

Prensa en XML: el negocio de la personalización y la conquista de nuevos medios

David Rodríguez Mateos, Vicente Luque Centeno, M^a Carmen Fernández Panadero, Carlos Delgado Kloos, Antonio Rodríguez de las Heras, Andrés Marín López, Carlos García Rubio, Luis Sánchez Fernández, J. Tomás Nogales Flores, Antonio Hernández Pérez

*Universidad Carlos III de Madrid
per@it.uc3m.es*

Resumen

Los periódicos electrónicos no han encontrado aún la forma de aprovechar completamente las posibilidades que les ofrecen los lenguajes de marcado, y en especial XML, así como la posibilidad de usar diversos medios de salida, con el fin de difundir a través de ellos una misma información (adaptando la profundidad de la misma a las posibilidades de cada soporte) consiguiendo que el producto sea rentable. El proyecto InfoMedia, que lleva a cabo un grupo de investigadores de la Universidad Carlos III de Madrid, se desarrolla sobre esta idea. Este texto expone los conceptos básicos (y algunas iniciativas ya desarrolladas) sobre los que gira el proyecto: la necesidad de contemplar los diversos medios (ordenadores, televisión, teléfono móvil...) accesibles mediante un mismo lenguaje, basado en XML; la posibilidad de establecer diferentes tipos de productos y, por tanto, diferentes formas de financiación y pago de los mismos; finalmente, la necesidad de determinar qué públicos distintos pueden existir y qué información necesitan.

Introducción

La práctica totalidad de los periódicos de Estados Unidos están perdiendo millones de dólares con sus ediciones en Internet. En España, ninguna empresa de medios de comunicación parece estar ganando dinero con la apuesta del periódico electrónico, a pesar de que El País es el tercer sitio web más visitado por los navegantes españoles y que entre los 50 sites más visitados de nuestro país figuran, según una encuesta a finales de 1999, 13 medios de comunicación: 9 periódicos, 1 cadena de televisión, 2 emisoras de radio y un grupo editorial que cuenta con diversos periódicos y revistas. A pesar de ello, todas las empresas de comunicación están convencidas de que una parte importante del futuro negocio estará en la red y todas contemplan seguir invirtiendo en estos medios. Por poner un ejemplo, todos los grandes portales incluyen, entre sus contenidos, noticias servidas obtenidas por medios de información, agencias de prensa sobre todo.

Las razones que explican que la mayoría de los periódicos pierdan dinero en sus versiones electrónicas son múltiples y variadas. La gran competencia entre medios que eran antes complementarios y ahora son competidores en Internet (periódicos y agencias de prensa ofrecen ya cierta información en el acto), la falta de penetración en España del PC en los hogares (un 38% frente al 97,5% de televisores), la carencia de destrezas informáticas (sólo el 11% de la población española accede habitualmente a Internet), la escasa interactividad de los periódicos, la dificultad en la navegación y, como no, el exceso de información son algunas de las causas que explican la falta de rentabilidad de los periódicos electrónicos. Mientras que la cantidad de noticias que recibe una persona a través de los medios no ha hecho más que crecer, el tiempo disponible para conocer estas noticias se pierde en muchos casos tratando de decidir qué noticias se van a recibir.

Se presentan a continuación algunas de las pautas que sigue el proyecto "InfoMedia: Un sistema multiplataforma para la publicación de información", financiado con cargo al Programa Nacional de Aplicaciones y Servicios Telemáticos, en el que se está desarrollando un sistema de publicación de información periodística con tecnología Internet, por parte de un equipo de ingenieros, periodistas y

documentalistas de la Universidad Carlos III de Madrid. El proyecto se basa en otro ya terminado, también financiado por el mismo programa, y titulado "El Periotrónico: Concepción y Desarrollo de un Periódico Electrónico Personalizado". En este primer proyecto, se desarrolló un prototipo de periódico electrónico personalizado basado en tecnología XML. *InfoMedia* continúa los trabajos realizados, extendiéndolos a lo largo de tres líneas de innovación:

1. Nuevas plataformas de presentación: (televisión digital, teléfonos móviles, ...). Aunque el parque de ordenadores crece continuamente, el ordenador no está tan extendido en los hogares como lo está el televisor. La aparición prevista del cable en España plantea la posibilidad de publicar contenidos a través del televisor gracias al cable y los "set-top box". Por otro lado, los teléfonos móviles no sólo superan en número recientemente a los fijos, sino que permiten un nuevo tipo de servicios derivados de las nuevas tecnologías (WAP, UMTS, ...) que les pone en el punto de mira como uno de los dispositivos más populares para el acceso a Internet en un futuro inmediato. En este proyecto *InfoMedia* se investiga en nuevas formas de presentar información a través de estas nuevas plataformas de acceso a Internet.
2. Rentabilización de la publicación electrónica. Hemos mencionado que casi todos los periódicos electrónicos están perdiendo dinero. La publicación electrónica no puede mantenerse a la larga sin un modelo de negocio apropiado, y éste pasa por introducir técnicas de comercio electrónico de una forma u otra. En la segunda de las líneas de este proyecto investigamos la financiación de la publicación electrónica.
3. Agentes de apoyo al lector. La aparición de Internet no ha hecho más que acrecentar el problema del exceso de información. Desde 1997 hasta ahora la práctica totalidad de los denominados portales de Internet y muchos medios de comunicación se han interesado por los proyectos de personalización de noticias con la finalidad de filtrarlas y presentar a su audiencia solo aquellas que más le puedan interesar. El objetivo, sin duda, es lograr ampliar la cuota de visitantes que reciben en sus servidores. En este proyecto se investigan agentes de personalización de información, junto con otros que ayudan al lector a hacer más agradable el acceso a la información.

A continuación, se realiza un breve repaso sobre la evolución actual de estos aspectos, así como su influencia y aplicación en el proyecto *InfoMedia*.

Nuevas plataformas

Hasta no hace mucho, la principal causa que limitaba la expansión de Internet y de la publicación electrónica era el hecho de que requería el uso de un equipo informático cuya presencia en los hogares era limitada, en comparación con otros aparatos domésticos, como por ejemplo, la televisión, la radio, el teléfono móvil, etc.

En la actualidad, y aunque el ordenador sigue siendo la principal fuente de acceso a Internet, están surgiendo otros mecanismos de acceso que cuentan con la ventaja de ofrecer una mayor penetración en los hábitos de consumo españoles: la televisión y el teléfono móvil.

Con la irrupción del ordenador personal con capacidades multimedia en el espacio doméstico algunos se precipitaron a augurar la decadencia del televisor, un "viejo" aparato electrónico ya, pues ha superado el medio siglo desde las primeras emisiones para los hogares norteamericanos en 1946. Tal decadencia, por contra, no se ha producido. Si durante años, la integración del televisor se produjo lentamente hasta conseguir ser el "electrodoméstico rey" (un 97% de penetración en hogares españoles, por ejemplo), otro tanto ocurre con el ordenador personal. De hecho, hoy en día se piensa que no es realista pretender, al menos a corto plazo, que la implantación de los ordenadores en los hogares se aproxime a la del televisor. Una de las posibilidades del uso masivo de Internet está, pues, en conseguir que la televisión sea uno de sus principales medios de acceso.

En los últimos tiempos es posible observar una intensa actividad alrededor de la integración de Internet y televisión. De hecho, en España ya hay emisoras de televisión digital, como Quiero TV, que ofrecen televisión y acceso a Internet a través de un televisor más o menos convencional, denominado "sep-top box". Este tipo de iniciativas actualmente existentes persiguen varios objetivos. Por un lado, la posibilidad de que los sistemas de televisión digital proporcionen acceso a Internet utilizando el televisor abre paso a un conjunto de potenciales usuarios a los que anteriormente no se podía acceder. Por otra parte, la tecnología Web ofrece nuevas formas de presentación de contenidos en televisión. En este punto, es de destacar que el mecanismo de enlaces que se utiliza en las páginas Web es una forma atractiva de producir programas de "televisión interactiva".

Otra vía de penetración, ya puesta en práctica, es el teléfono móvil. La puesta en marcha de teléfonos móviles con tecnología WAP, en la actualidad y, a medio plazo, mediante el sistema UMTS, ofrece acceso a todo tipo de información útil, incluidas noticias, mediante este dispositivo. Esta forma de visualización no precisa, por el momento, de grandes contenidos gráficos, hasta el punto de recordar a la navegación inicial de Internet a través de herramientas como *gopher*. En la actualidad, por ejemplo, el Grupo Telefónica dispone de empresas que ofrecen contenidos periodísticos propios, (provenientes de la cadena de televisión Antena 3 TV), y de empresas de telefonía móvil (Movistar). La colaboración de ambas permite ofrecer noticias en televisión, enviar los titulares de las mismas a través del teléfono móvil y, al tiempo, difundir parte de esas informaciones a través de una web propia.

Los diferentes medios de penetración de Internet (la televisión, la telefonía móvil, el propio ordenador) son radicalmente diferentes en cuanto a los modos de visualización, y aún es posible encontrar nuevos medios de difusión en un futuro. Para una empresa informativa, sería más rentable poder producir sus contenidos de una sola vez, en un solo formato, del que luego pudieran extraerse los contenidos necesarios para su difusión por cada medio (más breve y sin imágenes, en un teléfono móvil; con mezcla de imágenes de alta resolución, audio y texto, en un televisor, etc.). Esa posibilidad requiere el uso de un marcado de texto común para determinar todas las partes (textuales, visuales o sonoras) de cada noticia, de entre las que puedan seleccionarse determinadas partes para su difusión a través de uno u otro medio. La posibilidad de etiquetar contenidos permite al autor aprovechar una de las características de Internet como medio de difusión, la amplitud del ciberespacio. Al no haber limitaciones en el espacio físico, el creador no tiene por qué dirigirse a una audiencia genérica, puede estructurar sus documentos y decidir qué tipo de información puede mostrar a cada usuario (distintos contenidos o distintos niveles de profundidad por ejemplo). Este tipo de marcado es factible en XML, que especifica la estructura lógica de los documentos, utilizando además XSL, que permite especificar el estilo de la presentación y transformar a otros formatos de visualización. Una de las principales ventajas de usar XSL frente a CSS es la posibilidad de especificar un paso de transformación antes de aplicar el formato.

Con el mismo documento XML y diferentes hojas de estilo podemos generar diferentes versiones del documento con estructura y propiedades de formato que se ajusten a los diferentes espacios de información (versión impresa, versión on-line en redes de banda ancha, versión on-line en redes con poco ancho de banda); en particular, podemos convertir documentos XML a HTML, SMIL y quizá en un futuro a MHEG para mostrarlos mediante una set-top box en un televisor. Esta nueva funcionalidad puede resultar muy útil a los proveedores de información no sólo por la ventaja de separar los contenidos de la presentación, sino también por la posibilidad de reutilizar información en diferentes escenarios; por ejemplo, a partir de unos mismos contenidos XML sobre un producto, se podría extraer la especificación técnica del mismo, un manual de usuario, información comercial, etc. Después del paso de transformación las reglas de estilo especifican el formato de los documentos.

Un posible lenguaje, basado en XML, que permita las capacidades arriba indicadas, es NewsML, cuya propuesta, aún en preparación, está siendo realizada por el IPTC (International Press Telecommunications Council), un conjunto de empresas y organismos periodísticos para el desarrollo de normas y aplicaciones de telecomunicación relacionadas con la prensa. NewsML presenta la ventaja

de ser una norma de consenso... lo cual también ralentiza su formulación definitiva. Asimismo, NewsML está basado en propuestas anteriores sobre marcado de prensa electrónica, como XMLNews. Pero, dado que pretende servir como una referencia para todo tipo de contenidos periodísticos, deberá tener en cuenta algunas otras normas relacionadas, como el citado SMIL.

El proyecto *InfoMedia* parte de la experiencia en el uso de un lenguaje de marcado periodístico, JML (Journalism Markup Language), generado mediante XML. El futuro, dependiendo del grado de desarrollo de NewsML y de las necesidades del proyecto, pasará por integrar JML en NewsML, si la evolución de este último lo convierte en un estándar más o menos aceptado.

Comercio electrónico: Rentabilización

El periódico electrónico, como cualquier otra empresa, necesita ser rentable para poder subsistir. La rentabilidad de un periódico electrónico no va a venir dada únicamente por su atractivo tecnológico sino, principalmente, por el modelo de negocio adoptado. Éste debe permitir capturar la atención del lector ofreciéndole una información y unos servicios únicos, de modo que al lector le compense pagar por ellos.

Un periódico electrónico dispone de dos formas clásicas de financiación, de forma similar a lo que sucede con su antecesor en papel:

- Inclusión y cobro de publicidad a marcas anunciantes.
- Cobro de la información al lector, de manera puntual o por suscripción.

Las primeras publicaciones electrónicas se basaban en un modelo estático, propio de otros medios no interactivos, como la televisión: ciertas marcas anunciadoras pagaban grandes cifras de publicidad por alcanzar un mercado masivo. Este tipo de publicidad no se adapta fácilmente a un entorno personalizado, como Internet, en el que el lector puede buscar su propia información. Como resultado de esto, la mayoría de las compañías de publicación electrónica no han dado aún beneficios. El futuro de las publicaciones electrónicas pasa, por tanto, por desarrollar nuevos modelos de negocio que permitan cobrar al lector directamente, de forma que éste considere el cobro justificado y continúe siendo un cliente y recuperar el interés anunciador de las marcas ofreciéndoles acceso a un mercado personalizado.

Parece evidente que los soportes de difusión elegidos determinarán el tipo de información y, por tanto, el tipo de cliente. Pero, sean cuales sean ambos, soporte y usuario, habrán de tenerse en cuenta algunos factores a la hora de decidir tanto el precio de la información producida como el cliente que estaría dispuesto a pagarla (anunciante o usuario). Estos factores son:

- **Método de financiación:** Por medio de publicidad, por suscripciones, por pago directo por unidad informativa (sea la noticia, o un superconjunto o subconjunto de ella).
- **Pricing:** Qué precio se pone a las unidades informativas. Esto se hará de acuerdo con diversas estrategias y factores.
- **Instrumento de pago:** Cómo paga el lector. Con tarjeta de débito, de crédito, o con dinero electrónico. De uso general o restringido.
- **Aspectos de seguridad:** Cómo se realizan las transacciones seguras.
- **Copyright:** Cómo se garantiza la propiedad intelectual de ciertos artículos.

- **Inserción en prácticas habituales:** En un periódico la información sigue un determinado flujo utilizando ciertas herramientas que usan diversos formatos. Deben integrarse esas prácticas con la forma de financiación, es decir, con la inserción de publicidad, si corresponde, o con la oferta de esos contenidos mediante pago.

En este último punto, debe destacarse que ambos flujos, el de generación de noticias y el del pago de las mismas, deben estar integrados para permitir que una noticia sea ofrecida al instante. Diversos productos de este tipo están ya en el mercado. Debe citarse, por ejemplo, Insiight, que permite combinar el manejo de contenidos tanto para la edición papel como para edición de noticias en línea, así como la manipulación de publicidad (incluidos los anuncios por palabras, una de las principales fuentes de financiación de algunos periódicos en papel) y el uso de herramientas para el comercio electrónico relacionado con los contenidos (tanto informativos como de publicidad). Por su parte, ContentWare integra todo tipo de aplicaciones conocidas para la elaboración de los diferentes contenidos, permitiendo luego su difusión a través de diferentes medios y lenguajes, incluido WAP. La idea central de ContentWare es el uso de metadatos que permiten identificar los diferentes conjuntos de textos, así como el soporte de numerosos formatos, incluido XML.

Personalización

Para poder conseguir y mantener clientes en una publicación electrónica, debe saberse de antemano quiénes son y qué tipo de noticias quieren recibir de entre todas las posibles. En otras palabras, se debe tratar de personalizar, en lo posible, la información que se le ofrece, como valor añadido frente a los medios tradicionales (radio, televisión, prensa en papel).

Esta personalización debe entenderse en varios sentidos: por un lado, ofrecer al usuario aquellos contenidos que prefiera, de entre la multitud de informaciones posibles, para tratar de ganar su interés y, al tiempo, permitirle enfrentarse al exceso de información; por otro lado, ofrecer al usuario diferentes grados de profundización en la noticia, dependiendo del soporte elegido, del tiempo disponible, etc.; finalmente, debe ofrecerse al usuario la información deseada en un formato de visualización que haga la información comprensible y atractiva.

Los proyectos de personalización de prensa empezaron a comienzos de la pasada década. Ya en 1990, se constituyó en el Media Lab, del MIT, el consorcio "News of the Future", uno de cuyos objetivos era la personalización de noticias electrónicas para diferentes usuarios. De esa iniciativa, aún en marcha, han surgido proyectos como Fishwrap, basado en una combinación de noticias tomadas de agencias, junto con una serie de noticias universitarias del propio MIT. Técnicamente, Fishwrap era un sistema hipertexto integrado en el que el servidor Fishwrap intenta personalizar la edición del lector a través de un sistema de perfiles de usuarios.

El prototipo de este estilo que más interés despertó en su momento fue presentado en la reunión del G-7 que tuvo lugar en Halifax, Canadá en 1995: el Electronic News Delivery Project. El contenido del "periódico" fruto de este proyecto sería una colección integrada de noticias multimedia procedentes de diferentes fuentes y personalizadas por el lector; el control del empaquetamiento y envío de la información lo tendría el lector y no el emisor; la selección, clasificación y priorización de la edición puede ser complementada por terceras partes: editores, anunciantes o lectores pueden añadir noticias que recibirá otro lector; el diseño del periódico dependerá de las facilidades de visualización del software y el hardware del lector; el envío de las noticias se realizará a través de una tercera parte (compañía telefónica o de cable); y los anuncios serán un componente integral y no un añadido al periódico. Los anuncios podrán ser multimedia, estar asociados a las estructuras lógicas de documentos o a los tipos de lectores que accedan a cierto tipo de información.

También es interesante la experiencia llevada a cabo por NewsEdge. Aquí, las noticias provenientes de diversas fuentes son filtradas por un programa especial que atiende a ciertos criterios. Las noticias filtradas son procesadas a continuación por un equipo de editores especializados. Por otra parte, cada

lector especifica en una tabla los temas que son de interés para él en un determinado orden. Una vez elegidos estos temas, el lector tendrá acceso todos los días a su periódico personalizado. En este sistema se realiza un gran esfuerzo en el filtrado de la información, lo cual es de importancia capital. Sin embargo, la presentación de la noticia es muy simple y la misma noticia puede encontrarse bajo epígrafes distintos.

En nuestro país existen algunos proyectos interesantes pero están aún muy lejos de convertirse en productos capaces de crear verdaderos periódicos electrónicos personalizados. Los grandes medios de comunicación españoles aun no los han implementado. De hecho, pueden compararse los dos modelos tradicionales de personalización de información usados en Internet y las respectivas estrategias de personalización utilizadas por los citados medios:

1. El modelo de difusión selectiva de información, de los distribuidores de bases de datos tradicionales, muy comunes en el ámbito científico. El usuario del servicio informa al proveedor de información de las palabras claves en las que está interesado y, con la periodicidad acordada entre las partes, el proveedor ofrece al usuario del servicio las novedades que se han producido en la base de datos. Este tipo de servicios funciona desde los años 70 en bases de datos comerciales como Dialog. Algunas empresas, como MyNews o Prensadigital, ofrecen servicios de personalización de noticias de este tipo; en algunos casos, como MyNews, en formato HTML. Estos servicios, sin embargo, funcionan con un grado de personalización relativamente elevado pero con escasa interactividad con el usuario y con mínimas posibilidades respecto a la presentación de las noticias.

2. El modelo "Webcast", en el que están basados los famosos canales. En esta opción, el usuario elige las fuentes de información de las que quiere recibir noticias y de cada fuente elige las secciones de las que desea recibir noticias. El usuario puede, además, elegir cuándo desea recibir las noticias: todos los días, cada semana, cuando lo solicite expresamente, etc. Y cómo desea recibirlas: como mensaje de correo electrónico, verlas en un salvapantallas, o verlas con un navegador. Algunos de los principales medios españoles han optado por este medio, ya sea mediante canales propiamente dichos, accesibles a través del Explorer de Microsoft, que funciona de acuerdo al CDF (Channel Definition Format); o, simplemente, comenzando a crear contenidos originales para los servidores web (que denominan en algunos casos "canales") añadiendo a las noticias enlaces hipertextos, utilidades de búsqueda y funciones de cierta interactividad y personalización (envío diario de titulares por correo electrónico, posibilidad de consultar información personalizada sobre acciones de bolsa... así como chats, foros y encuestas).

El problema fundamental de los modelos de personalización actualmente existentes es que se basan fundamentalmente en HTML, un lenguaje muy limitado en cuanto a capacidad de presentación y estructuración de textos. Pocos medios han comprendido aún la importancia de utilizar de forma integrada los distintos formatos estándar aprobados para la publicación electrónica y que Internet, como nuevo medio de difusión, trae consigo un profundo cambio estructural en la forma en que se crean, se publican y se leen los documentos.

Conclusiones

El proyecto *InfoMedia* está considerando en la actualidad para su desarrollo los siguientes factores fundamentales:

- La necesidad de contar con lenguajes de marcado que permitan utilizar un mismo producto informativo en diferentes fuentes, creando el producto de una sola vez, pero permitiendo que cada una de las partes (ya sean textos, imágenes, sonidos...) sean ofrecidas por los diferentes medios de que se disponga, sin realizar modificación alguna sobre el contenido. NewsML puede ser el lenguaje que permita controlar todos estos contenidos de una sola vez; pero, en todo caso, el lenguaje desarrollado debería estar basado en XML.

- La certeza de que el producto resultante habrá de ser rentable y, aún más, de que deberá definirse cuál es esa rentabilidad. En la actualidad, a los posibles usuarios se les ofrecen productos cerrados (por ejemplo, el acceso a una base de datos con informaciones concretas), cuando los medios electrónicos podrán ofrecer contenidos muy concretos, que habrán de comercializarse de diferente manera, por ejemplo, mediante micropagos. Los procesos de producción, difusión y comercialización deben de estar integrados, de modo que los datos e informaciones puedan estar disponibles en el menor plazo posible y puedan rentabilizarse desde ese momento.
- Para que esa rentabilidad sea posible, habrá de estimarse qué tipos de usuarios se pueden tener, qué información quieren y cómo la quieren recibir. Ello supone plantearse el modo de lectura de la prensa en Internet, si ese modo varía respecto a la lectura en papel, cómo afecta al texto escrito la presencia adicional de imágenes y sonidos... y cómo traducir este conjunto a herramientas que generen los contenidos deseados, mediante interfaces válidas.

Agradecimientos

El trabajo en el que se basa este documento es financiado por el proyecto TEL99-0207 de la CICYT.

Bibliografía

AIMC . Resultados de la 3ª Encuesta AIMC a usuarios de Internet (octubre-diciembre/99).

<http://www.aimc.es/aimc/html/inter/informe99.html>

Eva Domínguez. "Herramientas para unificar la gestión de contenidos, edición y clasificados de prensa e Internet". En: *La Vanguardia Digital*, 21/6/2000.

<http://www.vanguardia.es/cgi-bin/noticialvd.pl?noticia=nexpo210600&seccion=pdg>

Eva Domínguez. "El texto es el primer reclamo en la información on-line". En: *La Vanguardia Digital*, 8/5/2000, <http://www.vanguardia.es/cgi-bin/noticialvd.pl?noticia=eyetrack&seccion=pdg>

IPTC. NewsML. <http://www.iptc.org/NMLIntro.htm>

MHEG Centre. <http://www.mhegcentre.com/>

Massachusetts Institute of Technology (MIT). News of the Future.

<http://nif.www.media.mit.edu/index.html>

NewsEdge. <http://www.newsedge.com/>

Openpages. ContentWare. <http://www.openpages.com/products/contentware.asp>

Systems Integrators Inc. Insight. <http://www.sii.com/products/productframe.htm>

Quiero TV. <http://www.quierotv.es>

Vicente Luque, María del Carmen Fernández, Carlos Delgado, Andrés Marín, Carlos García. "Concepción y desarrollo de un periódico electrónico personalizado", Jornadas Técnicas de RedIRIS 1998, Barcelona, Noviembre de 1998.

Watters, C.R; Shepherd, M.A; Burkowski, F.J. Electronic News Delivery Project. En: *Journal of the American Society for Information Science* 49, 2 1998 134-150.

World Wide Web Consortium. Extensible Markup Language. <http://www.w3.org/XML/>

World Wide Web Consortium. Synchronized Multimedia. <http://www.w3.org/AudioVideo/>

World Wide Web Consortium. Television and the Web. <http://www.w3.org/TV/>