

Document downloaded from:

<http://hdl.handle.net/10251/65171>

This paper must be cited as:

Gil Salinas, JA.; Pont Sanjuan, A. (2015). !Vaya tela! Atrapados por la Web. *Novática*. 41(234):102-104. <http://hdl.handle.net/10251/65171>.



The final publication is available at

<http://www.ati.es/novatica/2015/234/nv234sum.html>

Copyright Asociación de Técnicos de Informática (ATI)

Additional Information



!Vaya tela! Atrapados por la Web

José A. Gil y Ana Pont

Departament d'Informàtica de Sistemes i
Computadors, Universitat Politècnica de
València

{jagil, apont}@upv.es

Pocas aventuras en la historia de la ciencia y la tecnología han tenido un éxito tan rápido y masivo como la World Wide Web (WWW). Desde que en 1991 se hiciera de público acceso el primer servidor y se publicara la primera página web, tomando como base las ideas planteadas durante 4 años por Tim Berners-Lee en el CERN y desarrolladas con la colaboración de Robert Cailliau, la web ha evolucionado hasta llegar a ser hoy en día nuestra más importante interfaz tanto con el mundo real como el virtual.

Aquel proyecto, concebido inicialmente como un mero sistema para el intercambio de información científica (artículos e informes técnicos con texto e imágenes), ha ido experimentando continuos cambios en su aspecto, funcionalidad e infraestructura y ganando terreno en el ámbito social y económico, cubriendo hoy en día un amplio espectro de aplicaciones. Tareas cotidianas como leer la prensa, hacer compras, tramitaciones con la administración, enseñar, aprender, hacer el seguimiento y controlar la evolución de los pacientes, gestiones bancarias, conocer la predicción del tiempo, intercambiar fotografías, o simplemente departir con los amigos, son un claro ejemplo de cómo la web forma parte de nuestras vidas.

Pero, ¿qué más podemos esperar de ella? ¿Cuál va a ser la web de los próximos años? Para intentar dar una respuesta tenemos que analizar cómo ha ido cambiando en los últimos tiempos tanto desde el punto de vista de su contenido, estructura y de la infraestructura que la soporta.

El contenido. Los documentos de hipertexto fueron el origen de la WWW (en adelante la Web), consistiendo en mucho texto, estructuras básicas y soporte básico para los hiperenlaces. Con la aparición de las páginas web se extiende la definición de hipertexto añadiendo recursos hipermedia adicionales, como imágenes y audio, así como otros elementos que permiten elegantes diseños visuales, pero no muy interactivos y no muy diferentes a una página impresa.

Con la utilización de lenguajes como JavaScript, que permite la modificación de la página una vez cargada y la interacción con el usuario, se crea el HTML dinámico (DHTML) y con la incorporación de AJAX que permite interacciones múltiples con el servidor sin abandonar la página, las páginas web se transforman en Aplicaciones Interactivas. Los usuarios han dejado de ser espectadores y se convierten en actores de una Web en la que se fomenta el intercambio de información y la colaboración. Comienza así una nueva era de complejos gráficos de dependencia entre scripts, hojas de estilo, y marcado para determinar la forma en que deben dibujarse los píxeles en la pantalla. La mayoría de estudios sobre la web etiquetan estas épocas como Web 1.0 y Web 2.0 respectivamente. La moda de etiquetar obliga a seguir poniendo adjetivos a la Web, se habla de la Web 3.0 como la web semántica en la que mediante metadatos los contenidos pueden tratarse, no solo sintácticamente, sino también por su semántica. Esto, unido a técnicas de inteligencia artificial permite procesar mejor la información y generar un mayor conocimiento. A pesar de seguir inmersos en un importante debate sobre el término Web 3.0, su significado y definición, se habla ya de la Web 4.0 en la que el término ubicuidad se utiliza tanto para usuarios como contenidos. No estará tampoco exenta de polémica esta nueva acuñación.



Una de las características más importantes de la web actual y, sin embargo, no recogida en las etiquetas, es su capacidad para generar negocio y cambiar los paradigmas de la economía. No estamos refiriéndonos únicamente a la amplia difusión del comercio electrónico, cuyas cifras de ventas se incrementan vertiginosamente en todo el mundo habiendo movido alrededor de 1500 billones de dólares en 2014, sino también al cambio de modelo en la prestación de servicios de cualquier tipo y a su impacto en la economía. Quizá deberíamos comenzar a acuñar el término Web \$.\$.



Lo que realmente ha sucedido es que La Web se ha convertido en una interfaz universal, integradora para cualquier tipo de aplicación y servicio a través de Internet. Lo que empezó reemplazando a la primitiva aplicación de transferencia de ficheros (ftp) ha seguido fagocitando otras aplicaciones, y hoy en día, cualquier servicio y aplicación de internet tiene su interfaz web, desde la consulta y edición de correo electrónico, hasta aplicaciones como eDonkey (eMule) o BitTorrent pueden ser controladas mediante una interfaz web. Incluso dispositivos como routers, cámaras de vídeo, impresoras, etc. disponen de su interfaz web que permite su configuración y control de forma remota. Lo que se bautizó como la "Internet de las Cosas" (IoT en inglés) no hay duda que evolucionará hacia la "Web de los Dispositivos" que, en realidad, ya acecha a la vuelta de la esquina.

El nuevo HTML5, publicado en octubre de 2014, ha traído multitud de innovaciones al diseño de las páginas web. Algunas son meramente estructurales, pero otras permiten la incorporación de contenidos mucho más ricos, como los elementos multimedia <audio> y <video> o los elementos gráficos <svg> y <canvas> que abren la puerta al diseño 3D. Dispone también de nuevas funcionalidades para arrastrar objetos por el documento e incorpora elementos que permiten gestionar la Web Semántica. Pero lo mejor son las nuevas APIS tales como la “Offline Web Applications” que permite construir aplicaciones Web que funcionen mientras se está desconectado. U otras como la Geolocalización del dispositivo, los *WebSockets* que permiten comunicación mediante flujos de bytes bidireccionales, o los *WebWorkers* para incorporar múltiples hilos de ejecución en paralelo. Y como guinda, el futuro *System Information API* (aún en estado latente y de momento archivado a la espera de tiempos mejores) que permitirá acceder al hardware a bajo nivel (red, ficheros, CPU, memoria, puertos USB, cámaras, micrófonos...) y que posiblemente permita la implementación de un verdadero Sistema Operativo Virtual Ubicuo en la Nube.



La estructura. Si analizamos la estructura de las páginas, comprobamos que el tamaño medio de estas ha pasado de unos pocos KBytes a cerca de 2 Mbytes en 20 años y lo mismo ha ocurrido con el número de objetos (archivos únicos) contenidos en cada página. Si a mediados de los años 90 lo normal era encontrar unos 3 objetos por página, actualmente se están alcanzando los 200. Todo ello ha sido una consecuencia de los mayores anchos de banda en las redes de comunicaciones que, de media y en países con alto nivel de desarrollo tecnológico, crece alrededor de un 20% cada año; si bien ello no significa que el tiempo de transmisión en la red haya mejorado en un ritmo similar. La paradoja es que, a pesar de estas mejoras en las redes, la descarga de una página web es cada año cerca de un 30% más lenta. A pesar de lo discutible de medidas y muestras cuando se trata de Internet y de la web, son varios los estudios que hablan actualmente de tiempos medios de descarga por página de entre 9 y 13 segundos (con desviaciones de alrededor de 6 segundos). Otros análisis alertan de que los usuarios abandonan un sitio tras más de 3 segundos de espera, mientras que Google indica que para los sitios de comercio electrónico el tiempo medio admisible está en solo 2 segundos.



No podemos obviar que la creciente complejidad de las páginas web en cuanto a apariencia, estructura y tamaño también afecta negativamente al tiempo de representación (rendering) de la página en la pantalla de nuestro dispositivo, a la vez que requiere más potencia de procesador y mayor cantidad de memoria. Esto no debería ser un problema en un escenario en el que los computadores personales son cada vez más potentes pero la realidad es que hoy en día los dispositivos móviles como tabletas y teléfonos han desplazado a estos cuando se trata de navegar por la web. Frente a este problema muchos sitios web han desarrollado versiones menos

pesadas para usuarios de móviles que, si bien es cierto que se descargan más rápidamente, también es cierto que tienen menos funcionalidad y ofrecen un entorno menos familiar a los usuarios por lo que estos acaban prefiriendo los sitios originales. La tendencia actual nos lleva hacia el llamado *Responsive Web Design* (RWD), cuyas técnicas y pautas de diseño permiten adaptar los sitios web a los distintos entornos y dispositivos de navegación del usuario ofreciéndole a éste una Experiencia de Usuario (UX) enriquecida.



El verdadero asesino es la latencia. La paráfrasis de M. Satyanarayanan en el congreso ACM Mobicom '96 es todavía cierta (*"The entire focus of the industry is on bandwidth, but the true killer is latency."*). Sin embargo, la industria sigue enfocada en el ancho de banda, y eso ha sido precisamente lo que los Proveedores de Servicio de Internet (ISP *Internet Service Providers*) han usado en todas sus campañas publicitarias para hacerse con un trozo del pastel propagando la falacia de que cuanto mayor ancho de banda adquieran los usuarios mejor experiencia tendrán en el acceso a los contenidos.

A pesar de haber sido muchas las voces de científicos que han incidido sobre el gran peso de los tiempos de transmisión o respuesta (RTT) en el tiempo total de descarga de las páginas web, no se han producido cambios significativos en el protocolo hasta hace unos pocos años. Mike Belshe ha afirmado que *"si pudiéramos reducir el tiempo de respuesta (RTT) transatlántico, de 150 ms a 100 ms, esto tendría un efecto mayor sobre la velocidad de Internet que aumentar el ancho de banda de un usuario desde 3.9 Mbps a 10 Mbps o incluso a 1 Gbps."* Mike Belshe es co-inventor del protocolo SPDY y uno de los principales autores de la especificación HTTP/2.0. Ambas son las primeras propuestas, tras más de 20 años de uso del actualmente ineficiente protocolo HTTP, encaminadas a reducir los tiempos de respuesta mejorando, entre otras cosas, la gestión de las conexiones TCP. La versión 2.0 del protocolo HTTP ha generado muchas expectativas pues además de permitir peticiones concurrentes en una única conexión TCP, comprime cabeceras y elimina de datos de control innecesarios a fin de reducir el tráfico, además de mejorar la seguridad de las comunicaciones y permitir un rol activo de los servidores al otorgarles capacidad para enviar datos a sus clientes sin que estos los hayan solicitado previamente. Esta última mejora será de gran utilidad a la hora de implementar más eficientemente técnicas como el caching o la prebúsqueda de contenidos.



La arquitectura. Efectivamente, hemos tenido que recurrir a técnicas aprendidas y utilizadas eficientemente en la arquitectura de los computadores como el caching, la prebúsqueda o el procesamiento distribuido para poder contribuir, de una manera cercana, a mejorar las prestaciones de la web. El paradigma original cliente-servidor ha ido evolucionando constantemente incorporando nuevos elementos tales como *Proxies* y *Surrogates* que, mediante la explotación de técnicas de caching, han permitido, por una parte ahorrar tiempo

~ 4 ~

de latencia y ancho de banda acercando los contenidos a los usuarios finales, y por otra, ahorrar tiempo de cómputo o de consulta a bases de datos. Tal ha sido la importancia de estos métodos que han dado lugar a toda una industria para explotarlos, las Redes de Distribución de Contenidos (CDNs).

Las CDNs permiten reducir la latencia final mediante la réplica de los contenidos en servidores próximos a los usuarios finales a la vez que proporcionan una solución eficiente ante una creciente demanda de servicio ya que permiten repartir la carga global de peticiones entre múltiples servidores. Esta técnica que parece tan sencilla conlleva la solución de múltiples problemas periféricos, tales como mantener la consistencia entre las réplicas cuando se actualizan los contenidos del sitio original, redireccionar eficiente y transparentemente las peticiones y distribuir las peticiones de manera equilibrada. Las CDN de nueva generación contemplarán aspectos como la movilidad de los usuarios, personalización de contenidos, transferencias autónomas, ubicuidad y sobre todo serán mucho más eficientes en su gestión. Lo que es seguro es que este mercado irá en aumento y sus técnicas se utilizarán mucho más para popularizar contenidos a través del marketing.

Asimismo, en los últimos años han comenzado a utilizarse con éxito técnicas adicionales para enriquecer aún más la experiencia del usuario acelerando el acceso a las páginas web. Mediante el análisis de los patrones de acceso de los usuarios a los contenidos, es posible predecir sus accesos futuros y pre-buscar éstos antes de que sean realmente solicitados. Mecanismos similares pueden utilizarse para prevalidar los objetos ya existentes en las cachés, haciendo innecesario el intercambio de mensajes con el servidor original para garantizar la coherencia. Estas técnicas de prevalidación, implementadas ya con éxito en algunos de los productos de Google, pueden llegar a reducir notablemente la latencia percibida por los usuarios, al mismo tiempo que descargan a los servidores al hacerles llegar menos peticiones y, todo ello, sin necesidad de un consumo adicional de ancho de banda.



¿Y en un futuro? El futuro ya está aquí. La popularización de los dispositivos móviles y *wearables*, junto con la realidad aumentada dirán adiós al clic que será sustituido por gestos. EL scroll de las páginas, el diseño sensible (*responsive*) y el uso de vectores de imágenes permitirán seguir llenando las páginas de hermosas vista que escalarán rápidamente y sin perder calidad en distintos dispositivos. A su vez, las web simplificarán los diseños y los diseñadores se implicarán

cada vez más con la infraestructura para ayudar a que las descargas sean más rápidas, porque los usuarios, cada vez más experimentados, van a ser, también, cada vez más impacientes. El papel de la infraestructura va a ser decisivo. Réplicas de contenidos, técnicas de caching y prebúsqueda de la información se explotarán conjuntamente aprovechando las ventajas de los nuevos protocolos y estándares y, en directa connivencia con los diseñadores, cada vez más profesionales e implicados con la satisfacción del usuario final. Un futuro que pasa, inevitablemente, por desarmar al verdadero asesino: **la latencia**.

REFERENCIAS

- **Alex Pinto.** Web Page Sizes: A (Not So) Brief History of Page Size through 2015. YOTTA website. August, 2015 <http://www.yottaa.com/blog/application-optimization/a-brief-history-of-web-page-size>
- **CodexM.** Average Page Load Time of Top Ranking Websites in Google. March, 2010 <https://www.seochat.com/c/a/tag/clear-browsing-data/>
- **Ilya Grigorik.** High Performance Browser Networking. What every web developer should know about networking and web performance. O'Reilly Media, 2013
- **Ethan Marcotte.** Responsive web design. 2013. <http://www.theusers.info/dev/mobile/RESPONSIVE-WEB-DESIGN-Ethan-Marcotte.pdf>
- **Mike Belshe.** More Bandwidth Doesn't Matter (much). August, 2010. <https://docs.google.com/a/chromium.org/viewer?a=v&pid=sites&srcid=Y2hyb21pdW0ub3JnfGRldnxneDoxMzcyOWI1N2I4YzI3NzE2>
- **Brian D. Davison.** Web Caching and Content Delivery Resources. Ultima revisión 2012 <http://www.web-caching.com/>
- Internet World Stats. Usage and Statistics. <http://www.internetworldstats.com/>
- The Chromium Projects. <http://www.chromium.org/spdy/>
- IETF Datatracker. <https://datatracker.ietf.org/wg/httpbis/charter/>
- Hypertext Transfer Protocol Version 2 (HTTP/2) <https://www.rfc-editor.org/info/rfc7540>