

Periotrónico: XML en un periódico electrónico a la carta. Resultados y perspectivas.

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

Universidad Carlos III de Madrid
Avda de la Universidad, 30 - 28911 Leganés (Madrid)
<http://www.it.uc3m.es/~per>
per@it.uc3m.es

RESUMEN

El *Periotrónico* es un periódico electrónico a medida. Se fundamenta por una parte en el empleo de JML, un lenguaje basado en XML, para la descripción de noticias, y por otra en un modelo de personalización de contenidos basado en el plegado (sólo se muestran a priori aquellas noticias cuya temática es de interés expreso para el usuario). Parte de la tecnología de control de personalización se realiza en el cliente lo que reduce la sobrecarga tanto del servidor utilizado como de las redes de transmisión necesarias.

A partir de los resultados obtenidos, se pretende, en una segunda fase, el desarrollo de cuatro objetivos: utilizar un lenguaje estándar de marcas periodísticas, basado en XML, para la generación de noticias; fomentar la difusión inmediata de unos mismos contenidos periodísticos en diferentes soportes (PC, TV, etc.), mediante el uso de tecnologías como MHEG; favorecer la creación de comunidades virtuales, basándose en las preferencias aportadas por los clientes, que permitan el envío de contenidos periodísticos y publicitarios comunes; por último, mejorar la relación entre los diferentes contenidos informativos y las posibilidades de recuperación de información incluida en éstos. Tales factores configurarían este producto como un modelo rentable apto para el comercio electrónico.

Palabras clave: personalización, XML, tecnología cliente-servidor, periodismo electrónico, comercio electrónico

1. INTRODUCCIÓN

No es ninguna novedad la explosión de información disponible a través de Internet: estudios de febrero de 1999 [1] estiman la existencia de unos 800 millones de páginas Web accesibles. Entre este cúmulo de contenidos, son numerosos los periódicos que también han creado sus ediciones total o parcialmente accesibles a través de Web.

Buena parte de estas publicaciones, no obstante, coinciden en ser una versión equivalente de los contenidos impresos, tratando de simular la organización de éstos tanto conceptualmente, mediante secciones, como visualmente, emulando las columnas de un periódico.

Esta concepción no tiene en cuenta los diferentes tipos de lectores que acuden a una publicación o, al menos, no aprovecha la posibilidad de utilizar el soporte electrónico como método para distinguir los contenidos de una publicación y proporcionar a los diferentes lectores sólo aquello que quieren leer.

Con el proyecto *Periotrónico* [2] se ha pretendido, en una primera fase, concebir un periódico electrónico personalizado. Ello quiere decir que, sin perder de vista todas las posibilidades de obtener información por parte de un lector de periódico tradicional, se le ofreciera en primer lugar aquellos contenidos por los que estuviera más interesado.

Este objetivo, y todo el desarrollo que incluía, exigía la presencia de un equipo multidisciplinar en el que pronto se integraron ingenieros de telecomunicación, informáticos, periodistas y documentalistas, todos ellos pertenecientes a la Universidad Carlos III de Madrid: distintos enfoques sobre el tratamiento de la información que aprovecharon las sinergias propias de una misma institución investigadora.

A continuación, se describen las diversas tecnologías utilizadas en el desarrollo del *Periotrónico* (fundamentalmente, robots, XML y agentes), los elementos del prototipo final obtenido (un periódico electrónico con posibilidades de personalización, noticias ofrecidas mediante plegado y reparto de tareas técnicas entre cliente y servidor) y, a modo de conclusión, futuras líneas de investigación y desarrollo sobre el producto obtenido (una nueva versión con un lenguaje de marcado estándar de textos periodísticos, la creación de comunidades virtuales basadas en los perfiles de los usuarios, la difusión de los contenidos periodísticos en múltiples plataformas y una mayor relación entre los contenidos periodísticos) que permitieran hacerlo rentable.

2. DESARROLLO DEL PROYECTO

Internet y la tecnología Web ofrecen a las publicaciones electrónicas características especiales que las distinguen de las ediciones impresas. Por un lado, los mecanismos de hipertexto permiten un acceso directo a la información en contraposición a la lectura secuencial que suele seguirse en una publicación impresa. Además, las publicaciones electrónicas son instantáneas, interactivas, carecen de limitaciones del espacio físico en papel, son actualizables continuamente, permiten la inclusión de elementos multimedia (que pueden ir desde las imágenes de una cierta calidad hasta el audio o el vídeo) y son programables (y por tanto personalizables). Además de estas ventajas, las publicaciones electrónicas tienen menores costes de producción y de difusión.

En cuanto al proceso de creación, se ha producido un enriquecimiento de los contenidos. Al principio, la tecnología Web era utilizada para transmitir principalmente texto. Pronto surgió la necesidad de hacer más atractiva la transmisión de la información en el Web. Esto trajo consigo la aparición de gran cantidad de formatos multimedia: audio (.au, .mid, .rm, etc.), vídeo (.mov, .avi, .rm, etc.), etc. La inclusión de elementos multimedia y la posibilidad de actualización permanente han dotado a los documentos de la dimensión temporal.

Por otra parte, las posibilidades de plegado de información que ofrece el hipertexto rompen con el tradicional acceso secuencial a la información y los lenguajes de programación y de *scripting* permiten al usuario interactuar con el documento sin necesidad de conectarse con el servidor. Todas estas características del medio de difusión han hecho evolucionar el concepto de documento.

En el caso particular de los periódicos, es necesario encontrar un lenguaje que permita describir las noticias de forma adecuada distinguiendo los diferentes tipos de informaciones y los elementos de que constan. La posibilidad de etiquetar contenidos permite al autor aprovechar otra de las características de Internet como medio de difusión: la amplitud del hiperespacio. 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 muestra a cada usuario (distintos contenidos o distintos niveles de profundidad, por ejemplo).

2.1. XML

Uno de los motivos que evitaban la oferta de personalización era la dificultad en describir físicamente los contenidos periodísticos, es decir, en delimitar la estructura interna de los mismos: la rígida y corta lista de etiquetas de HTML, el lenguaje utilizado hasta ahora para la creación de hipertextos accesibles a través de Web, no se adaptaba a la variedad de partes en que podía dividirse cualquier noticia.

Hacía falta, pues, un lenguaje de marcado de textos periodísticos que, al tiempo, fuera accesible mediante la navegación por Web. El elemento clave de este lenguaje sería, a su vez, lo que podríamos llamar un metalenguaje: XML.

XML (*Extensible Markup Language*) [3] es un reciente estándar de marcado de datos estructurados para el *World Wide Web*, diseñado por el organismo regulador de tecnología del Web (W3 Consortium) [4], que proporciona la habilidad de organizar información compleja en una representación fácilmente manipulable por programas informáticos. XML es una simplificación del estándar SGML [5], formato ampliamente usado en los últimos años para la edición electrónica mediante marcado de texto. XML permite solucionar en el Web muchos problemas que HTML, también basado en SGML, no es capaz por sí mismo de solventar. Con XML se permite una mayor estructuración de los documentos Web, donde lo importante no es la presentación sino la correcta estructuración de los datos que forman parte del documento.

XML, con los nuevos estándares y extensiones que forman parte de su definición, permite una mayor flexibilidad en el mantenimiento de enlaces entre documentos, mediante XLink [6] y XPointer [7]; una mayor capacidad de transformación entre distintos formatos y de estilo, mediante XSL [8]; y una mayor facilidad de catalogación de documentos por sus metadatos, mediante RDF [9].

Asimismo, sin modificar la estructura lógica de un documento XML, tanto los navegadores como los usuarios pueden modificar la apariencia física del documento. Éste doble objetivo se consigue separando la definición de la estructura lógica (esto es, de sus posibles partes) del documento, en un fichero aparte, denominado DTD (*Document Type Definition*), y las posibles opciones de salida del documento, en otro u otros ficheros, denominados *hojas de estilo*.

El documento final solo deberá hacer mención, al comienzo, de la DTD en la que está basado y de las hojas de estilo con las que puede ser visualizado. El navegador tomará todos los datos y ofrecerá el resultado al usuario, permitiendo incluso que éste elija entre varias opciones de presentación.

La elección de XML como lenguaje para los futuros documentos del Periódico permitió la elaboración de un lenguaje (esto es, su DTD) basado en XML, para la descripción de estructuras de textos periodísticos electrónicos, ya que no existía ninguna iniciativa al respecto. Así surgió JML (*Journalism Mark-Up Language*), creado expresamente por el equipo del *Periódico*.

2.1.1. JML

Como hemos indicado, la gramática de los documentos

XML se especifica en ficheros llamados DTD (*Document Type Definition*). La figura 1 ofrece un fragmento reducido de la gramática DTD del lenguaje JML. En ella puede verse la manera de definir un documento JML como la secuencia de una cabecera (*head*) y un cuerpo (*body*). La cabecera almacena, a su vez, los metadatos de la noticia referentes al título, nombre del periodista o redactor, fecha de publicación, sección en la que aparece la noticia, periódico donde ha sido publicada y una lista de palabras clave separadas por comas. Todos ellos almacenan datos en formato textual. Más adelante, se define el contenido del cuerpo del documento y la etiqueta <p> para los párrafos. Esta última etiqueta puede contener un atributo que indica la importancia (*relevance*) del párrafo con el fin de resaltar adecuadamente el párrafo.

Figura 1: Fragmento de la gramática (DTD) JML

```
<!ELEMENT jml (head, body)>
<!-- ===== Document Head ===== -->
<!ELEMENT head (title, author,
  place, date, section, journal,
  keywords)>
<!ELEMENT title (#PCDATA |
  %phrase;)*>
<!ELEMENT author (#PCDATA)>
<!ELEMENT date (#PCDATA)>
<!ELEMENT place (#PCDATA)>
<!ELEMENT section (#PCDATA)>
<!ELEMENT journal (#PCDATA)>
<!ELEMENT keywords (#PCDATA)>

<!-- ===== Document Body ===== -->
<!ELEMENT body %Block;>
<!ATTLIST body
  %attrs;
  >
<!-- ===== Paragraphs ===== -->

<!ELEMENT p %Inline;>
<!ATTLIST p
  relevance CDATA #IMPLIED
  %attrs;
  >
```

El siguiente ejemplo (figura 2) muestra una noticia etiquetada en JML. El contenido del cuerpo ha sido recortado por razones de espacio.

Figura 2: Ejemplo de un documento JML

```
<?xml version="1.0"?>
<!DOCTYPE jml SYSTEM "jml.dtd">
<jml>
<head>
  <title>U.S. Apache helicopters due
    in Albania on Wednesday</title>
  <author>Clark Kent</author>
  <place>Washington</place>
  <date>April 20, 1999</date>
  <section>U.S.</section>
  <journal>CNN</journal>
  <keywords>NATO, Albania,
    Yugoslavia, attack</keywords>
</head>

<body>
  <subtitle>Pentagon narrows causes
    for downed stealth plane
  </subtitle>
  <p relevance="high"
    class="entradilla">
    The Pentagon said that a delayed
    contingent of 24 U.S. Apache attack
    helicopters would begin arriving in
    Albania on Wednesday.</p>
  <p>The helicopters had been expected
    in the Albanian capital</p>
</body>
</jml>
```

2.2. Plegado

Otro de los elementos clave de nuestro proyecto fue el modo en que podríamos ofrecer diferentes alternativas a distintos usuarios sobre un mismo contenido global: el periódico de cada día. Al comienzo del proyecto, se barajaron dos alternativas principales: por una parte, ofrecer al lector sólo aquellos contenidos en los que estuviera interesado; por otra, ofrecer al lector esos contenidos, convenientemente destacados, y referencias sobre contenidos alternativos que le pudieran interesar.

El *Periotrónico* se decantó por esta última opción: ofrecerle al usuario todo el periódico, pero mostrarle solo aquello que le interesara, con lo que evitábamos lo que Alfons Cornellà denomina *infoxicación*, sensación propia de aquellos que se enfrentan a un exceso de información por el poco tiempo de que disponen para enfrentarse a ella. Esta opción evitaba, por otra parte, que el individuo quedara aislado de informaciones que, si bien no eran de su interés *a priori*, si podrían serle de utilidad en un momento dado. De otra forma, el periódico que ofreceríamos estaría más cerca de una mera revista electrónica especializada, y se apartaría del conocimiento enciclopédico que, *per se*, lleva implícita la misma concepción de periódico.

Tal dicotomía (ofrecer todo, mostrar sólo lo que interesa) se consiguió mediante la técnica del *plegado*, que consiste en ofrecer, inicialmente, las secciones solicitadas por el cliente, con los titulares de cada sección, enlazados a las noticias correspondientes. En un principio, esa personalización se ha realizado ofreciendo la posibilidad de elegir las secciones y, aún más, los términos que deban aparecer en una noticia, o en una sección determinada (por ejemplo: "Telefónica" en la sección Economía) para que la noticia le sea mostrada en primer lugar.

Junto a esta opción, se ofrecen además las cabeceras de las demás secciones. Todas las secciones incluyen un botón mediante el cual se ofrecen todos los titulares de la sección (se despliega ésta) o se ocultan (se pliega la sección). La opción inicial de presentación, consiste en mostrar las secciones elegidas, desplegadas, y el resto de las secciones, plegadas.

Este procedimiento, pese a su simplicidad conceptual, se adapta perfectamente a los objetivos planteados: ofrecer todos los contenidos de un periódico sin abrumar al lector con múltiples posibilidades de lectura.

2.3. Personalización. Problemas técnicos

La mayor parte de los servicios personalizados accesibles por Internet en la actualidad consisten en la entrega de información personalizada para cada usuario que se crea dinámicamente en el servidor según el usuario la solicita (tecnologías CGI o ASP). Para esto se requiere tener información relativa a los usuarios en alguna base de datos accesible desde el servidor. Este sistema confiere a los administradores de estos sistemas un gran control sobre el perfil de los usuarios, lo cual puede ser deseable en un sistema de publicación electrónica, como un periódico. En muchos casos, la construcción de estas páginas se consigue combinando el contenido personalizado de los usuarios (datos de los usuarios o información obtenida a partir de esos datos de los usuarios) con la estructura física y visual de plantillas de HTML.

Estos sistemas se han mostrado efectivos para implementar la personalización. No obstante, hay razones que invitan a desarrollar soluciones de otras maneras:

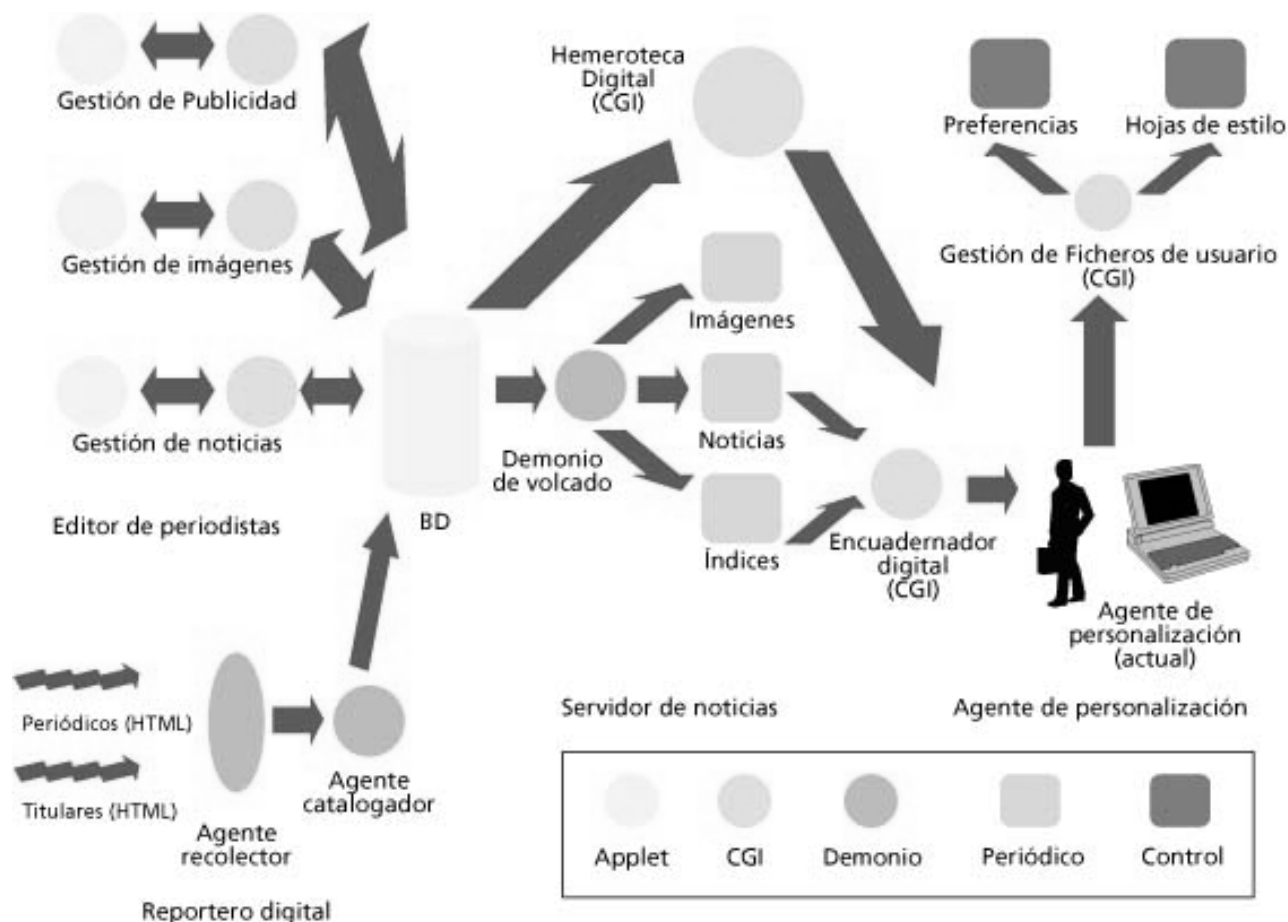
- La capacidad de procesamiento de las máquinas clientes está siendo desaprovechada. Mientras los servidores se encargan de realizar todo el procesamiento de personalización, los navegadores se utilizan como simples visualizadores de páginas (esquema de servidor grueso y cliente fino). En muchos casos, esta diferencia de distribución de tareas no supone un problema apreciable de rapidez de descarga y navegación (normalmente el problema de la lentitud de algunas páginas está más relacionado

con el ancho de banda de la red de comunicaciones y con la recarga de imágenes que con la presumible lentitud de procesamiento en el servidor). Sin embargo, un servidor Web que cree dinámicamente páginas para los usuarios tiene un número máximo de accesos concurrentes mucho más limitado que uno que simplemente entregue ficheros ya construidos. Este último tipo de servidores podrá hacer uso del sistema de caché de los *proxies* (sin los cuales las redes estarían prácticamente colapsadas de tráfico) y todo el cálculo que realicen estará relacionado con las direcciones de los documentos que entrega y no con sus contenidos. Por el contrario, un servidor basado en la creación dinámica de documentos, fácilmente tendrá un cuello de botella en algún recurso utilizado en la fabricación de las páginas (normalmente, una base de datos).

- Las últimas generaciones de navegadores de Internet incluyen plataformas de ejecución de aplicaciones Java que ejecutan en el cliente (*applets*) y de scripts que interactúan con el documento que las incorpora (rutinas de JavaScript o EcmaScript). Con estas tecnologías, es posible la implementación de agentes software que, embebidos en documentos HTML (o en un futuro XML) y descargados en el cliente, sirvan para ejecutar tareas que ayuden al usuario lector a manejar el documento de una forma más cómoda. Estas tareas normalmente son bastante sencillas y están bastante limitadas por restricciones de seguridad. Sin embargo, estos agentes pueden tener bastante control sobre la estructura del documento que las incluye, analizando y modificando sus propiedades conforme a los eventos que se originan por la intervención del usuario. Una ventaja añadida a este tipo de procesamiento radica en el hecho de que no se requieran conexiones adicionales al servidor pudiendo conseguir en algunos casos, dar la impresión de ser auténticas aplicaciones más que simples documentos. Una de las principales ventajas de estos agentes es que, al no necesitar mantener o abrir conexiones con el servidor, los usuarios perciben un grado de interactividad bastante elevado porque no sufren los retrasos de la red. Sin embargo, el almacenamiento de la información personal del usuario en el navegador mediante *galletitas* o *cookies*, aparte de tener numerosas restricciones y problemas de aceptación, imposibilita al usuario tener una versión uniforme de la documentación accedida si decide usar un ordenador distinto para conectarse.

Existen varias técnicas que permiten perfeccionar la creación dinámica de páginas en los servidores

Figura 3. Estructura interna del Perirotónico



aumentando la eficiencia de los mismos y disminuyendo el tiempo de creación de estas páginas (mayor número de páginas servidas por unidad de tiempo y mejor interactividad con el usuario). El empleo de procesos estables en el servidor (*servlets*), o la simple mejora del ancho de banda de las redes de comunicación o de la capacidad de procesamiento de los servidores puede sin duda, disminuir el tiempo de descarga de una página creada de forma personalizada en el servidor. En el *Perirotónico* se ha desarrollado una solución mixta de personalización compartida entre el cliente y el servidor que permita combinar la capacidad de procesamiento de ambas máquinas intentando aprovechar mejor el ancho de banda utilizado en la red y explotando las principales ventajas de cada técnica.

2.4. Combinación cliente-servidor

Así pues, la combinación del trabajo del cliente y del servidor permite la creación de un periódico personalizado y personalizable para cada usuario. Básicamente, la distribución de tareas realizada entre cliente y servidor es la siguiente:

1. Tareas del servidor:

- Selección de las noticias que deben formar parte del periódico personalizado de cada lector. El criterio de selección es personalizable según los gustos del usuario. Nuestro prototipo actual permite exclusivamente la selección de secciones completas, pero más adelante incorporará otros criterios más flexibles.
- Mantenimiento centralizado de la información de perfil y las preferencias de los usuarios. Esta información facilitará la realización de estudios estadísticos o la simple sustitución de la información de la personalización del lector en los casos en el que ésta no pueda ser almacenada en el navegador del cliente.
- Gestión de altas y bajas de usuarios, así como los datos de la suscripción.
- Mecanismos de acceso a las noticias que

no forman parte del periódico personalizado de cada usuario (*hemeroteca*).

2. Tareas del cliente:

- Detección y resaltado de las noticias que se estimen de más interés para el usuario de acuerdo con su personalización.
- Gestión y modificación de las preferencias de los usuarios
- Gestión del plegado y desplegado de las secciones y de los párrafos de las noticias con el fin de conducir primero al lector hacia la información más importante.
- Estructuración del contenido en la pantalla (*layout*) respecto de las preferencias de aspecto visual de cada lector (usando hojas de estilo).

2.5. Hemeroteca

La fecha de la información resulta especialmente relevante para el funcionamiento del sistema. Los documentos periodísticos tienen un periodo de vigencia informativa, más allá del cual pasan a un periodo de latencia, en el que las noticias no tienen ya valor informativo por sí mismas (han quedado superadas por nuevos acontecimientos, por ejemplo), pero pueden servir como contexto a otras más recientes. Finalmente, ciertas informaciones pierden todo su valor informativo para pasar a ser documentos históricos, si es que alcanzan ese valor.

Las noticias que han adquirido cierta antigüedad desde que fueron publicadas son eliminadas del servidor cuando se considera que han perdido interés para el público general. Sin embargo, siempre es posible extraer una versión de esa noticia accediendo directamente a la base de datos donde se almacena. El hecho de abrir la posibilidad de acceder a noticias que han dejado de ser publicadas y de realizar investigaciones en el periódico mediante búsquedas dirigidas por temas otorga al periódico electrónico un valor añadido de importancia tanto o más considerable en la medida en la que los contenidos de esa hemeroteca digital aumenten a lo largo del tiempo. Sin embargo, el acceso a esta base de datos es un proceso que requiere bastante capacidad de procesamiento y que sólo se puede realizar en el servidor (que es donde residen los datos).

Nuestra solución para este problema consiste en la implementación de un programa CGI de búsqueda en la base de datos por distintos criterios combinables entre sí: temas, autores, fechas, secciones, lugares...

3. NUEVAS PERSPECTIVAS

El proyecto *Periotrónico* ha servido para demostrar que los soportes electrónicos no solo ofrecen un acceso diferente a la misma información periodística tradicional, sino que posibilitan nuevas perspectivas para la lectura de prensa, es decir, para la búsqueda, el acceso y la comprensión de la información ofrecida por un periódico.

No obstante, este enfoque sólo se ha planteado a nivel básico, en lo que el equipo investigador entiende ya como la primera parte de un proyecto más amplio. En otras palabras: si el *Periotrónico* se concibió como un periódico electrónico personalizado, el desarrollo de la investigación y de los prototipos obtenidos han permitido plantear otros retos cuyo fin es el de manejar aún mejor la información contenida en un periódico, más allá de la mera elección de posibles contenidos entre todos los ofrecidos por la edición impresa convencional.

Para desarrollar este concepto, el mismo equipo investigador se ha propuesto continuar, en una segunda fase, con una versión ampliada del *Periotrónico*, denominada *Infomedia*.

Infomedia se planteará manejar toda la información contenida en un periódico, y no solo en una edición diaria del mismo, sino relacionándola con lo que ya se ha publicado en otras fechas, fomentando una concepción enciclopédica de la prensa electrónica. Pero, al mismo tiempo, se planteará el modo en que ese proyecto sea económicamente rentable. Ambos objetivos (mejor información, mayor rentabilidad) se desarrollarán a partir de cuatro tareas principales:

1. *La obtención y/o utilización de un lenguaje común de descripción de la información electrónica periodística.* Se trata de utilizar las amplias posibilidades de XML para la descripción precisa de contenidos periodísticos mediante el uso de un lenguaje de marcado de noticias.

En la primera fase del proyecto, como se mencionó anteriormente, no disponíamos de ningún lenguaje de marcado para textos periodísticos, por lo que se diseñó un lenguaje nuevo, JML, basado en XML. Dos años más tarde, nos encontramos con una propuesta de norma para este tipo de lenguaje, NITF [10] elaborada por un organismo dedicado al fomento de la publicación electrónica, con el apoyo de unas cincuenta asociaciones y empresas periodísticas de todo el mundo.

Bien como estándar de facto, bien como punto de partida, el empleo de una norma ya elaborada permitiría reducir el esfuerzo necesario para definir las etiquetas necesarias y fomentar la posibilidad de utilizar robots de búsqueda (que trabajarían con las etiquetas de marcado estándares) mucho más precisos que las actuales.

2. *La creación de comunidades virtuales.* Junto a la mera personalización en el envío de contenidos, existe otra posibilidad, ya explotada en productos accesibles a través de Internet (como los portales). Esta otra modalidad de filtrado de contenidos se conoce como *comunidad virtual*. En ellas, coinciden grupos de usuarios con similares intereses, y gustos comunes (cuya extensión debe ser definida). Esta concepción reúne varias ventajas:

- Permite poner en contacto a gente con afinidades que, de otra forma, no llegaría a relacionarse.
- Permite reducir los esfuerzos técnicos, parcial o totalmente, necesarios para mantener la personalización del periódico.
- Ofrece al periódico la posibilidad de conocer los tipos de público que pueden ser de interés para ciertos anunciantes y para ciertos contenidos. En este caso, el periódico puede ofrecer a los anunciantes diferentes públicos como objeto de su interés. Estos públicos no son tan numerosos en cantidad, pero si pueden ser más receptivos a la publicidad en cuestión, rompiendo con el concepto de publicidad generalista ofrecida por un periódico tradicional.

3. *Posibilidad de difusión multiplataforma.* El gran reto de la implantación masiva de la Web consiste en hacer llegar sus contenidos a través de plataformas más amigables para el público no acostumbrado al ordenador personal; o bien, hacia un público que puede (o que *sólo* puede) recibir información por canales alternativos.

En el primero de los casos, la televisión deberá ser el canal de difusión masiva de los contenidos accesibles en Web. El motivo fundamental de esta decisión, ya en desarrollo, es la enorme extensión de este electrodoméstico incluso entre sectores sociales tecnológicamente poco avanzados.

Ello significará que la información contenida en el futuro periódico electrónico debe pensarse para diferentes medios y, por tanto, con diferentes grados de complejidad (no es lo mismo, por ejemplo, la información accesible a través de un ordenador personal, de un televisor o de un teléfono móvil). 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). Por ejemplo, podremos convertir documentos XML a HTML, SMIL [11] y, en un futuro, a MHEG [12] para mostrarlos mediante *set-top boxes* en un televisor.

Por tanto, deberán establecerse claramente, dentro de un mismo texto, qué partes deberán ser accesibles por cada uno de los soportes utilizados, y cómo. Además, debe plantearse la posibilidad de utilizar un único medio de producción de contenidos con múltiples variedades de presentación a la hora de difundirlo, con el fin de amortizar gastos. Ello es posible utilizando alguna variedad de XML, que permita distinguir claramente las partes del texto, de un lado, y determinar cuáles de esas partes se mostrarán en cada soporte, como se presentarán (empleando hojas de estilo) y de qué modo se navegará dentro de ellas.

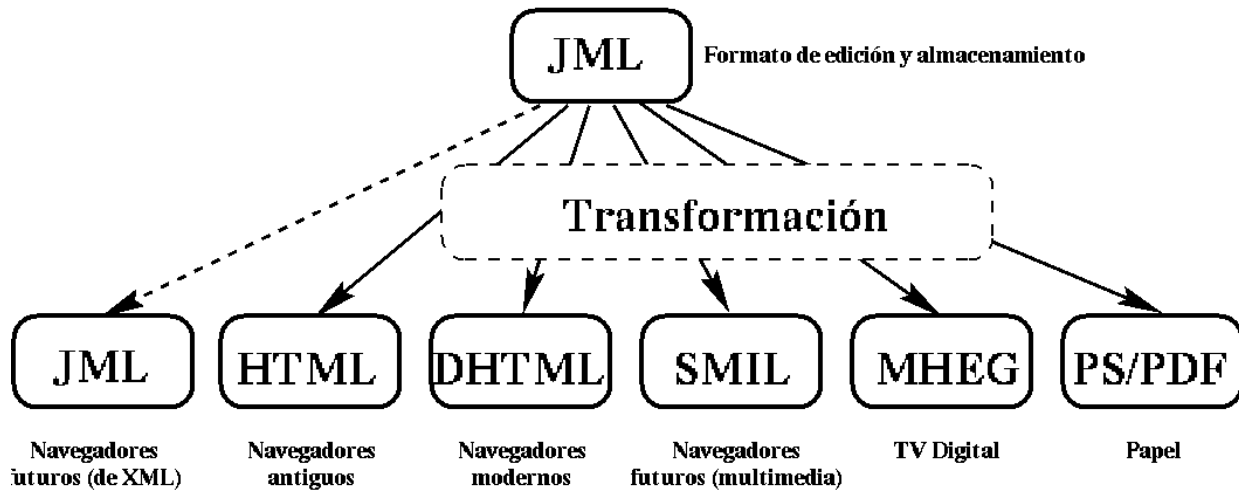
4. *Mayor y mejor relación entre los contenidos periodísticos.* La posibilidad de utilizar herramientas de búsqueda y recuperación de la información aplicadas a textos periodísticos permitiría relacionar contextualmente las diferentes informaciones publicadas por un solo medio periodístico e, incluso, por diferentes medios.

Ello haría posible una mejor información, ofrecida como servicio de valor añadido, de modo que el producto periodístico impreso difícilmente podría competir con el periódico electrónico.

Debe, pues, determinarse de qué modo se pueden establecer estas relaciones entre noticias y de qué forma se puede ofrecer al usuario la posibilidad de buscar entre tales relaciones. El objetivo final, en línea con otras investigaciones sobre recuperación de la información, consistiría en ofrecer al usuario la posibilidad de recuperar una serie de noticias sobre un tema concreto, más allá de la mera coincidencia de las mismas en contener una palabra o una expresión comunes, a partir de opciones de búsqueda entre las que se incluiría el lenguaje natural.

Figura 4: Transformación de texto en XML a diferentes formatos

[5] International Standard Organization: *Information*



En conclusión, el proyecto *Infomedia* constituirá la segunda fase de un ambicioso periódico electrónico cuyo fin será ofrecer información más completa, precisa, fácil de interpretar y adaptable a los intereses de los múltiples lectores de prensa diaria. Y, al tiempo, conseguir que un producto de este calibre permita a la institución que lo desarrolle obtener un beneficio económico.

AGRADECIMIENTOS

El trabajo en el que se basa este documento ha sido parcialmente financiado por el proyecto TEL97-0788 de la CICYT. Queremos agradecer las fructíferas aportaciones de nuestros compañeros Peter T. Breuer, Pilar Diezhandino, Natividad Martínez y Luis Sánchez de la Universidad Carlos III de Madrid. Agradecemos también la ayuda prestada por El PAÍS Digital.

REFERENCIAS

[1] Lawrence, Steve, y Giles, Lee: "Accessibility and Distribution of Information on the Web". *Nature*, 400, 107-109, 1999. Resumen de los propios autores disponible en <http://www.webmetrics.com>

[2] Delgado Kloos, Carlos (director): *El periotrónico. Concepción y desarrollo de un periódico electrónico personalizado* (acceso: 15-10-1999), <http://www.it.uc3m.es/~per>

[3] Tim Bray, Jean Paoli, and C. M. Sperberg-McQueen (eds): *XML: Extensible Markup Language (XML) 1.0, W3 Recommendation*, 10-2-1998 (acceso: 24-3-1999), <http://www.w3.org/TR/REC-xmlW3C>.

[4] W3 Consortium (W3C), <http://www.w3.org>

Processing - Text and Office Systems - Standard Generalized Markup Language (SGML) - ISO 8879:1986, First edition. Ginebra (Suiza), 15-10-1986.

[6] Maler, Eve, y DeRose, Steve: *XML Linking Language (XLink) Design Principles. World Wide Web Consortium Note*, 3-3-1998 (acceso: 15-10-1999), <http://www.w3.org/TR/1998/NOTE-xlink-principles-19980303>

[7] Maler, Eve, y DeRose, Steve: *XML Pointer Language (XPointer). World Wide Web Consortium Working Draft*, 3-3-1998 (acceso: 15-10-1999), <http://www.w3.org/TR/1998/WD-xptr-19980303>

[8] Deach, Stephen (editor): *Extensible Stylesheet Language (XSL) Specification. W3C Working Draft*, 21-4-1999 (acceso: 15-10-1999), <http://www.w3.org/TR/WD-xsl/>

[9] Lassila, Ora, y Swick, Ralph: *Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation*, 22-2-1999 (acceso: 15-10-1999), <http://www.w3.org/TR/REC-rdf-syntax/>

[10] International Press Telecommunications Council (IPTC): *The IPTC now offer an Extensible Markup Language version of the NITF News Industry Text Format*. 1-8-1999 (acceso: 18-10-1999), <http://www.iptc.org/iptc/xmlnitf.htm>

[11] Hoschka, Philipp (editor): *Synchronized Multimedia Integration Language (SMIL) 1.0 Specification. W3C Recommendation*, 15-6-1998 (acceso: 15-10-1999), <http://www.w3.org/TR/REC-smil/>

[12] R. Joseph, J. Rosengren: *MHEG-5: An Overview*. 6-11-95 (acceso: 24-4-1998), <http://www.fokus.gmd.de/ovma/mug/archives/doc/mheg-reader/rd1206.html>