

Cientes ricos colonizando la web semántica. Porqual

José Miguel Samper Sánchez, jmsamper@formacion2020.es
César Sancho Arribas, csancho@rorcual.com

Abstract. El objetivo más ambicioso del W3C, la web semántica, es un terreno por colonizar. Seis años después de la creación del estándar RDF, la idea de una red de datos entendibles por las máquinas y una comunidad de agentes explotando esos datos aún se ve lejana. La Web 2.0 propone avanzar hacia la web semántica como una realidad presente, dando pasos pequeños pero seguros, mediante técnicas como microformatos o RSS, en espera de tecnologías más avanzadas que necesitan un universo de datos puros. Para ello es necesario que los contenidos estén claramente aislados de la presentación. Los clientes ricos Ajax, Flash, Java o XUL permiten publicar datos puros sin renunciar a interfaces espectaculares. La aplicación Porqual, de código abierto, usa un cliente rico para mostrar contenidos XHTML y RDF.

1. La Web 2.0

Web 2.0 es un concepto publicitario ideado por O'Reilly [1] y MediaLive International [2] para referirse a las metodologías en boga en el desarrollo web. Las tecnologías que conforman la Web 2.0 son objeto de polémica, porque cada cual defiende su tecnología preferida o que cree que triunfará en los próximos años. Sin embargo, podemos tomar como referencia los conceptos de los que se habla en la Web 2.0 Conference [3], que se celebra anualmente desde el año 2004. Éstos son:

- Páginas estructuradas mediante XHTML y microformatos.
- Clientes web ricos, mediante Ajax, Flash, Java, SVG o XUL.
- Webs colaborativas: blogs, tags, folksonomías.
- Sindicación de contenidos: RSS.
- Paradigma REST y servicios web (SOAP).

Podemos leer una introducción sobre la Web 2.0 en la Wikipedia [4] y en el artículo “*What is Web 2.0?*” [5] de Tim O'Reilly.

2. Porqual

Porqual es una aplicación de código abierto (licencia GPL) promocionada principalmente por la empresa hispana Formación 2020. Porqual permite crear con facilidad sitios web que siguen las tendencias de la web 2.0. Las webs con Porqual tienen el contenido estructurado semánticamente usando XHTML, microformatos y RDF. La interfaz de usuario es totalmente Flash sin abandonar la accesibilidad, adoptando las recomendaciones de la WAI [6]. Un sitio web hecho con Porqual gestiona documentos mediante metadatos, usando la base de datos RDF [7] Sesame [8].

3. Clientes web ricos

El formato HTML es ideal para la presentación de documentos, habitualmente formados por textos estructurados e imágenes. Sin embargo, el soporte para interfaces de usuario del estándar HTML es escaso, limitándose básicamente a hipervínculos, botones y elementos de formulario. Por ello, cuando se ha necesitado una mayor interacción con el usuario o se ha querido mostrar contenido multimedia (animaciones, sonidos, videos) se han utilizado otras tecnologías que los navegadores modernos permiten ejecutar. Cuando se usan estas tecnologías adicionales hablamos de clientes web ricos. Veamos los más populares.

3.1. Javascript y AJAX.

Javascript permite la manipulación dinámica del contenido de la página, así como la respuesta a

acciones del usuario y a otros eventos. Al ser muy cercano al HTML es ideal para pequeños efectos sobre el contenido (por ejemplo, ocultar determinados contenidos, ordenar listas) y sobre la interacción del usuario (por ejemplo, ayudar a rellenar formularios).

Se llama AJAX al uso de Javascript para acceder a datos remotos sin recargar la página.

Los aspectos más negativos de esta tecnología son la dependencia del navegador (el código sólo funcionará en ciertas versiones de ciertos navegadores) y los problemas de ejecución (como pérdidas de memoria) asociadas a este lenguaje.

Por tal motivo, Javascript puede usarse para ayudar en la interacción de usuario, pero las páginas deberían ser perfectamente funcionales en un navegador sin Javascript.

Podemos ver un ejemplo de los complejos interfaces de usuario que se pueden hacer con AJAX en la web de Zimbar Incorporated [9]. Han realizado un avanzado webmail con multitud de opciones y una agradable experiencia de usuario. Sin embargo, la página no se puede ver sin un navegador compatible.

3.2. Java Applets y JWS.

Entre los primeros clientes ricos estuvieron los applets de Java. Actualmente la tecnología ha evolucionado mucho y además existe el sistema alternativo Java Web Start, que permite la ejecución remota de aplicaciones.

De entre los clientes ricos, los basados en Java son los que más potencia permiten, porque tienen detrás un lenguaje maduro y robusto con multitud de opciones.

Para mostrar la potencia de esta opción veamos la suite ofimática on-line realizada por la empresa ThinkFree Corporation [10].

3.3. Flash

Flash es una tecnología muy adecuada para mostrar contenidos multimedia así como para realizar interfaces de usuario. Debido a la vistosidad de los elementos Flash, esta tecnología vivió una época de gran popularidad con muchas páginas que usaban Flash sin necesitarlo, páginas que no eran accesibles ni usables. Tras esa época hubo un movimiento contra las páginas con Flash, sobre todo a partir de la publicación del artículo “*Flash 99% Bad*” de Jakob Nielsen [11].

Las ventajas de Flash sobre las otras tecnologías que se usan para clientes web ricos son:

- Es la más adaptada a la creación de interfaces de usuario.
- Permite fácilmente realizar animaciones y mostrar contenido multimedia.
- Casi todos los navegadores soportan Flash. Según nuestras estadísticas, el 95% de los navegadores soportan AJAX, el 97% soportan Java y el 99% Flash.
El plug-in de Java, además, ocupa más que el de Flash (16 Mb frente a 1 Mb) y tiene una instalación más compleja.
- Para realizar un interfaz en Flash no es necesario tener conocimientos de programación, por lo que es más fácil encontrar desarrolladores Flash que desarrolladores Java o AJAX.

Porqual utiliza Flash para mejorar la experiencia de usuario (presentar contenidos de manera más compacta y cargarlos más rápidamente) y la usabilidad de las páginas, pero sin que ello suponga perder las características de las webs en HTML (accesibilidad, indexación en buscadores, posibilidad de imprimir la página o ver el código fuente, uso de los botones del navegador).

3.4. Otros clientes ricos.

Aunque las tres tecnologías anteriores son las más usadas no son las únicas a la hora de desarrollar clientes web ricos.

Por un lado tenemos SVG [12], que es un estándar promovido por el W3C. Actualmente los

navegadores Opera y Firefox dan soporte a un subconjunto de la especificación SVG. La importancia creciente de los clientes ricos ha llevado al W3C a crear varios grupos de trabajo dedicados a estandarizar formatos y APIs para desarrollar clientes ricos [13].

Por otro lado tenemos tecnologías propias del navegador, como los ActiveX de Microsoft Internet Explorer o XUL de Mozilla. XUL es un lenguaje muy adecuado para hacer clientes web ricos; además el navegador Mozilla se encuentra disponible para plataformas Windows, Linux y Macintosh.

4. Contenidos semánticos

4.1. Estructura

Como hemos comentado anteriormente, HTML es un formato adecuado para la presentación de textos estructurados e imágenes. La popularidad de internet hizo que se buscara una mayor vistosidad de las páginas web, por lo que los navegadores fueron añadiendo características en sus implementaciones de HTML para facilitar el diseño gráfico de las páginas. Esto llevó a la existencia de muchísimas páginas que sólo se podían visualizar en una versión de un navegador, y donde el contenido y la presentación se mezclaban completamente, imposibilitando extraer información automáticamente de las mismas.

Afortunadamente en los últimos años el esfuerzo se ha concentrado en la estructuración de los contenidos y la separación de presentación y contenido. Esto se consigue con la adopción de los estándares XHTML y CSS.

Cuando se usa XHTML correctamente para estructurar el contenido, un agente automático puede entender partes del contenido de una página. Por ejemplo, si dentro de una página tenemos

```
<dt>Rocín</dt>
<dd>Caballo de mala traza, basto y de poca alzada.</dd>
```

un agente sabe que lo que se encuentra dentro del tag *dd* es el significado de la palabra que se encuentra en el tag *dt*. Cosa que no podría hacer si tuviésemos algo como

```
<tr>
  <td>Rocín</td>
  <td>Caballo de mala traza, basto y de poca alzada.</td>
</tr>
```

En la siguiente versión de la especificación HTML el W3C da un paso más hacia la estructuración y el significado de los contenidos. En XHTML 2.0 [14] hay menos elementos que no corresponden a contenido (como `
`, `` o `<map>`) y más soporte para significado. Por ejemplo, si queremos hacer una lista de enlaces (un menú) actualmente haríamos

```
<li><a href="info.html">Información</a></li>
<li><a href="content.html">Contenidos</a></li>
```

mientras que en XHTML 2.0 podremos hacer

```
<li href="info.html">Información</li>
<li href="content.html">Contenidos</li>
```

que es más correcto: cada elemento es una opción del menú. En la primera versión hay innecesariamente dos elementos para una opción del menú. La segunda es más fácil de entender por un agente automático en búsqueda de información o un cliente rico que muestre el menú en pantalla.

4.2. Accesibilidad

El seguimiento estricto de los estándares hace posible, además, que una página sea comprensible por diferentes clientes, ya sea un navegador moderno, un navegador antiguo, un teléfono móvil, un navegador para ciegos o un robot de un motor de búsqueda.

La accesibilidad se convierte así en una exigencia para las páginas de la Web 2.0. El desarrollo de clientes ricos no debe entorpecer el acceso a los contenidos. Por ejemplo, si un enlace en una página está dentro de un objeto Flash o se ejecuta mediante código Javascript puede ocurrir que ese enlace no se pueda seguir en un navegador sin soporte para esas tecnologías.

4.3. Web semántica

Comentábamos que el seguimiento estricto y correcto de los estándares facilita la comprensión de una página por un agente automático. El agente puede discernir algunas partes de la página, como títulos, párrafos o citas.

La Web Semántica es un proyecto que pretende que las webs sean comprensibles por máquinas, para poder ofrecer servicios avanzados (búsquedas avanzadas, relación de datos) a partir de la información que obtienen de las páginas. El lenguaje que se ha de usar para describir los contenidos de las webs es RDF [7].

Para muchos la Web 2.0 es la web semántica. Pero la web semántica debe existir paralelamente a la web actual [15]. Aunque esta iniciativa está impulsada por el W3C, actualmente no se ha implementando de una manera general. Lo que sí que se está extendiendo es el uso de pequeños subconjuntos, tal como puede ser RSS.

Podemos experimentar con algunas herramientas de la web semántica: el buscador semántico Swoogle [16], el visor de metadatos Dublin Core Viewer [17] y el navegador semántico Simili [18]. Las páginas hechas con Porqual están preparadas para la web semántica, como podemos ver con las utilidades anteriores.

4.4. Microformatos

Entre lo que tenemos ahora (XHTML) y lo que queremos en el futuro (RDF) hay posiciones intermedias. En XHTML se puede dar significado a algunos contenidos de la página. En la web semántica se podrá expresar el significado de cualquier elemento de la página. Sin embargo, como la web semántica queda lejos, podemos adoptar soluciones a corto plazo.

Una de estas soluciones son los microformatos [19]. Un microformato es una convención de uso de HTML. Por ejemplo, tenemos el microformato *rel-directory*, que se usa así

```
<a rel="directory" href="parent.html">Productos</a>
```

El atributo *rel* indica que el enlace anterior es un enlace a la página padre, es decir, a una página en la que hay un enlace a la página actual. El cliente rico de mi aplicación puede aprovechar el microformato para indicar al usuario con un icono que este enlace le lleva a la página padre.

Hay microformatos para blogs, contactos y otras cosas. Podemos ver otro ejemplo en [20].

Si un gran número de páginas usan un microformato vale la pena implementar un agente automático que entienda dichas páginas, del mismo modo que ha valido la pena implementar lectores de RSS porque hay muchas páginas que sirven RSS.

La especificación de microformatos se realiza colaborativamente y el uso de los mismos es lo que determinará cuáles triunfen y cuáles fracasen. Podríamos decir que los microformatos son la versión social del grupo de trabajo de HTML del W3C.

4.5 Paradigma REST

REST [21] es una colección de principios sobre la arquitectura de las aplicaciones web. Estos

principios benefician a la web semántica porque permiten que las páginas sean más fácilmente accesibles y comprensibles por agentes automáticos.

Por ejemplo, una URL debe definir únicamente un recurso, sin estado (sin necesidad de estar en una sesión web o tener ciertas cookies). Esto hace que un agente de usuario pueda acceder siempre a ese recurso e identificarlo correctamente, cosa que no ocurriría si la misma URL da acceso a diferentes recursos según las circunstancias.

5. Conclusiones

La adopción de las prácticas expresadas anteriormente harán posible una web donde

- Los usuarios usan con más facilidad las webs y disfrutan más usándolas. Los usuarios con discapacidades tampoco tienen problemas para acceder a los contenidos.
- Los agentes automáticos pueden comprender los contenidos de la web y utilizar esa información para proporcionar servicios avanzados.

Por lo tanto, no hay Web 2.0 sin cumplimiento de estándares y sin accesibilidad universal. Los clientes ricos deben usarse para ayudar a los usuarios, pero sin romper los estándares ni la accesibilidad.

Porqual está concebido para generar webs que cumplan estos principios. Las webs hechas con porqual tienen un cliente rico hecho en Flash, cumplen con los estándares del W3C, son totalmente accesibles, proporcionan contenido RDF para la web semántica, adoptan varios microformatos y siguen el paradigma REST.

6. Referencias

- [1] O'Reilly Media, <http://www.oreilly.com/>
- [2] MediaLive International, <http://www.mlii.com/>
- [3] Web 2.0 Conference, <http://www.web2con.com/>
- [4] Artículo Web 2.0 en la Wikipedia, <http://en.wikipedia.org/wiki/Web2.0>
- [5] Tim O'Reilly, "What is Web 2.0?", <http://www.oreillyn.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [6] Recomendaciones de la Iniciativa por la Accesibilidad Web (WAI), <http://www.w3.org/TR/WCAG10/>
- [7] Especificación RDF, <http://www.w3.org/RDF/>
- [8] Sesame, base de datos RDF, <http://www.openrdf.org/>
- [9] Completo webmail hecho con AJAX, <http://www.zimbra.com/> (ver "hosted demo").
- [10] Suite ofimática en un applet, <http://online.thinkfree.com/>
- [11] Jakob Nielsen, "Flash 99% Bad", <http://www.useit.com/alertbox/20001029.html>
- [12] Especificación de SVG, <http://www.w3.org/TR/SVG/>
- [13] W3C Rich Web Clients, <http://www.w3.org/2006/rwc/>. Incluye un grupo sobre composición de documentos web, otro sobre APIs para desarrollo de clientes ricos y otro sobre estandarización de formatos para clientes ricos.
- [14] XHTML 2.0, <http://www.w3.org/TR/xhtml12/>
- [15] "The semantic web: an interview with Tim Berners-Lee", <http://www.consortiuminfo.org/bulletins/semanticweb.php>
- [16] Swoogle, buscador semántico, <http://swoogle.umbc.edu/>
- [17] Dublin Core Viewer, <http://www.splintered.co.uk/>
- [18] Simile – Piggy Bank, navegador semántico del MIT, <http://simile.mit.edu/piggy-bank/>
- [19] Microformatos, <http://microformats.org/>
- [20] Ejemplo del microformato XMDP, <http://gmpg.org/xmdp/samplehtmlprofile>

Sancho, C. y Samper, J.M., *Webs con contenidos semánticos y clientes ricos. Porqual*

[21] Paradigma REST en la Wikipedia, <http://en.wikipedia.org/wiki/REST>