

## LA VISIÓN INTEGRADORA DEL PROYECTO QUATRO EN LA APLICACIÓN DE LA WEB SEMANTICA PARA GARANTIZAR UN INTERNET DE CONFIANZA

M.A. MAYER<sup>1</sup>, A. LEIS<sup>1</sup>, V. KARKALETIS<sup>2</sup>, P. ARCHER<sup>3</sup>, K. STAMATAKIS<sup>2</sup>, A. PEREGO<sup>4</sup> y P. RUIZ<sup>1</sup>

<sup>1</sup>Web Médica Acreditada. Colegio Oficial de Médicos de Barcelona. 08017 Barcelona. España.

<sup>2</sup>National Center for Scientific Research (NCSR) "Demokritos". Institute of Informatics & Telecommunications. Atenas. Grecia.

<sup>3</sup>Internet Content Rating Association (ICRA). Brighton. Reino Unido.

<sup>4</sup>University of Milan. Department of Informatic and Communication. Milán. Italia.

*En este artículo se presenta el proyecto europeo QUATRO (Quality Assurance and Content Description) que ha aplicado las tecnologías de la web semántica a los sellos de calidad o marcas de confianza. El proyecto ha definido un vocabulario que puede ser usado por cualquier sello de calidad y ha desarrollado una plataforma técnica que permite que cualquier tipo de sello pueda ser procesado por agentes web semánticos. La calidad de los contenidos sanitarios en Internet ha planteado la necesidad de implementar esta tecnología en los sistemas de acreditación médicos. La web semántica se constituye como el medio que permitirá, por un lado, a los profesionales sanitarios adaptarse a los cambios que continuamente incorpora Internet, y por otro, a los usuarios de salud, poder conocer y entender mejor la información sanitaria obtenida en este medio, contribuyendo a su educación sanitaria de una forma más activa.*

### 1. Introducción

La integración de las diversas herramientas que ofrece Internet tanto en el diseño de páginas web, en la utilización de los metadatos, los motores de búsqueda y los diferentes estándares en los que se está trabajando como la accesibilidad o el intercambio de datos, se constituyen como una oportunidad fundamental para poder acceder de la manera más eficiente a la gran cantidad de información que ya se encuentra disponible en este medio, tanto para los usuarios como para los profesionales. Entre los temas de mayor interés para los usuarios de Internet están todos aquellos relacionados con la salud. La calidad de estos contenidos es muy diversa, podemos encontrar información de medicina basada en la evidencia así como también recomendaciones de dudosa utilidad e incluso peligrosas. El efecto y el impacto que este hecho puede tener sobre la salud de los usuarios y pacientes es todavía difícil de conocer pero los expertos están de acuerdo en que se trata de un tema de salud pública cada vez más preocupante.

Respecto a la calidad de los contenidos sanitarios en Internet se han planteado y puesto en funcionamiento diversas propuestas con la idea de aportar soluciones a esta problemática. Entre ellas está la creación de sistemas de filtrado, acreditación y revisión de contenidos, que garanticen que las webs acreditadas disponen de unos criterios mínimos de calidad y

responsabilidad profesional que aporten la confianza necesaria a los usuarios de salud que acceden a dichas webs.

Por otro lado, entre los estándares que se están desarrollando desde el World Wide Web Consortium (W3C) [1] y que se basan en metadatos, debemos destacar todo los aspectos relacionados con la Web Semántica (WS) [2]. La web actual está basada fundamentalmente en documentos escritos en formato HTML. Este formato tiene una serie de deficiencias que no permiten una descripción ni una clasificación adecuada de sus contenidos y por tanto no es posible realizar capturas inteligentes, ni tratamiento de información de una forma estandarizada, lo que dificulta la comprensión correcta entre los conceptos “humanos” y lo que buscan y encuentran los motores de búsqueda en Internet. La WS permite la introducción de “metadatos” (datos sobre datos) en las páginas web, de tal forma que esas etiquetas de contenidos facilitan la comprensión y la estandarización de conceptos entre los diferentes recursos presentes a la web, los buscadores de información y las necesidades de los usuarios.

En esta línea se está desarrollando el proyecto europeo QUATRO (Quality Assurance and Content Description) [3] basado en la Web Semántica. En este proyecto participa el sistema de acreditación, Web Médica Acreditada (WMA) que Colegio Oficial de Médicos de Barcelona puso en marcha en 1999 [4]. La experiencia adquirida en estos años por WMA, en la acreditación de webs de contenido sanitario así como en la aplicación de la web semántica mediante su participación en el proyecto europeo MedCIRCLE [5], la ha situado entre las 5 iniciativas de acreditación más conocidas alrededor del mundo tal y como se ha publicado recientemente en un estudio internacional realizado por la Health on the Net Foundation (HON) [6,7], institución referente en esta actividad. Esta tecnología puede permitir a los profesionales sanitarios adaptarse a los cambios que continuamente incorpora Internet, y por otro, a los usuarios de salud, poder conocer y entender mejor la información sanitaria obtenida en este medio, contribuyendo a su educación sanitaria de una forma más activa.

## **2. Objetivos del proyecto QUATRO**

En Europa existe una gran variedad de iniciativas de calidad, algunas de cuales están financiadas por otros programas de la Unión Europea. En cada caso, después de un proceso de revisión, se concede el sello de acreditación a una web. El usuario debe estar en la web para conocer la existencia del sello de calidad y no siempre es fácil que lo visualice ya que sólo se muestra en la página de inicio. Para obtener más información sobre la acreditación, el usuario deber clicar en el sello o marca y se mostrará una entrada de una base de datos que confirma la validez del logotipo.

Sin embargo, estos sellos de calidad como los concedidos por WMA, deben ser suficientemente visibles para los motores de búsqueda y otros agentes, sino podría limitarse su utilidad. Estas etiquetas funcionan de manera aislada por lo que no pueden ser recopiladas, agregadas o utilizadas de otra manera por ninguna máquina.

Ante esta problemática el objetivo fundamental del proyecto QUATRO, es desarrollar una plataforma común para permitir que todos los tipos de sistemas de etiquetado y acreditación sean legibles por las máquinas sin que esto afecte a su proceso de evaluación ni a su independencia, ofreciendo a los usuarios finales la posibilidad de comprobar que el sello es fiable y cumple ciertos criterios. Los sellos de calidad, también llamadas marcas de confianza, pueden usarse en diferentes soportes de contenido digital, como páginas webs y teléfonos móviles. Estas etiquetas legibles por máquina permitirán a los motores de búsqueda mostrar

logotipos junto a los resultados de las búsquedas, y así los usuarios podrán ver inmediatamente qué webs han sido revisadas por las organizaciones o sistemas de acreditación. Además, los usuarios finales también deben tener la certeza de que la etiqueta sigue siendo válida y no se trata de una falsedad y que la información realmente es fiable.

La plataforma proporcionará soporte para facilitar a los proveedores de contenidos añadir sellos de calidad a sus webs y que sean “visibles” tanto para humanos como para los motores de búsqueda, se validará de forma automática el sello en lugar que el operador realice una comprobación su base de datos lo que permitirá la verificación del sello en tiempo real.

### *2.1. Necesidades de los usuarios*

Para conocer de forma adecuada las necesidades de los posibles usuarios, se ha llevado a cabo un estudio para saber lo que esperan éstos de los sellos de calidad. Se beneficiarían de la aplicación de las utilidades proyectadas en QUATRO:

- Motores de búsqueda y portales
- Generadores de contenidos
- Usuarios finales
- La industria
- Sistemas de acreditación y sellos de confianza
- Empresas del medio

Finalmente, en el estudio se observó que la confianza en las webs aumenta cuando aparece un sello de acreditación y es más probable que los usuarios finales visiten una web, si desde los resultados del motor de búsqueda se distingue si tiene sello. Los generadores de contenido perciben los sellos de calidad como un posible beneficio para sus modelos de negocio y están a la expectativa de ver cómo trabajará QUATRO y qué coste de implementación tendrá. La mayoría de los interesados creen que las webs de información sanitaria y educación son las tienen una mayor exigencia de sellos de calidad.

## **3. La arquitectura de QUATRO**

### *3.1. El Schema RDF de QUATRO para elementos comunes de los sellos de calidad*

QUATRO ha publicado un vocabulario con la intención de poder ser utilizado por una gran variedad de marcas y sellos de confianza. Este vocabulario se muestra como un conjunto de criterios de calidad con su respectiva definición para ser comprensible por los usuarios así como el “Schema XML/RDF” para el acceso y utilización por parte de los motores de búsqueda. Este vocabulario, que presenta una parte comprensible para los humanos y otra en formato RDF (Resource Description Framework) [8] para las máquinas, está dividido en cuatro apartados generales: criterios generales, criterios para garantizar la calidad de contenidos, cumplimiento de la legislación vigente para el comercio electrónico y publicidad, criterios propios de cada sistema de acreditación. Este vocabulario ha sido creado para ser lo más flexible posible y que cada sistema de acreditación pueda adaptarlo a sus necesidades y aplicación [9,10].

### *3.2. Plataforma del QUATRO*

El proyecto QUATRO ha desarrollado dos herramientas complementarias que ofrecen importantes mejoras de cara a permitir que un gran número de etiquetas de calidad diferentes sean altamente interoperables.

QUATRO proporciona a los usuarios que navegan por Internet, los medios necesarios para ser avisados cuando se detecte la presencia de etiquetas o sellos de calidad y de si son válidas o no. Las aplicaciones que permitirán esta funcionalidad al usuario final son:

- ViQ, es una barra de herramientas que reconoce la presencia de sellos en una web que está siendo visitada y añade el icono adecuado al navegador.
- LADI (Label Display Interface), es una interfaz de visualización de etiquetas para resultados de motores de búsqueda que anota los resultados de dichos motores

En la figura 1 podemos observar cómo ViQ presenta en la barra de herramientas del navegador las autoridades de acreditación reconocidas (en este ejemplo ICR, IQUA y WMA) y cómo acceder a más información sobre el RDF Schema propio de cada una de ellas.



**Figura 1:** Barra de herramientas ViQ de QUATRO

### 3.3. El servidor QUAPRO

QUAPRO desempeña un papel central en la arquitectura QUATRO. Se trata de una aplicación basada en un servidor que recibe solicitudes de ViQ o LADI, identifica los sellos de calidad, los evalúa y responde acordeamente. Si bien los usuarios finales sólo verán las herramientas ViQ y LADI, el QUAPRO se encargará de localizar sellos y valorar su validez.

El proceso paso a paso sería el siguiente:

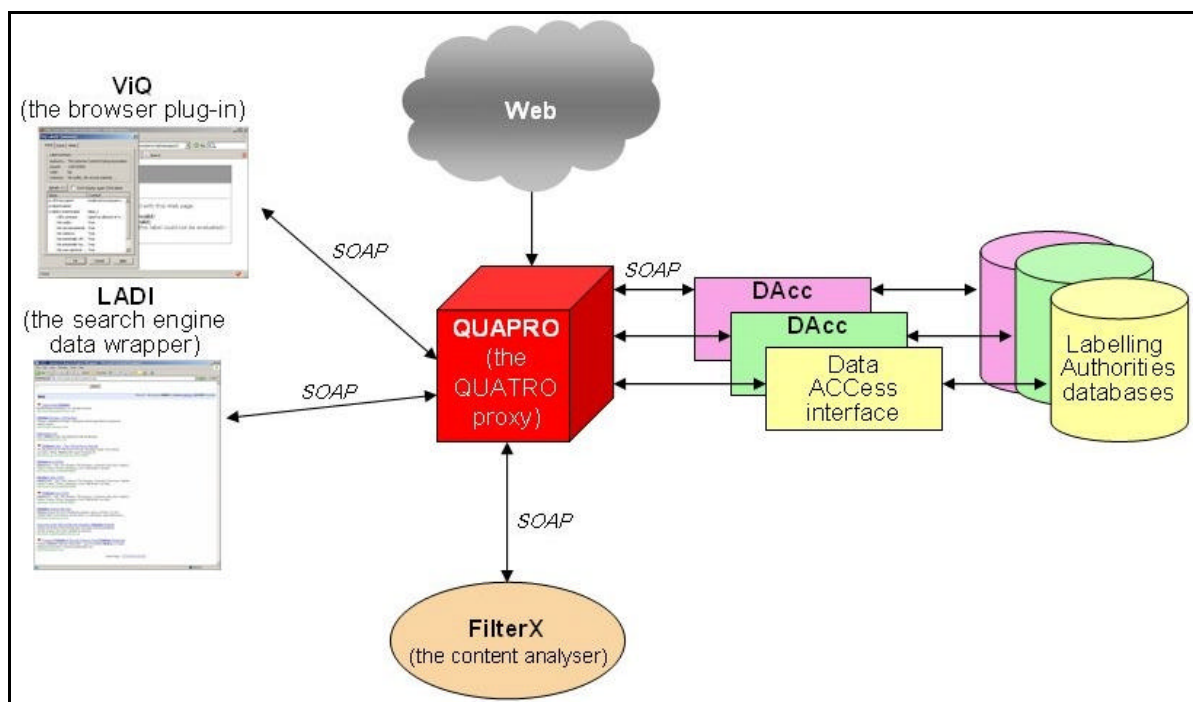
- ViQ es una extensión de navegador que muestra sellos cuando están presentes en un sitio web. Envía una única URL cada vez a QUAPRO y obtiene la respuesta apropiada que indica

si hay sellos presentes o no. Para cada sello identificado, ViQ muestra el gráfico pertinente que representa su estado de validez (válida, inválida o no verificable).

- LADI, es una interfaz para motores de búsqueda a la que puede accederse a través de ViQ (ya añadida en la barra de búsqueda del navegador). LADI reenvía la solicitud del usuario a través de los motores de búsqueda y muestra sellos de calidad, si los encuentra, junto a los resultados obtenidos.

- DAcc (Data Acces Interface: DAcc), interfaz de acceso de datos, permite a QUAPRO acceder a las bases de datos de las autoridades u organizaciones de acreditación y realizar la comprobación de la veracidad e integridad de los sellos. Este sistema permite a todas los sistemas de acreditación que quieran adoptar el modelo QUATRO, elaborar su propia Dacc usando el protocolo simple de acceso a objetos (SOAP).

- QUAPRO es la aplicación intermedia entre ViQ, LADI y las Autoridades de Acreditación y Calidad como son: WMA, IQUA o ICRA. Se comunica por un lado con LADI o ViQ y, por el otro, con las base de datos de las autoridades de etiquetado a través de sus DAcc. QUAPRO recibe URLs de la web a valorar y la visita en busca de si dispone información sobre alguno de los sellos reconocidos. Cuando encuentra uno o más sellos de calidad, QUAPRO indica su presencia, y a continuación, devuelve un mensaje que contiene información sobre la validez de este sello.



**Figura 2:** La arquitectura del QUAPRO

QUAPRO evidentemente juega un papel clave en el modelo de confianza ya que sólo trabajará y validará los RDF Schema de las organizaciones de acreditación que hayan sido previamente reconocidos, dificultando que los resultados se puedan imitar o copiar.

#### 4. Conclusiones

QUATRO propone el uso de las tecnologías de la web semántica para abordar el principal problema de los actuales sistemas de acreditación, es decir, la necesidad de una mayor visibilidad en los motores de búsqueda así como una mayor confianza en los sistemas de acreditación o filtrado de las páginas webs, un proceso que requiere un gran esfuerzo humano. Se espera que la tecnología resultante tenga un impacto significativo en los sistemas de acreditación médicos, ayudando al trabajo de los expertos e incrementando el número de webs acreditadas alrededor de Europa y su efectiva verificación. La web semántica se constituye como la herramienta necesaria para mejorar el conocimiento de la calidad sanitaria divulgada en Internet, contribuyendo a la educación en salud de los usuarios.

#### Agradecimientos

El QUATRO está financiado por el programa “Safer Internet Action Plan” de la Unión Europea y es producto de la colaboración de todas las instituciones participantes sin las que no sería posible su realización. Los participantes del proyecto son: Coolwave (Reino Unido), ECP.NL (Holanda), ERCIM, ICRA (Reino Unido), IQUA (España), NCSR “Demokritos” (Grecia), Pira International (Reino Unido), Universidad de Milán (Italia), WMA (España).

#### Referencias

- [1] World Wide Web Consortium (W3C). Página principal. Disponible en: <http://www.w3.org/>.
- [2] W3C. Semantic Web. Disponible en: <http://www.w3.org/2001/sw/>.
- [3] The Quality Assurance and Content Description Project (QUATRO). Página principal. Disponible en: <http://www.quatro-project.org>.
- [4] Mayer MA, Leis A, Sarrias R, Ruiz P. Web Médica Acreditada Guidelines: Reliability and Quality of Health information on Spanish-language websites. En: R. Engelbrecht et al (Eds). *Connecting Medical Informatics and Bioinformatics*. MIE 2005: Proceedings of the 19th International Congress of the European Federation for Medical Informatics (CD-ROM). Geneve, Switzerland; 2005. Vol I, No. 1, 2005. p.1287-92.
- [5] Mayer MA, Darmoni SJ, Fiene M, Köhler C, Roth-Berghofer TR, Eysenbach G. MedCIRCLE: Collaboration for internet rating, certification, labelling and evaluation of health information on the World-Wide-Web. En: *The New Navigators: from Professionals to Patients*. R. Baud et al. (Eds). IOS Press. Proc MIE 2003:667-672.
- [6] Analysis of 9th HON Survey of Health and Medical Internet Users. Disponible en: <http://www.hon.ch/Survey/Survey2005/res.html>.
- [7] Health on the Net Foundation. Página principal. Disponible en: <http://www.hon.ch>.
- [8] RDF Primer. Disponible: <http://www.w3.org/TR/rdf-primer/>.
- [9] QUATRO vocabulary. Disponible en: <http://www.quatro-project.org/vocabulary/1.0/>.
- [10] Mayer MA, Karkaletsis V, Archer P, Ruiz P, Stamatakis K, Leis A. Quality Labelling Medical Web Content. *Health Informatics Journal* 2006;**12**:81-87.