Creado el 25/01/2014
Última versión: 25/01/2014
Arturo Montejo Ráez, Universidad de Jaén
Comunidad moocTLH
Esta obra está bajo una licencia de Creative Commons Reconocimiento-CompartirIgual 3.0 Unported.

Retos y oportunidades de las TLH/PANORAMA ACTUAL

Web 1, 2 y 3.0

~~A todos nos suena el término "fidelización" o alguna vez hemos buscado la opinión de otros a la hora de comprar un producto, sea por Internet, sea en una tienda física. Veamos las líneas generales de cómo se construye un sistema de este tipo aprovechando las TLH.~~

Transcripción del vídeo

Hola, soy Arturo Montejo Ráez, del grupo de Sistemas Inteligentes de Acceso a la Información de la Universidad de Jaén, y vengo a resumiros los tres principales estados evolutivos de la World Wide Web. Empecemos donde comenzó todo esto.

El Laboratorio Europeo de Física de Partículas (CERN), es una institución que emplea a más de 2,400 personas, con unos 10,000 científicos e ingenieros visitantes desde cientos de países. Este enorme laboratorio investiga en sus instalaciones la naturaleza última de la materia que compone el universo. En este lugar tan activo y concurrido se desarrolló el proyecto ENQUIRE, por Tim Berners-Lee y Robert Cailliau hacia 1990.

ENQUIRE pretendía facilitar el intercambio de información entre los investigadores desarrollando el concepto de "hipertexto". Un hipertexto es un documento de texto donde algunos fragmentos son "hipervínculos", es decir, punteros que llevan a otros documentos. En 1992, el CERN publicó un informe, firmado por Berners-Lee y Cailliau, donde se describía la arquitectura de lo que denominaron la "World Wide Web", liberando esta tecnología al mundo gratuitamente. El software fue difundido a través de grupos de noticias y otros foros, extendiéndose por laboratorios de física de todo el mundo. De esta forma comenzó a ser posible "navegar" por este entramado de páginas enlazadas. Es el comienzo de la Web 1.0.

Su arquitectura es cliente/servidor, donde un software servidor proporciona documentos ante las peticiones de los programas clientes. La WWW se sustenta sobre tres estándares principales: uno es el formato en el que se escriben las "páginas web" (hipertextos), denominado HTML (Hyper-Text Markup Language), otro es la nomenclatura según la cual podemos identificar los documentos, denominada URL (Unified Resource Locator), y el tercero es el protocolo utilizado para que el programa cliente se comunique con el servidor, denominado HTTP (Hyper-Text Transfer Protocol). De esta manera, un "navegador web" puede buscar un documento HTML en la web gracias a la URL del mismo.

Como podemos imaginar, esto es un escenario donde la web se compondría de páginas estáticas de hipertexto. Dado que la arquitectura no indica la forma en que se generan esas páginas ¿podríamos tener en el lado del servidor "programas" que construyeran el HTML de las páginas según ciertas variables de entrada, por ejemplo campos en la URL, e información alojada en una base de datos? Sin duda. Esto es lo que se conoce como "web dinámica" y marcó el inicio de tecnologías CGI y los lenguajes web, como ASP o PHP, entre otros. Gracias a esto pudieron aparecer los primeros "servicios en la web", como el buscador AltaVista, en diciembre de 1995, o servicio de correo electrónico Hotmail, en julio de 1996.

La Web 1.0 se popularizó y rápidamente dio el salto de las universidades y los centros de investigación al usuario doméstico. Gracias a los primeros servicios de alojamiento web gratuitos, como Geocities a finales de 1994, muchos usuarios pudieron comenzar a publicar información en Internet usando la tecnología web. Por aquel entones era necesario en muchos casos crear las páginas localmente y luego subirlas al servidor para su publicación. Existían editores como FrontPage (1996) o Dreamweaver (1997) que facilitaban la tarea de crear páginas web sin necesidad de grandes conocimientos en HTML.

La creación de páginas web personales sería el germen de los “web-logs”, posteriormente denominados “blogs”. Algunos de los primeros blogs comenzaron su andadura allá por 1997, pero no sería hasta la llegada de los sistemas de gestión de contenidos (CMS) orientados a blogs que no comenzaría su explosión y popularidad. Un CMS es un software servidor que permite la gestión de los contenidos desde la propia interfaz web. De esta manera no es necesario un editor local para crear el contenido y luego subirlo al servidor, sino que la herramienta facilita una interfaz web para gestionar el contenido (el back-end) que luego será publicado y mostrado al mundo (a través de su front-end).

Uno de los servicios más populares para gestión de blogs es Blogger, lanzado en agosto de 1999. La sencillez con la que un usuario puede crear y mantener su blog era tal que la "blogosfera" creció exponencialmente, potenciando tecnologías derivadas como el RSS, para recibir actualizaciones, o los podcasts, para publicar audio, así como la participación mediante comentarios. Este fue el comienzo de la Web 2.0.

El diseño de la Web 1.0, incluso la dinámica, no era suficiente para poder desarrollar todo el potencial que la web podría ofrecer al usuario de a pie como productor y consumidor de información. Para ello era necesario "enriquecer" las capacidades que los navegadores web ofrecían como interfaz de programa. Esto se consiguió proporcionando dinamismo al navegador, por lo que surgió la necesidad de lenguajes de programación también del lado cliente. La solución fue Javascript.

Javascript fue primeramente usado por Netscape Navigator 2.0, en 1995, y permitía manipular el contenido de una página web desde el navegador, es decir, acceder al modelo de objetos del documento (DOM) inherente a HTML. La existencia de "eventos" asociados a funciones (al hacer clic en un botón o pasar sobre un enlace, por ejemplo) y la posibilidad de cambiar dinámicamente después el contenido de la página web abrió grandes posibilidades para el desarrollo de una experiencia de usuario más interactiva.

Al conseguirse experiencias de usuario más avanzadas fue posible desarrollar aplicaciones web más intuitivas y con mayores funcionalidades, convirtiendo al navegador en la "interfaz universal". Publicar en Internet y comunicarse a través de la web era ahora muy sencillo. Aparecen así servicios del tipo "2.0", como Wikipedia o YouTube.

Pero ¿que es la web 2.0? Pues es aquella que se convierte en un espacio para el desarrollo de comunidades y sistemas de colaboración entre usuarios. Es la "web social" donde todos estamos y donde comunicarse con otros es fácil y rápido. Este término, creado por Darcy DiNucci en 1999, engloba a esos servicios y tecnologías orientados a promover la web como un medio de diálogo social, donde el contenido es libremente creado por sus usuarios. Algunos ejemplos de los servicios y nuevas tecnologías que forman parte de esta Web 2.0 son las folksonomías, las wikis, los blogs, las redes sociales, los mashups y un largo etcétera. Esto ha provocado el surgir de grandes compañías, como Facebook, y es la semilla de tecnologías como la computación en nube, que ofrecen servicios intuitivos y transparentes incluso más allá de la propia web.

Nos queda para finalizar definir la Web 3.0, un concepto bastante ambiguo. A pesar de las discrepancias, en general parece haber un consenso que puede resumirse como el escenario donde nueva información es producida por máquinas, no sólo por humanos. Esto implica el uso de “agentes autónomos”, es decir, programas que son capaces de recabar información, procesarla y generar nueva información a partir de ésta. Visto de esta forma, la denominada Web Semántica se erige como el conjunto de tecnologías más prometedoras para alcanzar dicho objetivo.

También ideada por Tim Berners-Lee, la web semántica va más allá de páginas conectadas por hiperenlaces y trabaja por una web donde los nodos son datos (recursos) y los enlaces tienen un significado (relaciones). De esta forma, todo dato viene identificado por su URI (Uniform Resource Identifier), un concepto similar a la URL pero no orientado a páginas web. Los datos se relacionan unos con otros, creando una gran red de datos enlazados. Se pretende de esta forma convertir a la web en una base de datos universal mediante el establecimiento de “ternas”: sujeto – predicado – objeto. Sobre dicha red de datos es donde actuarían los agentes, resolviendo necesidades de información mediante motores de inferencia y razonamiento, más allá de la mera recuperación de páginas que hacemos ahora.

Tal vez el proyecto más avanzado en este sentido sea Linking Open Data, el cual está volcando un número creciente de bases de datos siguiendo los formatos de la web semántica y creando una estructura interconectada global con DBpedia, la versión de Wikipedia para la web semántica, como núcleo central.

Al igual que las bases de datos disponen de lenguajes de consulta, los datos enlazados pueden ser consultados usando un lenguaje: SPARQL. Podemos responder a preguntas cuyas respuestas no existen de forma explícita en la web, sino que son inferidas desde ésta. DBpedia ofrece algunos ejemplos de consultas en SPARQL como “¿cuáles son las parejas de los hijos de Margaret Thatcher?”, “¿se llama Michelle la esposa de Obama?” o “¿cuáles son los grupos de música heavy de los años 80?”. Otro ejemplo que puede resultar más informativo es el de “encuentra todos los países sin salida al mar con una población superior a 15 millones”, lo cual podría resultar interesante en escenarios como el militar o el comercial.

Las Tecnologías del Lenguaje Humano también juegan un papel importante en este paradigma, pues tareas como el “slot filling” posibilitan la identificación de datos estructurados (ternas de relaciones) a partir de documentos no estructurados. Hay una ingente cantidad de información donde las TLH pueden extraer datos y relaciones y contribuir en la construcción de una red de datos enlazados mayor.

La Web 3.0 otorga mayor estructura a la información que existe en la web, y el procesamiento de esos datos serviría para generar nueva información, nuevo conocimiento, a partir de un modelo global. Millones de agentes se alimentarán y poblarán esta maraña de datos, colaborando entre ellos para producir información de mayor nivel. Esto, sumado a la “Internet de las Cosas”, donde multitud de aparatos y sensores vuelcan sus datos en la web de forma autónoma, supondrá la exploración de nuevos servicios que aún no somos capaces de prever.

Tal vez sea el último paso hacia la “singularidad tecnológica” y, a partir de ahí, la civilización humana será transcendida por una inteligencia superior.

Te animo ahora a realizar los ejercicios y afianzar así los contenidos de esta lección.