y comunicación aplicadas a viajes y turismo. El congreso ENTER ofrece un foro único para el sector académico, la industria, los gestores de destino y de marketing y representantes de organizaciones gubernamentales para explorar el futuro de las TIC en el turismo a través de la investigación y el diálogo dinámico dentro de la red social de la comunidad IFITT. En el congreso se desarrollaron más de 150 presentaciones y asistieron más de 450 participantes. Podemos calificar a ENTER como un gran evento internacional en Turismo y Tecnología que reúne a investigadores, profesionales y gestores de los destinos turísticos.

El tema central del congreso de 2009 ha sido eTourism: desafíos para la dinámica de viajes y turismo. Se ha analizado este tema tanto desde el punto de vista de la tecnología aplicada a las empresas como desde el punto de vista de la tecnología aplicada a las organizaciones de destinos locales y nacionales encargadas de la gestión del turismo.

Entre muchos temas tratados destacamos los siguientes:

- Tecnologías aplicadas a la recomendación de sistemas de viaje.
- Creación de redes sociales y de viaje.
- Web 2.0 y Travel 2.0.
- Segmentación de las TIC.
- eTourism y discapacidad.
- Diseño de eTourism accesible.
- e-Learning para el Turismo.
- Distribución directa versus indirecta.
- La productividad basada en las TIC.
- Reingeniería del turismo.
- eTourism y Precios.
- Análisis web.
- Seguridad en línea, ePayments.
- eBranding y la integridad de la marca.
- Web 3.0 y Turismo 3.0.
- Servicios basados en localización.
- Sistemas inalámbricos y de WiMAX.
- Soluciones móviles.
- Usabilidad y diseño web.
- Convergencia multimedia.

Las ponencias del evento se publican en la editorial Springer <<http://www.ifitt.org/enter>>.

Entrevista a Gabriel Martí Fuentes

Empresario del sector informático; Responsable de tecnología de ATI

<gabi@atinet.es>

1. Visión

Últimamente hemos oído hablar mucho de que ya tenemos aquí la "Web 2.0". Unos insisten en que consiste básicamente en disponer de nuevas herramientas, otros en un cambio de actitud...

P: ¿Cuál es tu visión sobre la Web 2.0?

R: El paso a la Web 2.0 es una evolución de las necesidades de información de los propios usuarios. Se pasó del contenido estático a los sitios web mantenidos con un gestor de contenidos y finalmente a la participación y alimentación de contenidos por los visitantes. Sin este cambio de actitud no existiría la Web 2.0. Sí, es cierto que ha habido cambios tecnológicos necesarios para llegar a la Web 2.0, pero es más importante, desde mi punto de vista, el cambio generado por el comportamiento humano que el cambio tecnológico en sí.

P: ¿Crees que la Web 2.0 es una "revolución tecnológica" o una simple evolución a partir de lo que ya teníamos?

R: No diría que es una revolución, es una evolución, tanto tecnológica como social. Hace 20 años nadie imaginaba tener acceso global a la información, y hace 10 años con la tímida llegada de Internet, nadie se imaginaría estar explicando su vida en Internet a través de un blog o las redes sociales. Esto se produce por unos cambios de hábitos influenciados por la llegada de las tecnologías y las necesidades generadas.

P: Para participar en ella, ¿debemos adquirir hardware, software o ambas cosas?

R: Para participar en la Web 2.0 no se requiere de ningún hardware especial que no sea un ordenador con capacidad suficiente para tener una conexión a Internet y correr un navegador. Respecto al software es recomendable un navegador que no sea obsoleto, pero en principio cualquier navegador web sirve.

P: ¿Cuáles son las herramientas principales que podemos o debemos usar para participar en la Web 2.0?

R: Desde el momento que nos registramos en cualquier sitio web que permite la participación del usuario y expresar y compartir opiniones estamos participando en la Web 2.0, y no estamos utilizando ninguna herramienta más imprescindible que un navegador web.

Resumen: hace ahora unos cinco años un empresario y editor llamado Tim O'Reilly utilizó por primera vez un término que ha hecho fortuna para referirse a todo el fenómeno "participativo" que ha surgido del uso de la Web y de sus herramientas de "última" generación: Web 2.0. Desde entonces, y cada vez con mayor intensidad, todos queremos apuntarnos a este tema de moda, a veces sin tener las ideas claras del todo. En esta entrevista, Gabriel Martí nos ayuda a clarificar conceptos y nos relata sus experiencias alrededor de la Web 2.0. Al mismo tiempo, como ideólogo e impulsor de estos temas dentro de la Junta General de ATI nos relata el presente y el futuro de los proyectos de ATI en esta línea.

Palabras clave: ATI, ATInet 2.0, blog, Internet, participación en red, red social, Web 2.0.

Entrevistado

Gabriel Martí Fuentes es diplomado en Informática por la Universidad de Cambridge (Higher National Diploma of British Council). Es actualmente empresario del sector informático y trabaja como asesor y consultor, especializado en redes, Internet y desarrollo web. Es socio de ATI y miembro de la Junta Directiva General, ejerciendo como Responsable de tecnología de la Asociación.

Autor de la entrevista: Llorenç Pagés Casas, director de *Novática*.

Pero si queremos ser realmente activistas de la Web 2.0, un primer paso sería crear un blog para dar nuestras opiniones sobre cualquier tema que creamos importante para nosotros, y sobre todo, permitir la participación de nuestros lectores aceptando comentarios; si no tendría poco, es decir nada, de Web 2.0.

Otra herramienta de la Web 2.0, que no estaba en los inicios pero que en estos momentos tiene mucho peso es el caso de las redes sociales que permiten compartir virtualmente nuestros pensamientos, opiniones, comentarios, fotografías e incluso videos, pero esta vez con un grupo concreto de usuarios que son nuestras "amistades". Pero Web 2.0 no solo es esto; subir un video de Youtube, fotos a Flickr, una presentación a Slideshare, crear y compartir un mapa en Google Maps, o cualquier otra acción que signifique publicar contenido y compartir es participar en la Web 2.0.

P: ¿En qué consiste ese cambio de actitud del que tanto se habla?

R: El cambio de actitud viene dado por una voluntad (y casi necesidad) por parte de los usuarios de expresar sus opiniones, de hacerlas valer y mostrar una parte personal de sí mismos al mundo.

Si no hubiera la voluntad de subir contenido y compartirlo las herramientas disponibles para la Web 2.0 serían totalmente inútiles, por lo tanto van acompañadas de este cambio de actitud.

P: Si un profano te pregunta ¿dónde y**cómo puedo introducirme en la Web 2.0?... ¿Qué le dirías?**

R: Para conocer más de la Web 2.0 lo más práctico es participar en ella, pues participando se va conociendo más sobre la mecánica de su funcionamiento. Pero como siempre existe ese miedo a lo nuevo, recomendaría los siguientes enlaces para conocer un poco más:

Unas presentaciones en Slideshare, una Web 2.0, donde los usuarios ceden sus presentaciones para que otros las vean, compartan y comenten:

<<http://www.slideshare.net/genisroca/introduccion-a-la-web-20>>.

<<http://www.slideshare.net/jquemada/introduccion-al-web-20>>.

<<http://www.slideshare.net/siruze/introduccion-a-la-web-20-presentation>>.

Un artículo interesante con un par de videos explicativos:

<<http://blog.federicosanchez.info/2007/11/08/introduccion-a-la-web-20-para-empresarios-que-apanas-encienden-la-pc/>>.

Su definición en la Wikipedia (otro ejemplo de Web 2.0):

<http://es.wikipedia.org/wiki/Web_2.0>.

Un PDF con una orientación marcadamente técnica:

<<http://www.di.uniovi.es/~labra/cursos/Web20/web20.pdf>>.

Como se puede ver, es la propia Web 2.0 la que

nos provee de la información de su propia definición.

P: ¿Crees que lo que hay alrededor de Web 2.0 es una simple moda?

R: No. También se pensó que eran una moda los primeros teléfonos móviles y ahora son una necesidad. Quizá tenga su tiempo de vida, hasta que evolucione, quien sabe a otra tecnología, pero se está convirtiendo en una cosa normal para el usuario doméstico y algo que deberían de tener en cuenta las empresas. Es un medio para expresarse.

P: ¿De verdad tenemos ya la Web 2.0 cerca de su máxima expresión o queda aún mucho por desarrollar?

R: El término Web 2.0, es un término "muy informático", por lo de las versiones de las aplicaciones, pero en realidad no se pasó del 1.0 al 2.0, antes hubo una 1.1, una 1.5, o incluso una 1.9, ya que las propias herramientas de la Web 2.0 fueron evolucionando. Y ahora puede que estemos en la Web 2.2 o 2.3, quien sabe, es complejo de medir. Hasta que un buen día aparezca la Web 3.0 de la que ya se habla desde hace algunos años.

P: ¿Hasta qué punto el fenómeno blog es característico de la Web 2.0?

R: Es una de las primeras herramientas que popularizó la Web 2.0, y permitió que cualquier persona sin necesidad de tener conocimientos técnicos, excepto las nociones básicas de utilizar un editor de textos, pudiera publicar contenido en la web de manera fácil y rápida, y con la misma facilidad que escribía en su procesador de textos habitual. Podríamos llamarlo "la primera piedra".

2. Experiencias

P: Mucha gente abre un blog para contar sus experiencias vitales, pero así corremos el riesgo de que todos hablemos (escribamos) a la vez y nadie esté escuchando (leyendo)... ¿hasta qué punto puede ser eso cierto?

R: Un blog requiere mucha dedicación, y con el tiempo se abren y cierran muchos blogs, pero se escribe mucho o poco, cualquier *blogger* contribuye al enriquecimiento de la información que fluye por la Web 2.0, y en este caso a la "blogosfera". Al mismo tiempo, la mayoría de bloggers, no se limitan solo a escribir, sino que lee también muchos otros blogs, se documentan, comparan, incluso hace referencia a otros blogs en el suyo, ya sea para criticar, alabar, corregir o comentar. Es como un intercambio de opiniones de manera abierta y pública.

P: ¿Qué opinas del fenómeno Wikipedia?

R: Es una gran herramienta. La consulto si no

a diario, sí varias veces a la semana. Pero hay que contrastar los datos ya que es un espacio abierto y cualquier persona puede escribir allí una información que pueden ser correcta o pueden ser errónea. Puede que con el tiempo evolucione y los autores requieran de cierta certificación y/o validación, con el riesgo de perder participación y popularidad.

P: ¿Cómo es posible que digan que la Wikipedia (que podemos editar todos) es más fiable que la Enciclopedia Británica (solamente editada por expertos).

R: Es increíble, pero cierto. Puede parecer una incongruencia, pero a pesar de ser un medio donde participa cualquier persona, experta o no, el hecho de tener miles (o millones) de personas que además de leer el contenido, lo pueden modificar y corregir, acelera la corrección de informaciones erróneas. Es por aquella frase de que "cuatro ojos ven más que dos", pero además en este caso "cuatro manos, corrigen más rápido que dos". Además la versatilidad del sistema *wiki*, permite volver a versiones anteriores y hacer seguimientos en los cambios, hecho que permite cambios rápidos en los contenidos.

P: Las comunidades en red y las redes sociales son otro aspecto de la Web 2.0. ¿Hasta qué punto pueden enriquecer la interacción con respecto a las comunidades presenciales?

R: Hace unos años se decía que Internet iba a crear generaciones de personas más cerradas, menos sociables, que se encerrarían en sus casas y no saldrían para nada porque desde su ordenador lo podrían hacer todo sin necesidad de cruzar la puerta.

Pero en este caso el efecto es todo lo contrario ya que las redes sociales y las comunidades virtuales en general, han permitido poner en contacto a las personas a grandes distancias. Personas comunicándose que quizá en unos casos se conocieran ya antes, y en otros que se han conocido mediante la red, y esto se ha traducido en un aumento de las relaciones sociales e incluso en una extensión del círculo de relaciones (que no amistades, que es muy diferente). Esto ha permitido que se generaran posteriormente necesidades de encuentros de "Networking presencial", pero ya con unos objetivos más específicos y con un radio de acción mayor.

P: ¿Qué aplicación, servicio o web podrías citarnos como mejor ejemplo de adopción de la filosofía 2.0?

R: Es difícil. Hay muchas y cada una en un sector concreto. Pero seguro que en este momento cualquier persona que esté leyendo estas líneas estará pensando en Facebook. No es la primera, ni será la última, pero es la del

momento actual.

P: ¿Participan realmente nuestras empresas de sectores tradicionales del fenómeno 2.0?

R: Puede que algunas sí, pero en general no. Se está extendiendo el fenómeno, pero más a nivel particular que a nivel empresarial. Las empresas tradicionales necesitan ver un rendimiento a corto plazo y lo traducen siempre en beneficio económico. Supongo que también es cuestión generacional, y las empresas creadas por jóvenes emprendedores tendrán una visión muy diferente y arriesgarán invirtiendo más en tecnología Web (no solo Web 2.0) que en medios tradicionales.

P: ¿A qué herramienta/servicio propio de la Web 2.0 le ves más proyección en el ámbito empresarial?

R: El blog es un primer paso, pero creo que no hay un servicio "tipo" de Web 2.0 para todas las empresas. Cada empresa debe de buscar su camino y los elementos que mejor le convengan. De hecho, viendo la evolución, el blog se queda casi como algo prehistórico.

3. Futuro

P: ¿Qué aspectos de la Web 2.0 crees que vamos a ver evolucionar en el futuro?

R: Creo que las tecnologías irán convergiendo, y la Web como tal desaparecerá y se fusionará con la radio, televisión y telefonía móvil.

Hace años que se oye hablar de televisión en el teléfono móvil, pero la fórmula aún no ha cuajado. Quizás es que el modelo tenga que cambiar. Tendremos una televisión y una radio "a la carta" con sugerencias de nuestra propia red social, pero llegará a nuestro ordenador, dispositivos móviles, a la puerta de la nevera, etc.

P: Hemos oído hablar a grandes expertos de que a la larga Internet tenderá a eclipsar a los medios de comunicación tradicionales. ¿Hasta qué punto es la Web 2.0 la punta de lanza de esta tendencia?

R: No soy experto en medios de comunicación, pero sí estoy seguro de que los propios medios estudian la manera de aprovechar las web 2.0 en su favor como un complemento a su medio tradicional. No es raro que cualquier periódico, radio o TV, tenga su blog, su canal en Youtube, un perfil en Facebook, ya sea como grupo o página, para de esta modo llegar de una manera más directa a su público consumidor. Es más, en muchos casos crean sus propias herramientas (*widgets*) para que su información sea fácilmente accesible. Un ejemplo es la web de la cadena de TV "La Sexta" <www.lasexta.com> que provee de

widgets o pequeñas aplicaciones <http://www.lasexta.com/widgets> para poder incorporar información de sus programas, o las noticias, dentro de una web, un blog o en el propio escritorio.

P: El uso de Internet creció primero entre las universidades y los particulares. Sólo unos años después aparecieron sus aplicaciones empresariales. ¿Crees que va a pasar lo mismo con el fenómeno 2.0?

R: Sí. La Web 2.0 ha llegado primero entre los usuarios particulares, y las empresas han seguido con el modelo tradicional, pero tarde o temprano las empresas irán cambiando su concepto y se introducirán más utilizando la Web 2.0. Aunque si tardan mucho tendrán que pensar ya en la Web 3.0.

P: Hablando de "Web 3.0", dicen que ahora nos viene una "Web inteligente"... ¿Hasta qué punto crees tú en ello?

R: Hay diferentes creencias sobre la Web 3.0. Hay quien cree que será la "Web Semántica" y hay quien cree que será una "Web 3D" con proliferación de sitios de mundos virtuales al estilo de Second Life. Supongo que los dos conceptos se extenderán, pero a mí me gustaría una mezcla de esto, y déjame poner un par de ejemplos:

Supongamos que es el cumpleaños de un amigo mío y no sé qué regalarle. Yo entraría en mi red social y tecleando su nombre podría obtener una lista de sugerencias según sus gustos y preferencias, y además enlaces a diferentes sitios web donde puedo encontrar los regalos que hacerle.

En otro orden de cosas, supongamos que me quiero ir de vacaciones a un lugar remoto del cual no conozco nada. Poniendo el nombre del lugar se me abriría una página con un resumen de las informaciones más importantes (no enlaces directos, sino un resumen de todo lo encontrado) y además una composición en 3D de algunos paisajes confeccionada a partir de un repositorio fotográfico extraído de la colaboración de los propios usuarios.

Me gustaría creer que será así. De hecho ya se está hablando también del concepto Web 4.0, y no sé si lo que he descrito está más cerca de la Web 4.0 que de la Web 3.0. En cualquier caso, son simplemente numeraciones, pero todo eso se irá traduciendo en una evolución constante donde en algunos casos será difícil decir si nos encontramos en el 3.0 o en el 4.0

Me gustaría que vieras un gráfico de Radar Networks (ver figura 1) donde se indica gráficamente las diferentes evoluciones en términos de "Web X.0". Sabemos dónde estamos, sabemos hacia dónde van las tendencias, pero no sabemos exactamente dónde llegaremos

porque eso está muy condicionado por las propias respuestas y costumbres de los mismos usuarios.

Esta red social, pretende además tener una conexión directa con la Bolsa de Trabajo de manera que esta misma aplicación, cuando

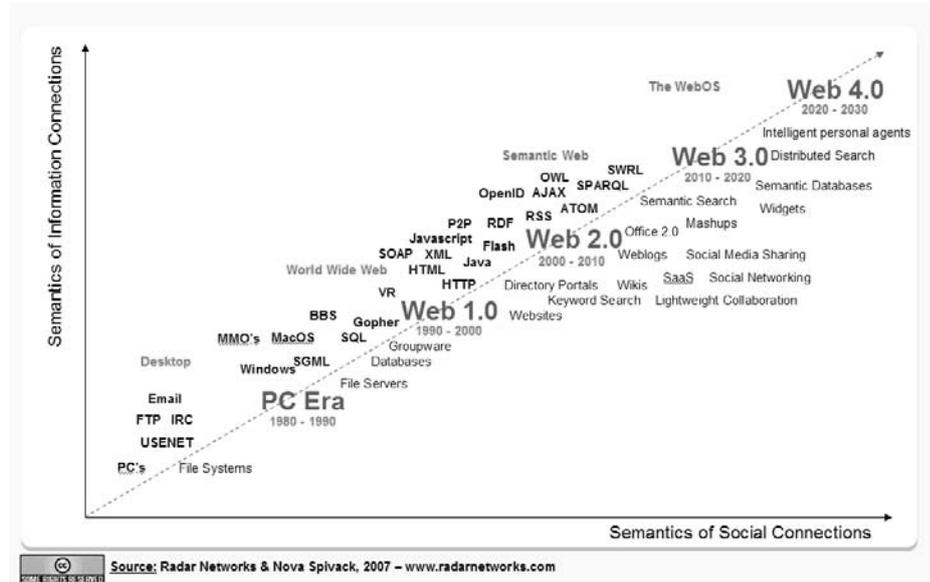


Figura 1. Gráfico evolutivo de la Web.

4. ATI: camino hacia una ATInet 2.0

P: Tú eres el principal ideólogo e impulsor en la Junta de ATI de lo que llamamos ATInet 2.0. ¿Puedes explicarnos los rasgos principales de tu idea y lo que podría llegar a ser ATInet 2.0?

R: ATInet 2.0 es el nombre clave que le dí a un conjunto de ideas y medidas para renovar los servicios de Internet de los socios cuando me incorporé a la Junta Directiva General de ATI.

Este conjunto de medidas abarcaba puntos como la Web, los blogs (que se incorporaron mientras formaba parte de la Junta Directiva del Capítulo de Catalunya), los foros, el Webmail, la Bolsa de Trabajo, y otros temas como videos y Youtube y las redes sociales.

Ampliando uno de los puntos anteriores, surgió el proyecto ATI Network que centraba su interés exclusivamente en las redes sociales.

En dicho documento, y basándome en las características de diferentes redes sociales como (Xing, LinkedIn, Ryze, y FaceBook) explicaba el objetivo del proyecto que es crear una red social exclusiva para los socios de ATI. La necesidad es evidente, ya que dado el volumen de socios que tiene ATI nos conocemos muy poco, y el objetivo de la creación de esta red es potenciar el conocimiento entre socios, crear sinergias, lazos de colaboración y dar a conocer el perfil de cada uno de nuestros socios no solo a nivel interno sino también a nivel externo a modo de ficha de curriculum de manera que pudiera servir para los socios que tengan la necesidad de buscar un puesto de trabajo.

reciba una demanda de trabajo realice una búsqueda dentro de los perfiles de los socios seleccionando los coincidentes con los parámetros requeridos y que éstos puedan recibir las ofertas directamente a sus buzones.

De hecho, estos puntos anteriores son solo algunas pinceladas, de un proyecto que está aprobado pero pendiente de iniciarse por falta de recursos, tanto humanos como económicos.

P: ¿Qué tenemos hasta ahora? ¿Están sentadas ya las bases de ATInet 2.0?

R: Tenemos un proyecto aprobado, con un borrador de lo que debe ser inicialmente, y como base para partir de este hacia un crecimiento mayor. Pero como he dicho anteriormente, no hemos iniciado todavía este proyecto.

Aun así, tenemos una Bolsa de Trabajo muy mejorada y que pretende ser parte de este proyecto futuro, puesto que creemos firmemente que la vinculación debe existir. También tenemos una sección de Blogs, donde cualquier socio puede crearse su blog y contribuir con sus conocimientos y experiencias, o simplemente utilizarlo como página personal.

Cuando se me ocurrió la propuesta yo no tenía ningún blog, no había escrito en ninguno anteriormente. Mi participación se había limitado a los escritos dentro del foro, y al poner en marcha el software de blogs, que se hizo casi en plan experimental, me ví obligado moralmente a abrir mi blog y a intentar mantenerlo vivo, cosa que no hago tanto últimamente. Ahora me gustaría cambiar el software por uno más potente y flexible y estamos

valorando diferentes posibilidades.

También disponemos desde hace unos meses del dominio "atinet.es" el cual se vinculó a los servicios Google Apps, y que han estado en fase de pruebas y usado por unos cuantos socios. No todo el mundo es fan incondicional de Google; yo tengo que confesar que sí lo soy, y esta propuesta vino a raíz de otro problema que padecemos muchos socios. El servidor de correo, se queda corto de capacidad en la mayoría de los casos, así que con este dominio usando los servicios de Google ya no se debe de sufrir por este motivo puesto que se dispone de una capacidad de 7Gb para cada buzón, además del potente servicio de filtro de correo y antisпам, chat, Docs, Sites, etc.

No pretende ser un sustituto del servidor de correo ati.es, sino un complemento. En estos momentos muchos socios lo están usando y desde la cuenta de atinet.es reciben y envían tanto correo de este dominio como con el de ati.es sin temor a perder ningún correo por sobrecarga del buzón de correo de ati.es.

P: ATI como asociación profesional, y en particular sus canales de comunicación, funciona como aglutinadora de información procedente de múltiples fuentes (noticias del sector, informes, anuncios de cursos y conferencias, etc.). Este es uno de los principales servicios a los socios. ¿Hasta qué punto ATInet 2.0 puede permitir mejorar esta clase de servicio?

R: Queremos que ATI sea un punto de referencia, y esta referencia la dan los socios con sus conocimientos y experiencia. Solamente con que un 10% de los socios colaborara escribiendo artículos en blogs, como ya he dicho no necesariamente personales sino blogs temáticos contribuidos por diferentes socios, tendríamos la base de blogs profesionales más extensa de este país con un amplio abanico de especializaciones. Además la intención es poder tener referenciados y catalogados a todos los profesionales de nuestra asociación, hecho que mejoraría y complementaría la bolsa de trabajo en los dos sentidos.

P: Soy socio de ATI, ¿qué puedo hacer desde ahora mismo para empezar a participar en ATInet 2.0?

R: Pues precisamente esto, participar. Necesitamos socios activos, y no solo en temas 2.0, su voz se puede oír también asistiendo a las asambleas y diferentes reuniones convocadas por ATI.

Pero centrándonos en ATInet 2.0, pueden participar abriendo un blog, o colaborando si desean en blogs temáticos o de grupos de trabajo. Un blog no tiene por que ser de un solo socio, puede ser compartido y coparticipado por varios socios para tratar un temática específica.



Figura 2. Gabriel Martí (derecha) presentando sus proyectos ante la JDG de ATI. De izquierda a derecha, algunos miembros de la Junta: Luis Fernández Sanz, Rafael Fernández Calvo, José Onofre Montesa, Josep Molas (presidente) y Olga Pallás.

También se han ido creando grupos en diferentes redes sociales para paliar momentáneamente la falta de nuestra propia red social y dar a conocer nuestra asociación.

Tenemos un grupo en XING <https://www.xing.com/net/ne_atiasociaciondetecnico>, red que por sus características se acerca bastante a la idea de la red social de perfil profesional que deseamos para ATI.

También tenemos un grupo en LinkedIn <http://www.linkedin.com/groups?gid=87964&trk=hb_side_g>, y finalmente un grupo en Facebook <<http://www.facebook.com/home.php#/group.php?gid=26961102102&ref=ts>>. Este último se creó con la intención de integrar no solamente a los socios, sino para que cualquiera interesado en ATI pueda dar a conocer sus actividades. Así fue como posteriormente, viendo las diferentes capacidades de Facebook, se ha procedido a crear una página <<http://www.facebook.com/pages/Asociacion-de-Tecnicos-de-Informatica-ATI/33618141402?ref=ts>> que permite importar noticias RSS directamente desde la web, por lo que ayuda mejor a la difusión de nuestra asociación.

P: ¿Cómo puedo contribuir y cómo podré llegar a contribuir en el futuro a extender esta idea en beneficio de todos?

R: Siendo crítico y participando activamente, ya que estos servicios se implementan por y para los socios.

Pediría a los socios que conocen esta temática y entornos (que seguro que hay muchos)

que nos hagan llegar su opinión y visión de posibles mejoras. Una asociación se nutre y crece con las colaboraciones de todos los socios, y así es como debe de ser. Si hay alguien interesado en contribuir en la mejora de nuestros servicios estaremos encantados.

P: ¿Qué fuerza puede llegar a tener ATI como asociación profesional apoyándose en una ATInet 2.0 potente y mediante la participación activa de sus socios?

R: Mucha fuerza. Puede ser una base de referencia, tanto para empresas, como profesionales y clientes directos, para la búsqueda de profesionales para sus proyectos. Las empresas pueden encontrar nuevos talentos, los socios profesionales independientes podrán encontrar a otros socios colaboradores, y los clientes potenciales pueden encontrar a profesionales que les ayuden a llevar a cabo sus proyectos.

Cambios en Secciones Técnicas de Novática

En consonancia con esta época de inicio del año en la que solemos renovar objetivos y proyectos, hemos impulsado y realizado algunos cambios en nuestras Secciones Técnicas:

■ Como cambio más relevante tenemos la puesta en marcha de una nueva sección llamada "*Tendencias tecnológicas*", una sección desde la que veremos las tecnologías de una manera "transversal", es decir respondiendo a cuestiones acerca de cuáles son las tecnologías que cobran más relevancia en el presente o que presumiblemente la van a tener en el futuro. Los coordinadores de esta sección serán **Alonso Alvarez García** (anteriormente coordinador de la sección "*Entorno digital personal*") y **Gabriel Martí Fuentes** (responsable de tecnología de ATI).

■ Tenemos renovación total en la sección "*Interacción Persona-Computador*". Abandona la coordinación de la sección **Julio Abascal González** para centrar su colaboración con ATI en la representación de la Asociación en el TC13 de IFIP (ver página 4 de este mismo número). Mientras que, siguiendo con la idea de iniciar una colaboración larga y fructífera entre ATI y AIPO (Asociación Interacción Persona-Ordenador), se incorporan en representación de esta última asociación **Pedro Latorre Andrés** y **Francisco Gutiérrez Vela**.

■ En la sección "*Software Libre*" **Israel Herráiz Tabernero** sustituye a **Pedro de las Heras Quirós**. Israel compartirá las responsabilidades de esta sección con **Jesús González Barahona**.

■ Finalmente, **Andrés Martín López** traslada su foco de colaboración, en consonancia con su dedicación profesional actual, desde "*Lenguajes Informáticos*" a "*Entorno digital personal*", donde compartirá tareas con **Diego Gachet Páez**.

A Julio y Pedro, quienes dejan su colaboración con **Novática**, queremos agradecerles todos los esfuerzos que nos han dedicado. Una cálida bienvenida y el deseo de una fructífera interacción a los recién incorporados: Gabriel, Pedro, Francisco e Israel. Mientras que esperamos que Alonso y Andrés sigan sintiéndose bien a gusto desde su nueva sección.

Programación de Novática

Por acuerdo de los Consejos Editoriales de **Novática** y **UPGRADE**, los temas y editores invitados de las restantes monografías del año 2009 serán, salvo causas de fuerza mayor o imprevistos, los siguientes:

Nº 198 (marzo-abril): "Sistemas de Información Geográfica". Editores invitados: **Irene Compte** (Universitat de Girona), **Jordi Guimet** (Institut Cartogràfic de Catalunya), y **Monica Wachowicz** (Wageningen University, Países Bajos).

Nº 199 (mayo-junio): "Software Libre en la empresa". Editores invitados: **Jesús González Barahona** y **Teo Romera Otero** (Libresoft, Universidad Rey Juan Carlos), y **Böjrn Lundell** (Universidad de Skövde, Suecia).

Nº 200 (julio-agosto): "Calidad en los Sistemas de Información". Editores invitados: **Luis Fernández Sanz** (Universidad de Alcalá) y **Darren Dalcher** (National Centre for Project Management, Reino Unido).

Nº 201 (septiembre-octubre): "Gestión de identidades y privacidad". Editores invitados: **Javier López Muñoz** (Universidad de Málaga) y **Miguel Soriano Ibáñez** (Universitat Politècnica de Catalunya).

Nº 202 (noviembre-diciembre): "2010, mirando al futuro tecnológico". Editor invitado: **Alonso Alvarez García** (Observatorio Tecnológico de Telefónica I+D).

Socios institucionales de ati

Según los Estatutos de ATI, pueden ser socios institucionales de nuestra asociación "*las personas jurídicas, públicas y privadas, que lo soliciten a la Junta Directiva General y sean aceptados como tales por la misma*".

Mediante esta figura asociativa, todos los profesionales y directivos informáticos de los socios institucionales pueden gozar de los beneficios de participar en las actividades de ATI, en especial congresos, jornadas, cursos, conferencias, charlas, etc. Asimismo los socios institucionales pueden acceder en condiciones especiales a servicios ofrecidos por la asociación tales como Bolsa de Trabajo, cursos a medida, *mailings*, publicidad en *Novática*, servicio ATInet, etc.

Para más información dirigirse a <info@ati.es> o a cualquiera de las sedes de ATI. En la actualidad son socios institucionales de ATI las siguientes empresas y entidades:

AGENCIA DE INFOR. Y COMUN. COMUNIDAD DE MADRID
 AGROSEGURO, S.A.
 AIGÜES TER LLOBREGAT
 ALMIRALL PRODEFARMA, S.A.
 BARCELÓ CORPORACIÓN EMPRESARIAL, S.A.
 BBR INGENIERÍA DE SERVICIOS, S.L.
 BURKE FORMACION, S.A.
 CÁLCULO, S.A.
 CENTRO DE ESTUDIOS VELAZQUEZ S.A. (C.E. Adams)
 CHOICE, S.A.
 CONSULTORES SAYMA, S.A.
 DEPARTAMENT D'ENSENYAMENT DE LA GENERALITAT
 DIGITAL PARKS
 ELOGOS, S. L.
 EPISER, S.L.
 ESPECIALIDADES ELÉCTRICAS, S.A. (ESPELSA)
 ESTEVE QUÍMICA, S.A.
 FUNDACIÓ BARCELONA MEDIA - UNIVERSITAT POMPEU FABRA
 FUNDACIÓ CATALANA DE L'ESPLAI
 FUNDACIÓN SAN VALERO
 GRUPO BAMESA
 GRUPO CORPORATIVO GFI INFORMÁTICA, S.A.
 IN2
 INFORMÁTICA Y COMUNICACIONES AVANZADAS, S.L.
 INSTITUT D'ESTUDIS CATALANS
 INSTITUT MUNICIPAL D'INFORMÀTICA
 INVERAMA
 ITUM TECHNOLOGY S.A.
 KRITER SOFTWARE, S.L.
 METASINCRON
 NTR - NET TRANSMIT & RECEIVE, S.L.
 OCCIDENTAL HOTELES MANAGEMENT, S.A.
 ONDATA INTERNATIONAL, S.L.
 PRACTIA CONSULTING, S.L.
 QRP MANAGEMENT METHODS INTERNATIONAL
 SOFTWARE, S.L.
 SADIEL, S.A.
 SCATI LABS, S.A.
 SISTEMAS TÉCNICOS LOTERIAS ESTADO (STL)
 SOCIEDAD DE REDES ELECTRÓNICAS Y SERVICIOS, S.A.
 SOGETI ESPAÑA, S.L.
 SOPORTES, SISTEMAS, SOFTWARE, S.L.
 SOS, S.A.
 TECNOLOGIA Y CALIDAD DE SOFTWARE, S.A.
 TISA ORDENADORES, S.A.
 TRAINING & ENTERPRISE RESOURCES
 T-SYSTEMS ITC Services España S.A.
 UNIVERSIDAD ANTONIO DE NEBRIJA
 UNIVERSIDAD DE EXTREMADURA - E. POLITÉCNICA DE CÁCERES
 UNIVERSITAT DE GIRONA
 UNIVERSITAT OBERTA DE CATALUNYA