

# Servicio de Selección de Noticias basado en Mashup de Contenidos con CMIS

José M. Jiménez, Guillermo Hernández  
División I+D+i  
Informática Gesfor (Grupo Gesfor)  
Avda Manoteras, 32  
28050 Madrid  
jmjimenezt,ghernandezc@grupogesfor.com

Carlos Á. Iglesias, David Jiménez  
División I+D+i  
Germinus XXI (Grupo Gesfor)  
Avda Manoteras, 32  
28050 Madrid  
cif, djimenezc@germinus.com

**Resumen**—El artículo presenta la propuesta y resultados del proyecto *Contenidos a la Carta*, que propone el uso de tecnología de mashups para la selección de contenidos (noticias), combinado con el uso de la especificación CMIS (*Content Management Interoperability Service*) para la integración de repositorios heterogéneos de contenidos. El artículo presenta la arquitectura propuesta, que define operadores de mashup específicos para contenidos, para lo que se ha extendido la herramienta de creación de mashups MyCocktail.

**Palabras Clave**—CMIS, noticias, mashups, contenidos, REST

## I. INTRODUCCIÓN

La web 2.0 con fenómenos como YouTube ha abierto sin duda una era de los contenidos, tanto en su creación como en su consumo. Han caído las barreras financieras, tecnológicas y culturales que limitan la creación de conocimiento. Sin embargo la “nueva” economía sigue obedeciendo a las viejas reglas y lo que se hace abundante se devalúa. ¿Dónde está ahora el valor? Posiblemente en la presencia en Internet, y en la capacidad de filtrado, la agregación y la remezcla de contenidos y servicios, así como en la conexión intelectual y emocional con los usuarios.

En un mundo donde más gente aporta datos y produce información y conocimiento, una gran parte de estos contenidos presentan escaso interés y/o calidad. La necesidad de localizar, combinar y posicionar los contenidos no está restringido a usuarios finales, sino que es también una necesidad empresarial para los proveedores de contenidos, que necesitan combinar sus contenidos propios y externos para poder recuperar la inversión realizada en su creación y añadirles valor.

El proyecto *Contenidos a la Carta* investiga y experimenta en técnicas y herramientas que faciliten la composición de ofertas personalizadas de contenidos, en este caso noticias. También investiga en técnicas y herramientas de posicionamiento de contenidos en español, así como técnicas de rastreo y protección de los contenidos en español en la red.

El proyecto se centra en el ámbito de las noticias de prensa y de la problemática de un proveedor de contenidos como la Agencia EFE, primera agencia de noticias en español y cuarta agencia mundial de noticias.

Para la composición de contenidos, el proyecto **Contenidos a la Carta** [2] investiga la aplicación del reciente estándar CMIS (*Content Management Interoperability Services*) [7] para proporcionar un servicio de interoperabilidad funcional

entre los diferentes repositorios de contenidos de Agencia EFE. Mediante la aplicación de técnicas semánticas, permitirá ofrecer interoperabilidad semántica entre sus metadatos. El proyecto también investiga en la aplicación de técnicas de mashups de contenidos que permitan combinar y adaptar los contenidos para innovar en el proceso de composición de nuevos contenidos, mediante una interfaz gráfica de usuario.

*Contenidos a la Carta* pretende también investigar en el rastreo de contenidos en Internet para detectar copias. Este tema es altamente relevante para garantizar los derechos de los proveedores de contenidos. Aunque en el mercado existen productos comerciales, como *Attributor*, estos productos están en inglés y las adaptaciones al castellano son muy pobres.

El proyecto experimenta sobre nuevos métodos para incrementar la eficacia y la eficiencia de las empresas dedicadas a la creación, transformación y distribución de contenidos, aumentando su competitividad y aumentando la presencia de contenidos digitales de calidad en Español en Internet, al ofrecer nuevos canales de distribución de noticias innovadores y flexibles, adaptables a los continuos cambios que aparecen en el mundo de la gestión de contenidos.

El resto del artículo se estructura como sigue. La sección II describe el estándar CMIS. A continuación, la sección III revisa el estado del arte en tecnologías de mashups y, en concreto, de la herramienta de creación de mashups MyCocktail. Posteriormente, en la sección IV se ilustra la solución propuesta mediante una descripción de la arquitectura y un caso de uso en la sección V. Por último, se recogen las conclusiones y trabajos futuros en la sección VI.

## II. EL ESTÁNDAR CMIS

El estándar CMIS (*Content Management Interoperability Services*, Servicios de Interoperabilidad de Gestión de Contenidos) [7] ha sido impulsado por IBM, EMC y Microsoft y respaldado por Opentext, Oracle, Alfresco y SAP, y viene a resolver uno de los mayores problemas que las empresas han acusado con respecto a la gestión de su información empresarial, que es la integración de repositorios de contenidos.

A través de un juego común de servicios, CMIS permite interactuar con los diversos repositorios de gestión de contenidos, sin importar quién es el fabricante del repositorio o cómo éste está implementado.

El objetivo de este estándar es permitir que las aplicaciones puedan trabajar con cualquier tipo de repositorio de

contenidos de manera uniforme y busca asegurar la interoperabilidad de las aplicaciones que usan múltiples repositorios de contenidos.

Para resolver tales problemas, CMIS define un modelo de dominio para interactuar con repositorios ECM (*Enterprise Content Management*) haciendo uso de Servicios Web. Provee un gestor de contenidos para modelos de datos de dominios específicos, un conjunto de servicios genéricos que actúan en ese modelo de datos y varios protocolos para acceder a esos servicios, incluyendo SOAP (*Simple Object Access Protocol*) y REST/Atom (*Representational State Transfer*) [6].

Dentro del modelo de dominio, CMIS define un modelo de datos, donde se especifican los elementos necesarios para trabajar con un gestor de contenidos, i.e. el repositorio y los objetos básicos que componen un repositorio: documentos, directorios, relaciones o la política administrativa. Además de definir con el modelo de datos los elementos del repositorio con los que operará CMIS, también es necesario definir en el modelo de dominio un conjunto de servicios genéricos que actuarán en ese modelo de datos. Estos servicios ofrecen las operaciones típicas de gestión de contenidos en un repositorio, tales como creación, búsqueda, edición o borrado de contenidos, conexión y desconexión a un repositorio, etc [7].

Dado que el objetivo del estándar es facilitar el intercambio de información y documentos entre entornos y repositorios documentales diferentes, eliminando los problemas de migración entre una plataforma y otra, y facilitando que coexistan sistemas de diferentes fabricantes (permitiendo la federación, por ejemplo), se hace necesario definir, además del modelo de dominio, una serie de API's de comunicación, orientadas a la definición de servicios Web, y un protocolo de publicación Rest/Atom que puede ser usado por aplicaciones para trabajar con uno o más gestores de repositorios de contenidos u otros sistemas. Los protocolos de comunicación empleados en el estándar son REST/Atom y SOAP Web services.

Para que sea posible ese intercambio de información entre repositorios documentales, es necesario que esos repositorios implementen el estándar CMIS. Alfresco, empresa participada por SAP, actualmente implementa esta especificación en su última herramienta lanzada al mercado.

Alfresco es un sistema de administración de contenidos de código abierto, que proporciona gestión de documentos, herramientas de colaboración, gestión de contenidos Web, además de otras muchas funcionalidades. La arquitectura de Alfresco está basada en tecnologías de código libre tales como Spring [23], Hibernate [20], Lucene [19], modernos estándares como JSR-168, JSR-170 [13], servicios Web, Java Server Faces [21] y contribuciones de la comunidad de software libre. Alfresco incluye un repositorio de contenidos, un framework de portal web para administrar y usar contenido estándar en portales, un sistema de administración de contenido, capacidad de virtualizar aplicaciones web y sitios estáticos vía Apache Tomcat, búsquedas vía el motor Lucene y flujo de trabajo en jBPM. Alfresco está desarrollado en Java [3].

Alfresco Labs3 [3], la última versión de código abierto de Alfresco lanzada al mercado, incorpora una serie de servicios Web basados en CMIS que cubren las funcionalidades

especificadas en dicho estándar.

CMIS debería hacer que los ECM pudiesen relacionarse mediante la tecnología de mashups, construyendo aplicaciones más ricas y rápidas. En definitiva, se espera que CMIS se convierta en una revolución en la gestión de contenidos equiparable a la que supuso SQL en el mundo de las bases de datos [14].

### III. TECNOLOGÍA DE MASHUPS

Los mashups son composiciones de contenidos y servicios obtenidos de diferentes fuentes que se presentan de manera homogénea ofreciendo un valor adicional respecto a los datos presentados por separado. En la web existe un número creciente de mashups que se ha visto acentuado por la aparición de APIs que facilitan la tarea de la reutilización del código [24].

Este aumento de los mashups junto con la tendencia cada vez mayor de la creación de contenidos por parte de los usuarios en Internet motiva la aparición de herramientas para la fácil creación y edición de mashups.

Según Gartner [10], los mashups han pasado del puesto 6 en 2008 al puesto 5 en 2009 en el ranking de tecnologías emergentes, destacando su penetración en las Empresas. Gartner predice que el 80% de las nuevas aplicaciones estarán basadas en mashups. Actualmente, hemos identificado las siguientes tendencias en tecnologías de mashups. Para cada una de ellas están surgiendo herramientas de creación y edición de mashups.

- *Mashups de Datos y Servicios* – Coleccionan datos y servicios de diferentes orígenes y los mezclan en una interfaz gráfica común. Algunas herramientas para la creación de este tipo de mashups son: YahooPipes [17], Popfly [9], MyCocktail [22].
- *Mashups de Procesos* – Permiten combinar diversos procesos como alertas o envío de correos. En la actualidad existen pocas herramientas para la creación de este tipo de mashups y, a excepción de OPUCE [15], la mayoría son propietarias: Serena Software [25], K2-BlackPearl [16].
- *Mashups de Interfaz de Servicios* – Permiten la creación de gadgets de una forma sencilla, evitando la programación y utilizando interfaces gráficas usables por usuarios no muy experimentados en estos entornos. Morfeo Fast [8].
- *Mashups de Navegador* – Son aplicaciones en forma de plugin de navegadores que permiten extraer información de páginas web para ser combinada en forma de mashups. Permite consumir la información de las páginas de forma personalizada, enriqueciendo la experiencia del usuario. Debido a la novedad de este tipo de mashups las herramientas existentes son poco intuitivas y presentan muchas limitaciones. Intel Mash Maker [12], Ubiquity [26], y Piggy Bank [5] son algunos ejemplos.

#### A. Herramienta de Mashups MyCocktail

MyCocktail [22], Romulus Mashup Builder, es una aplicación web que proporciona al usuario una interfaz gráfica para construir mashups fácilmente, permitiendo al usuario

desarrollar mashups de manera más rápida, incrementando así la productividad.

Esta herramienta permite al usuario combinar información proveniente de diferentes servicios, que puede ser modificada con operadores y más tarde presentada con una gran variedad de renderizadores. Todo este proceso se lleva a cabo mediante una interfaz gráfica de usuario de fácil manejo, que permite combinar componentes arrastrando y soltando. El tiempo que lleva desarrollar un mashup se reduce así considerablemente y se mejora la productividad.

MyCocktail permite a los diseñadores y programadores usar servicios sin preocuparse de detalles de bajo nivel. Los usuarios solo tienen que manejar una serie de herramientas que se proporcionan y MyCocktail hará las peticiones a los diferentes servicios.

MyCocktail está basado en Afrous [18] y proporciona tres tipos distintos de componentes, que combinados dan lugar al mashup:

- *Servicios.* Pueden invocarse varios servicios REST por defecto, como del.icio.us, Yahoo Web Search, Google AJAX Search, Flickr, Twitter, Amazon, etc.
- *Operadores.* La información obtenida se puede procesar por medio de operadores. Por ejemplo, es posible ordenar, filtrar o agrupar información según parámetros.
- *Renderizadores.* La información se puede presentar de diversas formas: HTML, diagramas estadísticos (gráfico de tarta o de barras), Google Maps.

Los pasos que suelen seguirse para contruir un mashup son:

- 1) El usuario obtiene información de uno o varios servicios.
- 2) Los datos obtenidos pueden filtrarse y procesarse usando los operadores proporcionados por la herramienta para extraer información útil para el mashup.
- 3) La información resultante puede mostrarse en HTML, en diagramas estadísticos o en mapas usando los renders.
- 4) Se exporta el mashup en uno de los diversos formatos que ofrece MyCocktail: JavaScript, HTML, Google Gadget o Netvibes Gadget.

La figura 1 ilustra el aspecto general de la herramienta de mashups. En ella se pueden observar dos partes bien diferenciadas. En la izquierda tendríamos un ventana donde se listan los servicios disponibles, los operadores y los renders. En la ventana de Servicios, el usuario obtendría contenidos a través de llamadas a los servicios Web de los Sistemas de Gestión de Contenidos y la de los principales portales Web2.0 como Flickr, Google Maps, del.icio.us, Twitter o buscadores como Google y Yahoo.

Los operadores permitirían realizar operaciones con los resultados obtenidos de los servicios. Con estos operadores el usuario podrá realizar mashups de contenidos. Cabe destacar que se incluirán operadores específicos para noticias. Por último los renders nos permitirán mostrar y publicar el resultado en varios formatos, entre ellos HTML, Javascript o como un Gadget. La otra parte que forma la herramienta de mashups será la ventana central donde se desplegaran los operadores, renders o las llamadas a los servicios.

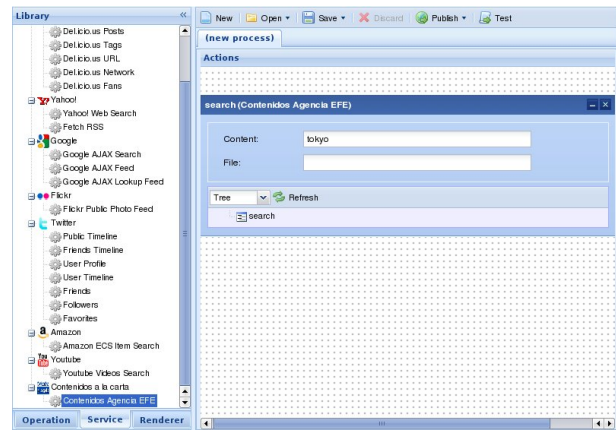


Fig. 1. Herramienta de Construcción de Mashups MyCocktail

#### IV. ARQUITECTURA DE CONTENIDOS A LA CARTA

Esta sección introduce brevemente el proyecto en el que se enmarca la investigación, Contenidos a la Carta, así como la arquitectura propuesta para la realización de mashups de contenidos.

La meta de este proyecto es mejorar la capacidad de ofrecer noticias personalizadas a determinados tipos de usuarios y automatizar determinados procesos relacionados con la publicación de noticias, como pueden ser la traducción, la difusión y la protección de los contenidos relativos a la noticia. Esto supondrá un avance significativo en la industria de los medios de comunicación, ya que permitirá a éstas disponer de unos contenidos más ricos y de mayor calidad, a la par que aumenta las posibilidades de difusión de sus noticias a lo largo de la Internet. Además, la plataforma de Contenidos a la Carta pretende conseguir que el uso de este tipo de sistemas se extienda en todas las empresas de este sector, ya que se requieren pocos conocimientos técnicos para la composición de noticias personalizadas y de gran calidad.

Contenidos a la Carta es, por tanto, un proyecto orientado a la investigación de métodos de selección, personalización y difusión de noticias mediante la reutilización de diversos contenidos multimedia alojados en sistemas de gestión de contenidos heterogéneos a través del estándar CMIS. Esto será posible gracias a la plataforma de Contenidos a la carta, que proporcionará herramientas de tipo mashup, adaptables a las necesidades y requerimientos de cada usuario, para integrar y componer noticias a partir de diferentes orígenes de información. Para poder utilizar contenidos de sistemas de gestión de contenidos diferentes de una manera flexible, en el marco de proyecto se tiene previsto desarrollar una capa de homogeneización de servicios de gestión de contenidos. Esta capa de homogeneización será posible a través del estándar CMIS.

La arquitectura del proyecto Contenidos a la Carta está formada, por tanto, por una herramienta de Mashups y por uno o varios sistemas de gestión de contenidos. De esta forma, el usuario tendrá ante sí una plataforma donde podrá recuperar contenidos de fuentes heterogéneas a través de servicios basados en CMIS, podrá trabajar con ellos formando noticias personalizadas a través de los operadores específicos para noticias que se han diseñado en la herramienta de mashups y los publicará en diferentes formatos de representación, ya

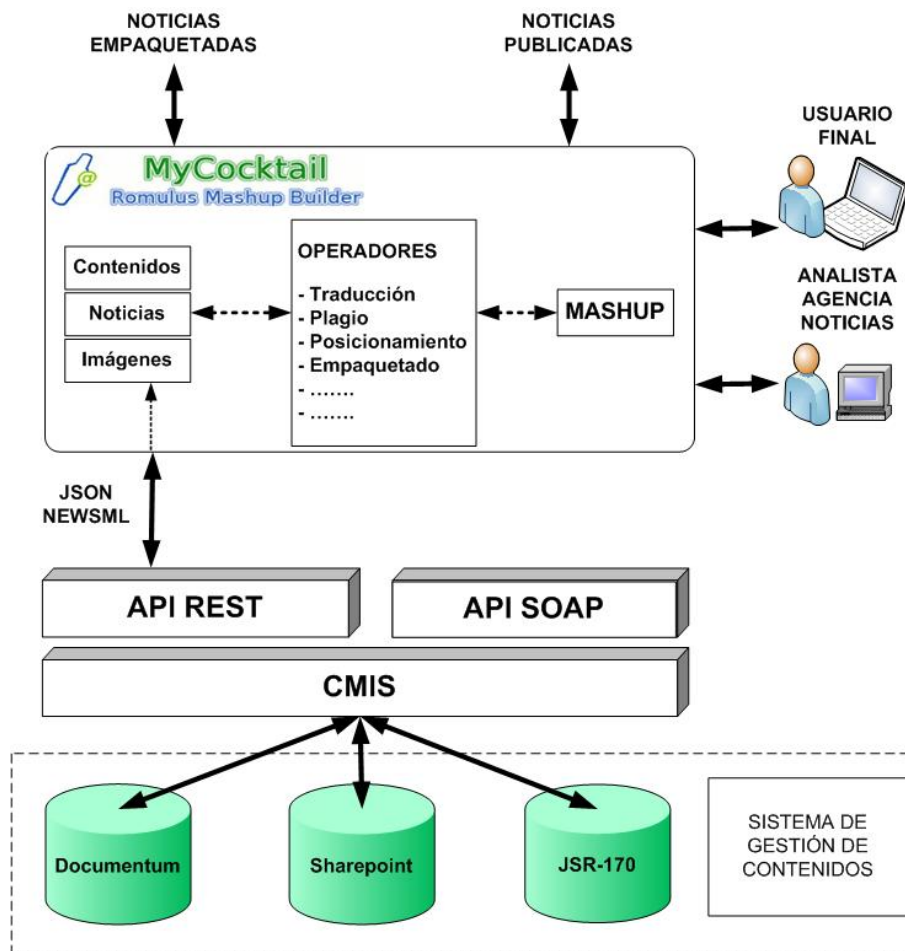


Fig. 2. Arquitectura de Contenidos a la Carta

sea HTML, XML, etc. En la figura 2 podemos observar el esquema de la arquitectura general de Contenidos A la Carta.

Uno de los dos elementos principales de la arquitectura de Contenidos a la Carta son los *Sistemas de Gestión de Contenidos*. La idea principal del proyecto es comunicar estos sistemas con la herramienta de mashups, de tal manera que podamos operar con los contenidos almacenados en los diferentes repositorios y crear nuevos contenidos mezclando algunos de ellos.

Dado que el proyecto está enfocado a la experimentación en la mejora del proceso de oferta personalizada de noticias, es importante tener en cuenta que los contenidos que se manejarán en la herramienta de mashups y se almacenarán en los repositorios serán noticias, por lo que se han querido adaptar los modelos de tipos de contenidos de los Sistemas de Gestión de Contenidos a algún estándar de noticias, que favorezca el intercambio de noticias como NewsML.

NewsML [11] es un estándar desarrollado por el International Press Telecommunications Council (IPTC [1]) que define un formato común para el intercambio de noticias, sin tener en cuenta el tipo de contenido multimedia en el que se presenta la noticia. Este intercambio de noticias es un método que permite no solo transmitir los contenidos de noticias, sino además describir el contenido de manera abstracta, mediante metadatos, con información relativa acerca de cómo manejar las noticias de una manera apropiada, así como su empaquetado e, incluso, el tipo de transmisión que se llevará a cabo.

NewsML es un lenguaje de contenedores de noticias digitales. Es decir, con NewsML no pueden crearse noticias en un formato concreto: es preciso disponer de ellas previamente en otros formatos. Su utilidad es transportar paquetes de contenidos periodísticos, sea cual sea su formato y su forma de difusión: texto, imagen y sonido. NewsML, permite contener más de un paquete informativo (o, simplificando, más de una noticia) en un solo documento. La estructura básica de un documento NewsML se muestra en la figura 3.

```
<NewsML>
  <NewsItem>
    <NewsComponent >
      <ContentItem>
        (Aquí puede el contenido de una pieza
        informativa, en cualquier formato, o una
        referencia a ese contenido, que esté físicamente
        en otro fichero.)
      </ContentItem>
    </NewsComponent >
  </NewsItem >
</NewsML>
```

Fig. 3. Formato NewsML

Básicamente:

- Cada pieza está incluida en un elemento ContentItem.
- El conjunto de varios ContentItem (un texto, una foto, un vídeo y un gráfico) pueden formar parte de un NewsComponent (una noticia).
- Un NewsItem puede contener varias noticias, varios NewsComponent.

Un documento NewsML puede llevar además toda una serie de metadatos que, en resumen, pueden ser de tres tipos:

- Datos relativos a la transmisión del documento NewsML en conjunto: quién lo envía, a quién, cual es su prioridad, la fecha de caducidad, etc.
- Datos sobre el documento en cuestión, o sobre partes determinadas: por ejemplo, en cada una de las piezas que se incluye, pueden añadirse elementos que describan el tema, o los protagonistas, o las relaciones de esa pieza con otras en el mismo documento, etc.
- Datos sobre cómo se normalizan los datos incluidos en otros elementos de NewsML. Por ejemplo, si en un NewsItem se incluye una noticia sobre fútbol, se puede incluir un elemento, o Topic, que describa temáticamente el deporte sobre el que trata la noticia. Ese elemento puede incluir un código; será entonces, un elemento añadido, o Catalog, el que indique cual es la clasificación o vocabulario del que forma parte el código, y dónde se encuentra la clasificación completa, en Internet.

La mayor parte de los metadatos que forman parte de NewsML pueden situarse en múltiples lugares del documento, aplicados a una sola pieza o a varias. Como puede deducirse, NewsML tiene dos características básicas:

- Una estructura modular, que además permite que las piezas estén situadas físicamente dentro del documento NewsML o fuera de él, unidas mediante referencias a objetos externos.
- La posibilidad de realizar una descripción estructural, aunque sea solo para distinguir cada una de las piezas (pero no sus partes), y una descripción semántica, también de tipo general.

La herramienta de mashup permite a los usuarios la creación de aplicaciones que manejen las fuentes de datos de manera sencilla y flexible, dando la posibilidad de aplicar un amplio conjunto de operadores que transformarán y combinarán los contenidos. Las entradas de estos operadores, que a partir de este momento denominaremos “*pipe*”, pueden ser la salida producida al procesar una o varias entradas por parte de otro operador o un servicio de datos proporcionado por la capa de interoperabilidad CMIS.

Los servicios que ofrece la capa de interoperabilidad pueden ser servicios Web XML o servicios REST. A las salidas de datos, además de generar *pipes*, también se les puede aplicar un proceso denominado renderización de contenidos, que no es más que aplicar un proceso que produce una salida visualizable del contenido por algún tipo de cliente estándar (Navegador Web, Navegador Móvil). Mediante las operaciones de renderización se permitirá seleccionar el tipo de dispositivo sobre el que se visualizará el contenido.

La comunicación entre los Sistemas de Gestión de Contenidos y la herramienta de mashups se realizará mediante peticiones por parte de la herramienta de mashups a los servicios Web de los Sistemas de Gestión de Contenidos para recuperar contenidos. La herramienta de mashups se encargará de buscar y recuperar contenidos en forma de noticia, procesarlos y permitir al usuario que realice un nuevo contenido a partir de la mezcla de varios de ellos.

La herramienta de mashups tiene definida como entrada ficheros JSON, que posteriormente procesa y devuelve el

resultado en diversos formatos, por lo que es necesario que las respuestas de los servicios Web de los repositorios ECM sean en formato JSON. En caso de que alguno de dichos repositorios no soporte trabajar con formatos JSON, será necesario incluir un conversor de este formato.

Este es el caso de los Sistemas de Gestión de Contenidos Alfresco Labs3 utilizados en Contenidos a la Carta. En este caso, se deberá incluir un conversor, como Apache Abdera [4], para la traducción de atom/xml (el formato utilizado por Alfresco Labs3) a JSON. La elección de Alfresco Labs3 como Sistema de Gestión de Contenidos para Contenidos a la Carta es debido a que es uno de los primeros gestores de contenidos de código abierto que implementa el estándar CMIS.

Dentro del proyecto, hemos identificado los siguientes operadores para la realización de mashups de contenidos:

- **Consulta.** Los servicios de consulta de noticias diseñados constan de un operador de búsqueda textual de contenidos, un operador de consulta avanzada basada en metadatos y un operador de búsqueda multilingüe, donde el usuario tendrá la posibilidad de buscar noticias en diversos idiomas.
- **Operaciones de posicionamiento.** La herramienta de mashups dispondrá de un operador de extracción de metadatos de la noticia y sugerencia de términos para su posicionamiento en buscadores (SEO, *Search Engine Optimization*), logrando que las noticias tengan un mayor alcance y difusión. Este operador permite automatizar la tarea de catalogación y extracción de palabras clave, lo que en el contexto de las noticias resulta fundamental ya que permite evitar demoras en la publicación de las mismas.
- **Rastreo y protección de contenidos.** El uso de este operador servirá para detectar copias de las noticias, garantizando así los derechos del autor de las noticias. Este operador utilizará técnicas de detección de plagio comparando el mayor o menor grado de coincidencia entre fragmentos de las diferentes obras según diferentes propiedades, tales como la frecuencia de palabras, el uso de un tipo u otro de palabras o aparición de erratas.
- **Traducción de noticias.** Este operador facilitará al usuario la traducción de noticias, tanto del texto que la forma como los metadatos de la noticia, lo que permitirá realizar búsquedas multilingües.
- **Empaquetado.** Una vez generado un “*pipe*”, este operador empaquetará el conjunto de contenidos de noticias, ya sea artículos de texto, imágenes, vídeo o cualquier tipo de contenidos multimedia, en un package para su difusión y exportación a NewsML. La exportación a este formato de intercambio de noticias, junto a los operadores de posicionamiento y traducción, hacen que la difusión de las noticias sea mucho mayor que la obtenida utilizando los métodos de publicación tradicionales.

Por último, cabe destacar que la plataforma Contenidos a la carta integra un módulo que permite generar aplicaciones adaptables bajo la tecnología de Mashup, también se permite a los usuarios seleccionar el formato de visualización que se aplicará a los contenidos mediante filtros de renderización, esto permitirá que los usuarios puedan visualizar los contenidos en distintos dispositivos. Esto quiere decir que la in-

formación del contenido se puede separar de la representación del contenido.

V. CASO DE USO

En esta sección se desea mostrar al lector un ejemplo de aplicación y uso de Contenidos a la Carta, para que vea las mejoras que puede ofrecer esta plataforma al proceso de creación de noticias personalizadas por parte de las agencias de noticias.

Anteriormente, cuando hablábamos de la arquitectura de la plataforma señalábamos que utilizaríamos el estándar NewsML-G2 para modelar los diferentes tipos de contenidos que soportará el repositorio. Esto es así porque las agencias de noticias hacen uso de este formato estándar de intercambio de noticias para la creación y definición de contenidos. Contenidos a la Carta aprovechará este formato de intercambio de noticias, además de la tecnología CMIS, para automatizar, en la medida de lo posible, todo el proceso de creación de una noticia y centralizar dicho proceso a través de una única plataforma.

En este ejemplo nos situaremos en el papel de un periodista que desea realizar un artículo sobre destinos turísticos en Italia. Para crear dicho artículo, nuestro periodista utilizará una serie de contenidos que tiene a su disposición, como son una serie de reportajes de texto sobre distintas ciudades y algunas fotografías tomadas por los fotógrafos de la agencia. La secuencia de pasos que el periodista deberá seguir estará compuesta por la búsqueda de los artículos de texto e imágenes, la reunión de dichos contenidos en una misma noticia y por último, subirlo al repositorio de gestión de contenidos, donde la noticia, aprovechando las funcionalidades de los repositorios ECM, podría ser revisada por un supervisor y, posteriormente, apobar su publicación. Mediante la plataforma de Contenidos A la Carta todas estas operaciones pueden ser realizadas desde un mismo interfaz, mediante el cual se podrá operar con los contenidos almacenados en los repositorios de contenidos, creando nuevas noticias a partir de ellos y dándolas de alta en el sistema.

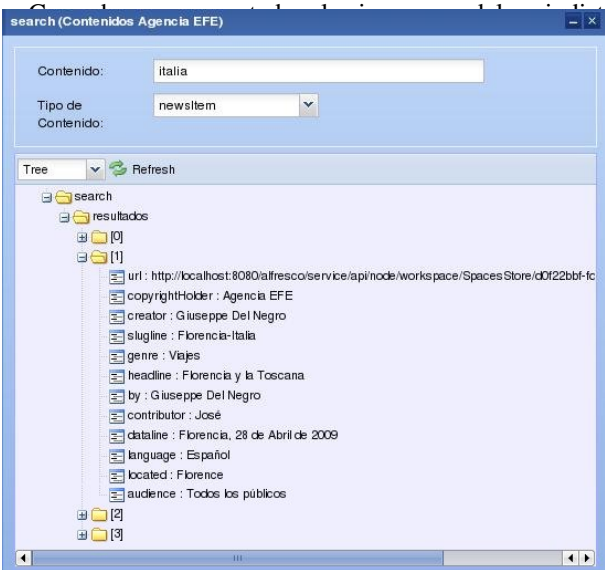


Fig. 4. Búsqueda de Noticias

Mediante este servicio de Búsqueda de Noticias, la

plataforma realizará una llamada al repositorio ECM, apoyándose en el estándar CMIS, buscando coincidencias entre la palabras clave que el usuario ha introducido y los metadatos de las noticias. El repositorio de contenidos devolverá una lista, en formato JSON, de las noticias que contiene algún metadato relacionado con las palabras clave que el usuario ha introducido.

Una vez que el periodista ha seleccionado algunos artículos con los que le gustaría construir el reportaje, podrá utilizar el servicio de empaquetado de noticias para ello. Mediante este servicio se podrá construir, a partir de varios contenidos, una noticia completa publicable. En la figura 5 podemos ver que el usuario podrá indicar los metadatos del nuevo reportaje. Una vez que el servicio se lleve a cabo, el sistema automáticamente dará de alta el reportaje en el repositorio donde, como hemos dicho anteriormente, podrá entrar en una cadena de revisión

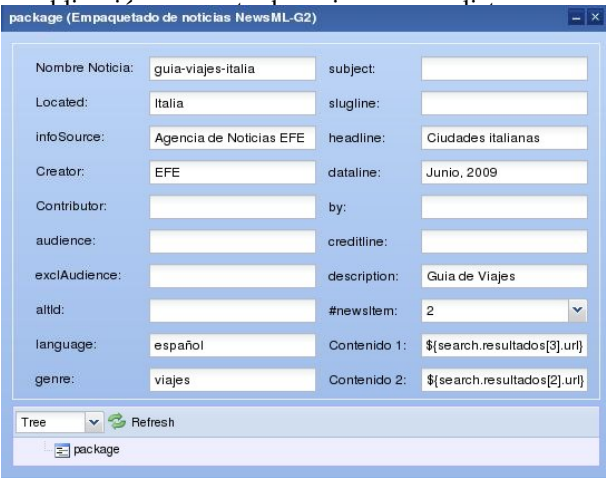


Fig. 5. Empaquetado de Noticias

Las noticias creadas mediante la plataforma de Contenidos A la Carta pueden crear rápidamente noticias personalizadas ricas en contenido mediante una sola plataforma. Otra de las ventajas es que además de aprovechar los recursos propios de la agencia, desde la misma plataforma se podrán acceder a



Fig. 6. Mashup de contenidos del repositorio y Google Maps

Como ejemplo de esta capacidad de aprovechar todas las posibilidades que ofrece Internet, en la figura 6 se muestra

cómo el usuario puede utilizar sus contenidos y, por ejemplo, la API de Google Maps para realizar, mediante uno de los servicios definidos en la plataforma, un mapa interactivo con las noticias (en este caso reportajes) distribuidas por su lugar de origen.

## VI. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo de investigación se ha presentado el proyecto Contenidos a la Carta, basado en tecnologías de mashups e interoperabilidad de contenidos para la composición de ofertas personalizadas de contenidos.

El trabajo ha definido la arquitectura del proyecto, formada, principalmente, por una herramienta de Mashups y por uno o varios sistemas de gestión de contenidos. La tecnología de mashups se ha postulado como una de las tecnologías emergentes en la gestión y personalización de contenidos. Los operadores de contenidos para realizar mashups presentados están ofreciendo resultados interesantes, y grandes posibilidades si los combinamos con otros operadores existentes. Por ejemplo, podemos filtrar noticias y mostrar estadísticas gráficas de autores o mostrar las noticias en un mapa según el lugar de la noticia.

Actualmente Contenidos a la Carta se encuentra en fase de desarrollo y tiene también la intención de lograr potenciar el posicionamiento de las noticias en buscadores, mediante técnicas SEO (*Search Engine Optimization*) [27].

## AGRADECIMIENTOS

Este proyecto ha sido financiado por el Ministerio de Industria, Turismo y Comercio, dentro de la convocatoria 2/2008 del subprograma Avanza I+D, como proyecto de Desarrollo Experimental (TSI-020501).

## REFERENCIAS

- [1] International press telecommunications council, consorcio que agrupa a las más importantes agencias de noticias y empresas de comunicación, disponible en <http://www.iptc.org/>.
- [2] Web de contenidos a la carta. disponible en <http://alacarta.germinus.com>, 2008.
- [3] Alfresco. Alfresco labs3, un sistema gestor de contenidos de código abierto, disponible en <http://www.alfresco.com>, 2008.
- [4] Apache. Web del proyecto apache abdera, una implementación abierta de atom, disponible en <http://abdera.apache.org/>, 2009.
- [5] P. Bank. Piggy bank, 2009.
- [6] R. Cover. Content management interoperability services (cmis), información disponible en <http://xml.coverpages.org/cmish.html>. Technical report, OASIS, 2008.
- [7] EMC Corporation, IBM Corporation, and Microsoft Corporation. *Content Management Interoperability Services, borrador del estándar disponible en http://www.alfresco.com/about/cmish/cmish-draft-v0.5.zip*, 8 2008.
- [8] M. FAST. Morfeo fast, disponible en <http://fast.morfeo-project.eu/>, 2009.
- [9] M. P. Fly. Microsoft pop fly, disponible en <http://www.popfly.com>, 2009.
- [10] Gartner. Information technology research and advisory company, disponible en <http://www.gartner.com>, 2009.
- [11] K. Holland. IPTC Standards: EventsML-G2 version 1.1, NewsML-G2 version 2.2, SportsML-G2 version 2.0. Guide for Implementers. Technical report, IPTC Standards. International Press Telecommunications Council, 2009.
- [12] I. M. Maker. Intel mash maker, disponible en <http://mashmaker.intel.com>, 2009.
- [13] S. Microsystems. Jsr 170: Content repository for java technology api, disponible en <http://jcp.org/en/jsr/detail?id=170>, 2009.
- [14] J. Newton. Hacia la estandarización ecm con cmis, artículo disponible en <http://www.techweek.es/gestion-documental/opinion/1003918003401/estandarizacion-ecm-cmis.1.html>, Noviembre 2008.
- [15] P. Opuce. Proyecto opuce, disponible en <http://www.opuce.tid.es/>, 2009.
- [16] K. B. Pearl. K2 black pearl, disponible en <http://www.k2.com/en/displaycontent.aspx?id=903>, 2009.
- [17] Y. Pipes. Yahoo pipes, disponible en <http://pipes.yahoo.com>, 2009.
- [18] A. Project. Afrous project web site, disponible en <http://afrous.com/>, 2009.
- [19] A. L. Project. Apache lucene project web site, disponible en <http://lucene.apache.org/java/docs/>, 2009.
- [20] H. Project. Hibernate project web site, disponible en <http://www.hibernate.org/>, 2009.
- [21] J. Project. Javaserfaces project web site, disponible en <http://java.sun.com/javaee/javaserfaces/>, 2009.
- [22] R. Project. Mycocktail web site, disponible en <http://www.ict-romulus.eu/web/mycocktail>, 2009.
- [23] S. Project. Spring project project web site, disponible en <http://www.springsource.org/>, 2009.
- [24] Salesforce. Mashups: The what and why, 2007.
- [25] S. Software. Serena software, disponible en <http://www.serena.com/>, 2009.
- [26] Ubiquity. Ubiquity, disponible en <http://labs.mozilla.com/projects/ubiquity/>, 2009.
- [27] Wikipedia. Posicionamiento en buscadores, 2009.