



Comunicación

123

PODCASTING BASADO EN WEB SEMÁNTICA: UNA ESTRATEGIA PARA LA DIVULGACIÓN DE CONTENIDOS PARLAMENTARIOS

Francisco Chávez Gutiérrez

Jefe de la Unidad de Informática
Parlamento de Canarias

Elena Sánchez Nielsen

Profesora del área de Ciencias de la Computación e Inteligencia Artificial
Universidad de La Laguna

Palabras clave

Podcasting, Web Semántica, Tesoros, Gobierno Electrónico.

Resumen de su Comunicación

El uso de podcasting puede proporcionar un nuevo mecanismo en el acceso de los ciudadanos a la información. Actualmente, la tecnología de podcasting está siendo empleada como una herramienta efectiva y económica para la divulgación de contenidos de audio y vídeo sobre dispositivos móviles y ordenadores personales. Su esencia se centra en crear contenidos multimedia para audiencias que deseen escuchar y ver cuando ellos deseen, donde ellos deseen y como ellos deseen.

El presente trabajo se enmarca dentro de la propuesta de estrategias para el incremento de la eficiencia y efectividad de las funciones parlamentarias en relación con la divulgación de contenidos parlamentarios a los ciudadanos. Con este fin, en este trabajo proponemos una solución que utiliza el paradigma de Web Semántica para proporcionar a los ciudadanos suscripciones a contenidos parlamentarios acordes a sus preferencias y obteniendo de esta manera, únicamente los segmentos multimedia de su interés. En particular, mediante la utilización de tesauros describimos el contenido de la información audiovisual e introducimos tres nuevos componentes en los sistemas tradicionales de podcasting: (i) catalogación, (ii) división y (iii) distribución personalizada. Usando estos elementos, establecemos cómo presentar resultados podcasting acordes a las solicitudes de contenido de los usuarios.

PODCASTING BASADO EN WEB SEMÁNTICA: UNA ESTRATEGIA PARA LA DIVULGACIÓN DE CONTENIDOS PARLAMENTARIOS

1. Introducción

La utilización de las tecnologías de la información y comunicaciones en la Administración Pública permite ofrecer al ciudadano nuevos servicios tales como la declaración de impuestos a través de Internet, oficina virtual de denuncias así como servicios de información y tramitación ofrecidos a través de los sitios Web de las Administraciones Públicas.

El objetivo del presente trabajo es mejorar la eficacia y difusión de contenidos parlamentarios a los ciudadanos mediante la utilización de la tecnología de podcasting y la incorporación del paradigma de Web Semántica. Asimismo, dicha solución puede ser extendida a otros ámbitos de la Administración Pública. En esta comunicación, describimos la estrategia y arquitectura para la obtención de sistemas podcasting basados en el conocimiento de contenido específico en contextos parlamentarios, basando nuestra visión general en el paradigma de Web Semántica [4] y en particular en la noción de tesauros. En concreto, proponemos la utilización de recursos basados en tesauros con la finalidad de estructurar, indexar y proporcionar contenido audiovisual, introduciendo una nueva aproximación que consta de cuatro procesos: (i) codificación y almacenamiento, (ii) catalogación, (iii) fragmentación, y (iv) entrega personalizada. Mediante la utilización de estos procesos, proporcionamos un nuevo sistema que permite distribuir los contenidos de las sesiones parlamentarias en función de las características específicas de contenido seleccionadas por el usuario.

2. Tecnología Podcasting y Web Semántica

En la actualidad, la tecnología de podcasting está siendo empleada como un medio efectivo y eficaz para la divulgación de contenido Web sindicado entre proveedores de contenido y usuarios finales.

Básicamente, las interacciones entre los proveedores y usuarios consisten en un proceso de dos etapas. Inicialmente, el proveedor de contenido hace accesible los archivos de audio y/o vídeo sobre un servidor Web disponible. Dichos archivos se conocen como un episodio de un podcast. Seguidamente, el proveedor de contenido reconoce la existencia de dichos archivos y la referencia en otro fichero conocido como feed. El feed representa una lista de episodios accesibles. Esta lista es generalmente publicada en formato RSS [1] (Really Simple Syndication) que proporciona además otras informaciones tales como: la fecha de publicación, títulos y algunas descripciones de texto de las series y cada uno de sus episodios. El feed es generalmente limitado a una lista resumida de los episodios más recientes. Por otra parte, el usuario utiliza un programa software denominado podcatcher con el propósito de determinar la ubicación de los episodios más recientes y descargarlos automáticamente sobre su ordenador personal o reproductor portátil. Los episodios descargados podrán ser reproducidos en cualquier momento que se desee y cuántas veces se desee.

Actualmente, la utilización de un canal de suscripción de entrega automática de nuevos contenidos es lo que distingue básicamente a un podcast respecto a una simple descarga o streaming en tiempo-real. De esta manera, la utilización de suscripciones permite a los usuarios coleccionar diferentes programas desde una diversidad de fuentes para escucharlos y visualizarlos offline en cualquier instante de tiempo y lugar.

El paradigma de Web Semántica representa una extensión de la Web actual basada en nuevos diseños de servicios y agentes inteligentes [6] capaces de procesar información con relaciones semánticas, apoyándose en la utilización de ontologías [5].

En este contexto, nuestro objetivo es triple: (i) por una parte, utilizar la tecnología de podcasting como mecanismo para la divulgación de manera eficaz de los contenidos de las sesiones parlamentarias a los ciudadanos, (ii) por otra parte, emplear el paradigma de Web Semántica, con el propósito de dotar de mayor significado a los vídeos que contienen los contenidos parlamentarios y así poder atender de forma más precisa las demandas de información de los ciudadanos acerca de los contenidos parlamentarios, proporcionando una respuesta apropiada y específica a los ciudadanos y ampliando de este modo la efectividad de los contenidos proporcionados y finalmente (iii) la automatización del sistema de podcasting mediante la utilización concreta de tesauros que abarca todos los ámbitos de actividad de las Comunidades Europeas.

En el siguiente apartado se describe la arquitectura y componentes del sistema podcasting para la divulgación de contenido audiovisual de sesiones parlamentarias y finalmente se muestran las principales conclusiones y líneas de trabajo futuras a seguir.

3. Divulgación de Contenido de Sesiones Parlamentarias

Nuestro objetivo es ofrecer un sistema que permita distribuir los contenidos de las sesiones parlamentarias del Parlamento de Canarias, según las preferencias indicadas por los ciudadanos que indican los temas que son de su interés, usando un modelo de vídeo podcasting.

Los contenidos de las sesiones que se incluyen son las plenarias, dado su carácter público, tal como se indica en el artículo 71 del Reglamento del Parlamento de Canarias [3]. Las sesiones plenarias incluyen a los participantes, debates e iniciativas y temas a ser discutidos.

Con este fin, presentamos una infraestructura que modifica el modelo habitual de vídeo podcasting, para que los ciudadanos reciban únicamente los fragmentos de vídeo que están relacionados con los temas que son de su interés. Con esta aproximación aportamos dos ventajas nuevas al ciudadano: (i) minimizar el tiempo de búsqueda para un tema específico y (ii) reducir el tamaño de la información (datos) que tiene que recuperar.

La Figura 1, muestra la arquitectura resultante de la solución propuesta. Esta arquitectura está compuesta de los siguientes procesos que se detallan a continuación:

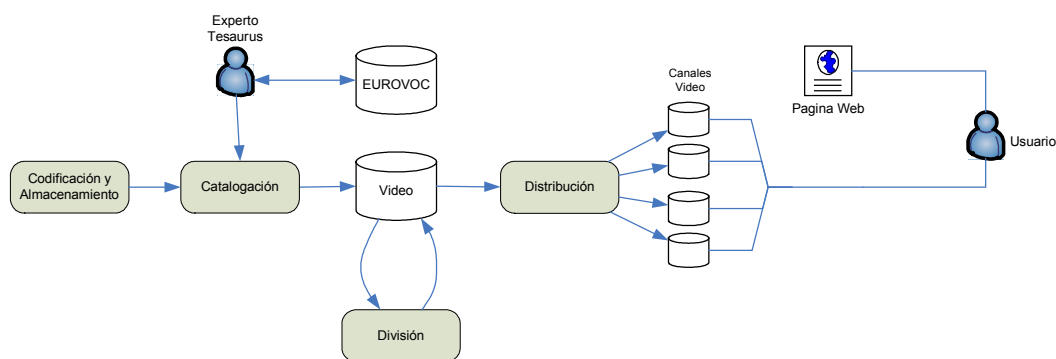


Figura 1. Arquitectura de podcasting para la divulgación de contenidos específicos de sesiones parlamentarias

3.1 Codificación y Almacenamiento

El contenido de las sesiones plenarias es digitalizado usando tarjetas de captura de audio y vídeo Osprey, en un sistema Windows Media, que se encarga de convertir la ingesta de audio y vídeo al formato que permitirá su procesado posterior.

El almacenamiento de todo el contenido multimedia se realiza en una plataforma SAN (storage-area network). La elección de dicha plataforma permite que el sistema ofrezca respuestas adecuadas para el manejo y gestión del vídeo. Al mismo tiempo, permite incrementar fácilmente la capacidad de almacenamiento que siempre resulta problemático en un proyecto relacionado con digitalización de vídeo.

Además de las sesiones plenarias, el módulo de codificación y almacenamiento, digitaliza las sesiones de comisiones, para lo cual se dispone de más equipos con tarjetas de captura de audio y vídeo, que se encargan de digitalizar cada sesión de comisión que se produce simultáneamente. No obstante, en el modelo que estamos describiendo sólo se tratarán posteriormente las sesiones plenarias. Debido al tipo de evento que se trata, es necesario asegurar una buena calidad de vídeo. En cuanto al audio, una calidad de radio es suficiente. Con este fin, conviene realizar un buen estudio sobre la calidad y tamaño del vídeo a escoger, así como los bitrates para la calidad del sonido debido al tipo de reproductores que están vinculados con el uso del podcasting actualmente.

3.2 Catalogación

El catalogador analiza el contenido audiovisual de la sesión con el fin de realizar en cada instante de tiempo las anotaciones correspondientes. Este proceso puede realizarse en tiempo real mientras está teniendo lugar la sesión plenaria, o bien, posteriormente cuando dicha sesión ya ha finalizado. Las anotaciones realizadas por el catalogador identifican los siguientes aspectos: nombre de la persona que interviene, iniciativa parlamentaria y motivo de la intervención. Al mismo tiempo el catalogador describe el tema que se está tratando en cada momento, mediante el uso de una serie de conceptos que se describen a través de un conjunto de descriptores o términos. Estos descriptores son obtenidos mediante la utilización de un tesoro (un modelo básico de ontología que consiste en una base de datos de términos jerárquicos con relaciones semánticas y genéricas). Todas las anotaciones realizadas por el catalogador se almacenan en una base de datos, donde al mismo tiempo, el sistema incluye información complementaria de la sesión (nombre, fecha, orden del día, etc.)

El catalogador sigue convenciones internacionales para catalogar la información audiovisual mediante la utilización del tesoro plurilingüe Eurovoc [2]. Actualmente, dicho tesoro abarca todos los ámbitos de actividad de las Comunidades Europeas, así como permite indexar los documentos en los sistemas de documentación de las instituciones europeas y de sus usuarios. El Parlamento Europeo, la Oficina de Publicaciones de las Comunidades Europeas, los parlamentos nacionales y regionales de Europa, las administraciones nacionales y determinadas organizaciones europeas se sirven actualmente de él. Las diferentes áreas de actividad abarcadas por el tesoro Eurovoc son: vida política, relaciones internacionales, comunidades europeas, derecho, vida económica, intercambios económicos y comerciales, asuntos financieros, asuntos sociales, educación y comunicación, ciencia, empresa y competencia, trabajo y empleo, transportes, medio ambiente, agricultura, silvicultura y pesca, sector agroalimentario, producción, tecnología e investigación, energía, industria, geografía y organizaciones internacionales.

Cada asunto es identificado mediante una etiqueta de comienzo y fin. Mediante la utilización del Eurovoc, el catalogador describe el tema tratado entre cada etiqueta de comienzo y fin, representando de manera inequívoca el concepto del asunto que se está tratando en cada momento.

Los conceptos que identifican los temas tratados, se describen mediante un conjunto de descriptores o términos del Eurovoc. Cada descriptor es representado mediante una clasificación jerárquica

de dos niveles que representa (i) un campo, identificado por dos dígitos numéricos y títulos mediante términos, p.e: 10 Comunidades Europeas y (ii) un micro tesoro, identificado por cuatro dígitos numéricos, p.e: 1011 Leyes Comunitarias, donde los dos primeros dígitos hacen referencia al ámbito que contiene el micro tesoro. Las relaciones semánticas entre los descriptores abarcan notas de ámbito (algunos descriptores pueden ser acompañados de notas, clarificando el significado del descriptor), relaciones entre sí del micro tesoro, relaciones de equivalencia, relaciones jerárquicas y relaciones asociativas.

Actualmente, el tesoro Eurovoc abarca 21 ámbitos diferentes, 127 micro tesoros, 6439 descriptores, 6448 relaciones jerárquicas y 3501 relaciones asociativas.

3.3 División

Tras el proceso de catalogación, las anotaciones realizadas por el catalogador permiten la identificación del comienzo y fin de los asuntos tratados.

Cada vídeo de cada sesión plenaria es dividido mediante un proceso automático, en vídeos más pequeños que tendrán el tamaño de la duración correspondiente a cada asunto tratado.

Una vez realizado el proceso de división, se realiza la asignación de cada fragmento de vídeo al canal correspondiente. Mediante la consulta de las anotaciones realizadas en la fase de catalogación, se recuperan los descriptores que corresponden a cada fragmento de vídeo. Por otro lado, la estructura jerárquica del tesoro, permite obtener el área de actividad al que pertenece cada descriptor. De esta forma, cada fragmento de vídeo de cada asunto tratado, es clasificado en un área de actividad del tesoro Eurovoc.

3.4 Distribución personalizada

Los canales de suscripción a los cuales los ciudadanos pueden sindicarse, se basan en las 21 diferentes áreas de actividad abarcadas por el tesoro Eurovoc.

Con el fin de que los ciudadanos conozcan los canales disponibles, se publica una lista en una página Web. Cada canal representa un feed, es decir, un fichero XML basado en el uso de la especificación RSS. Esta especificación es el formato usado en podcasting, permitiendo transformar la forma en que millones de personas consumen noticias e información. La creciente adopción de RSS, permite que cada vez se produzca con menor frecuencia la lectura de la información directamente desde una página Web.

En nuestro modelo de podcasting, el sistema genera diariamente el fichero RSS para cada canal, con el fin de reflejar la disponibilidad de los nuevos fragmentos de vídeo (episodios). En cada fichero se mantendrá la referencia a los últimos 10 episodios de vídeo disponibles para cada canal.

En este contexto, los ciudadanos pueden sindicarse a los canales en los que estén interesados. Una vez sindicado al canal, el programa (podcatcher) de cada ciudadano, determinará la localización de los episodios de vídeo más recientes para cada canal suscrito y automáticamente lo descargará al ordenador del usuario. El episodio de vídeo descargado podrá ser posteriormente visualizado o transferido a un dispositivo portátil.

4. Conclusiones y líneas futuras de trabajo

Este trabajo presenta nuestro enfoque de divulgación de contenidos de sesiones parlamentarias mediante la utilización de tecnologías de podcasting. La incorporación del paradigma de la Web Semántica permite ofrecer nuevas oportunidades así como prestaciones al ciudadano. En concreto, la utilización de tesoros permite representar los contenidos audiovisuales de las sesiones parlamentarias mediante la utilización de términos que guardan entre sí relaciones semánticas y genéricas, representando de manera unívoca, el contenido de éstos, con el fin de servir tanto para la catalogación, como para la recuperación de fragmentos específicos de las sesiones parlamentarias.

Nuestra línea de trabajo futuro estará centrada en extender el modelo propuesto para permitir a los ciudadanos la realización de búsquedas y preguntas sobre los temas tratados en las sesiones plenarias, así como la evaluación de nuevas técnicas de optimización que produzcan contenido de vídeo más acorde a las diferentes posibles audiencias.

5. Referencias

- [1] RSS 2.0 Specification. <http://blogs.law.harvard.edu/tech/rss>
- [2] The Office for Official Publications of the European Communities. Eurovoc Thesaurus. <http://europa.eu.int/celex/eurovoc/>
- [3] Reglamento del Parlamento de Canarias. <http://www.parcan.es/pub/reglamento.pdf>
- [4] W3C: Semantic Web. <http://www.w3.org/2001/sw/>
- [5] C. Brewster, K. O'Hara, S. Fuller, Y. Wilks, E. Franconi, M. A. Musen, J. Ellman, S. B. Shum. Knowledge Representation with Ontologies: The Present and Future. IEEE. Intelligent Systems. January/February 2004 [Vol. 19, No. 1] pp. 72-81
- [6] D. Fensel, W. Wahlster, H. Lieberman, J. Hendler. Spinning the Semantic Web. The MIT Press. February 2003.