Linked Data Methodologies for Managing Information about Television Content

Applying Linked Data principles in the OntoTV system, in order to improve the collection processes and the way television information is accessed.

José Luis Redondo-García, Vicente Botón-Fernández, Adolfo Lozano-Tello Universidad de Extremadura. Quercus Software Engineering Group Escuela Politécnica, Campus Universitario s/n, 10071

Cáceres, Spain
jluisred@unex.es, vboton@unex.es, alozano@unex.es

Abstract— Onto TV is a television information management system designed for improving the quality and quantity of the formation available in the current television platforms. In order to achieve this objective, Onto TV (1) collects the information offered by the broadcasters, (2) integrates it into a ontology-based data structure, (3) extracts extra data from alternative television sources, and (4) makes possible for the user to perform queries over the stored information.

This document shows the way Linked Data methodologies have been applied in OntoTV system, and the improvements in the data consumption and publication processes that have been obtained as result. By the one hand, the possibility of accessing to information available in the Web of Data has made possible to offer more complete descriptions about the programs, as well as more detailed guides than those obtained by using classic collection methods. On the other hand, as the information of the television programs and channels are published according to the Linked Data philosophy, it becomes available not only for OntoTV clients, but also for other agents able to access Linked Data resources, who could offer the viewer more fresh and innovative features.

Keywords-television; Linked Data; Onto TV; ontology; consuming; publishing.

I. Introducción

Hoy en día es cada vez mayor el número de plataformas y de canales de televisión disponibles, por lo que no resulta inmediato para el televidente decidir qué quiere ver en cada momento. Es cierto que algunos proveedores ofrecen información sobre los programas que emiten, pero esta información suele ser demasiado escasa, y no se ofrecen la posibilidad de ejecutar operaciones avanzadas sobre ella, como las recomendaciones de contenidos.

Por ello, la definición de un sistema gestor de información sobre contenidos televisivos que dé solución a estos inconvenientes puede resultar muy beneficioso para los televidentes y para su experiencia televisiva. Este sistema debe ofrecer un servicio de información más sencillo y universal, superior en funcionalidades a las soluciones existentes. Cuando la información sea escasa, debe acceder a recursos externos para completarla de forma rápida y transparente. De esta forma

se consigue un punto de acceso común a información sobre contenidos televisivos que los clientes pueden utilizar desde un móvil, un decodificador, o un ordenador personal. En investigaciones anteriores en esta misma línea [1], se propuso la creación del sistema OntoTV para dar respuesta a estas necesidades. Este sistema recolecta información sobre contenidos televisivos de varias fuentes y la representa por medio de ingeniería del conocimiento y ontologías.

Sin embargo aún existen inconvenientes relacionados con la forma en que OntoTV maneja la información. Por un lado, el sistema utiliza un tipo de componentes software, denominados "Crawlers", que recuperan información de ciertas fuentes que ofrecen su información de manera poco estructurada, como las páginas HTML. Este proceso normalmente conlleva altos costes computacionales, y obliga a que los algoritmos de lectura de cada fuente de datos se adapten exclusivamente a las particularidades de las mismas. Por otro lado, sólo los clientes compatibles con las especificaciones de OntoTV pueden acceder a la información recolectada.

En esta situación, la metodología Linked Data [2] para la publicación y consumo de datos está ganando en presencia e importancia en la Web. Se trata de una serie de principios para estructurar y enlazar datos, de forma que resultan más útiles y fáciles de reutilizar por terceros. Esta metodología está construida sobre estándares ampliamente utilizados en la Web, como URI y HTTP, por lo que la información compartida de esta forma se convierte en accesible tanto para humanos como para máquinas. Al final, este conjunto de datos de diferentes fuentes, enlazados y fácilmente accesibles, conforman lo que se conoce como Web de Datos.

El principal objetivo de esta investigación es aplicar la metodología Linked Data en OntoTV, para mejorar los procesos de recolección de información y la experiencia televisiva del usuario. Para ello se han diseñado componentes para el consumo de información televisiva de fuentes Linked Data, concretamente datos sobre películas. Además, la información televisiva que se almacena en la base de conocimiento se publicará ahora según principios de Linked Data, por lo que estará disponible en la Web de Datos para que muy diversos agentes puedan hacer uso de la misma.

II. EL SISTEMA ONTOTV

El sistema OntoTV (ONTOlogy-based management system for digital TeleVision) es un sistema gestor de datos sobre contenidos televisivos que permite al televidente acceder a la información sobre los programas que han sido, están siendo, o serán emitidos en las distintas plataformas digitales. Debido a que incorpora mecanismos adecuados de adquisición de datos, el sistema puede ofrecer al usuario descripciones detalladas de contenidos y le permite realizar operaciones avanzadas de búsqueda y recomendación. El sistema OntoTV fue presentado anteriormente en [3] donde se mostraron los funcionalidades básicas del mismo:

- Integrar toda la información posible sobre contenidos televisivos, incorporando mecanismos para recolectarla desde las diversas fuentes existentes.
- Representar los datos extraídos utilizando ontologías, de forma que sea posible realizar razonamientos e inferencias que generen nuevos conocimientos [4].
- Ejecutar sobre la base de conocimiento operaciones que sean atractivas para el usuario, como las búsquedas y las recomendaciones, con un alto grado de personalización.
- Permitir al usuario interactuar con el sistema. El terminal cliente envía peticiones para la ejecución de operaciones y recibe los resultados de las mismas. Tendrán acceso clientes con decodificadores basados en MHP, Google TV, Media Centers, etc.

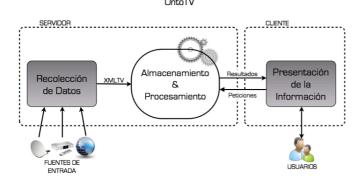


Figura 1. Esquema de los módulos principales de OntoTV

La Figura 1 muestra un esquema del sistema OntoTV. El módulo de "Almacenamiento y Procesamiento" está formado por la ontología de contenidos televisivos y por los distintos algoritmos de búsqueda y recomendación que se ejecutan sobre ella. Los módulos de "Recolección de Datos" y de "Presentación de la Información" serán descritos con mayor detalle a continuación, pues sobre ellos se aplicarán los principios Linked Data que pretenden mejorar el consumo y publicación de datos en el sistema.

A. Módulo de Recolección de Datos

Este módulo realiza la lectura directa de datos desde las fuentes contempladas en el sistema. El proceso consiste en interpretar el formato que utiliza una determinada entrada y

transformar la información leída al formato XMLTV que utiliza el sistema.

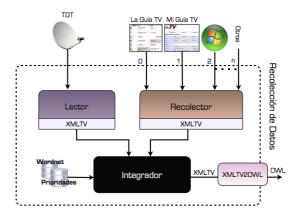


Figura 2. Módulo de Recolección de Datos en OntoTV

En la Figura 2 se pueden observar los distintos componentes que forman el módulo de recolección. Los de tipo "Lector" extraen los datos que los proveedores ofrecen en las plataformas de televisión digital soportadas por el sistema. Los componentes de tipo "Recolector" se conectan a servidores externos, normalmente utilizando el protocolo TCP/IP, para obtener otras guías de programación e información complementaria. Finalmente el componente denominado "Integrador" incluye en un solo archivo de datos toda la información en XMLTV procedente de los dos componentes anteriores. Los duplicados de descripciones acerca de un mismo programa en distintas fuentes son detectados y resueltos según mecanismos descritos en [2]. Siguiendo este esquema, OntoTV accede a las siguientes fuentes de datos televisivas:

- Información incluida en la señal de la TDT, mediante el procedimiento descrito en [1].
- Acceso a "La Guía TV" , un portal Web con información relativa a contenidos televisivos emitidos por los principales canales de televisión. Son necesarias conversiones de HTML a XMLTV.
- Acceso a "Mi Guía TV" ². Se trata de un portal Web de características parecidas al del caso anterior.
- Acceso a la guía de Windows Media Center. Microsoft ofrece guías de programación para los principales canales de televisión en España.

Como se puede observar, todas estas fuentes proporcionan información acerca de contenidos televisivos. El problema es que las estrategias de consumo de información utilizadas en cada caso son distintas: el acceso a la TDT se realiza interpretando tablas DVB-SI, la información de las paginas Web se extrae de ciertas etiquetas HTML, etc. Además de esta falta de uniformidad en los métodos de lectura, los procesos involucrados en la misma suelen ser computacionalmente costosos porque la información no se encuentra suficientemente estructurada.

¹ http://www.laguiatv.com/

² http://www.miguiatv.com/

B. Presentación de la Información

La arquitectura de tipo cliente-servidor que ha sido utilizada en OntoTV hace posible que distintos terminales de acceder funcionalidades televisión puedan a sus independientemente de las características particulares de los mismos. Esto resulta muy beneficioso hoy en día dada la enorme variedad de dispositivos disponibles en el mercado: decodificadores MHP, televisores compatibles con la plataforma Google TV, móviles con sistema operativo Android, etc. Para todas estas plataformas es posible desarrollar una aplicación cliente, denominada "OntoTV-Client", que realice las funciones propias de presentación de información al usuario. Las únicas premisas son tener conexión a Internet para la comunicación cliente-servidor, y disponer en cada plataforma de librerías específicas para la lectura de eventos de usuario, la generación de interfaces gráficas, y el intercambio de mensajes entre cliente y servidor.

Sin embargo, este intercambio de información se lleva a cabo utilizando un tipo de mensajes y una secuencia de comunicación que han sido fijadas de antemano y son únicas y exclusivas para el sistema OntoTV. Para que un cliente establezca comunicación válida con el servidor debe implementar ese conjunto de peticiones/respuestas particular.

Tabla 1. Intercambio de Mensajes en un Cliente de OntoTV

Tipo de Mensaje	Mensajes	
	Salida	Entrada
Gestión de contenidos	Solicitud de búsqueda de contenidos en base a ciertos criterios. Solicitud de descripción detallada de un contenido.	Lista de contenidos que satisfacen los criterios especificados. Descripción de un contenido en particular.
	Petición de guía de programación personalizada	Lista de contenidos acorde a los gustos del televidente
Datos de Usuario	Envío de eventos locales (pulsación de botones, navegación por menús, etc Envío de información contenida en el menú de preferencias explícitas	Perfil de usuario almacenado en el Servidor
Conexión con el Servidor	Solicitud de apertura de conexión. Solicitud de cierre de conexión.	Confirmación de conexión correcta. Confirmación de cierre de conexión.

La Tabla 1 recoge los mensajes más destacados que el cliente OntoTV envía y recibe del servidor. Para su implementación se ha utilizado el protocolo HTTP e intercambio de ficheros en formato XMLTV sobre TCP/IP.

El problema de esta aproximación es que, a pesar de conseguir independencia de la plataforma utilizada por el consumidor, éste debe implementar este conjunto específico de mensajes siempre, incluso cuando el agente no es propiamente de un cliente de OntoTV sino otra entidad que puntualmente necesita información televisiva. Por ejemplo, un portal Web que pretende mostrar información variada al usuario podrá consultar en OntoTV datos sobre horarios de emisión de programas, pero sólo si incorpora toda la lógica de comunicación propia de un cliente OntoTV.

III. APLICACIÓN DE METODOLOGÍAS LINKED DATA

El análisis del funcionamiento de OntoTV realizado anteriormente pone de manifiesto los problemas existentes en el mismo. Por un lado, los mecanismos tradicionales de extracción de información de las fuentes televisivas contempladas en el sistema se muestran poco eficientes debido a la heterogeneidad en los métodos de acceso y a la poca estructuración de la información. Por otro lado, sólo los clientes compatibles con OntoTV pueden acceder a la información que éste almacena. En este apartado se pretende solucionar estos inconvenientes mediante la aplicación de los principios de consumo y publicación de datos Linked Data [5], siguiendo la línea iniciada por otros sistemas televisivos que también utilizan tecnologías semánticas, como Notube [6], [7].

A. Consumo de Datos con Linked Data

En esta sección se muestra cómo se han incorporado estrategias de consumo de datos Linked Data en el módulo de "Recolección de Datos", para mejorar la cantidad de información sobre contenidos televisivos disponible en el mismo. Concretamente, se describe el funcionamiento del nuevo componente "LinkedData Movies" (ver Figura 3), que accede a recursos Linked Data, identifica información sobre determinadas películas, y completa las descripciones de las mismas en la guía de programación XMLTV.

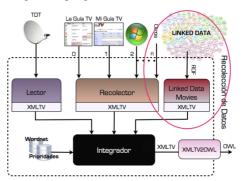


Figura 3. Módulo de Recolección extendido para acceso a Linked Data

1) Alternativas para el Acceso a la Web de Datos

Se han identificado varias alternativas para el acceso a información sobre contenidos televisivos en la Web de Datos, que se muestran a continuación:

- Acceder al dataset LinkedMDB³. Ejecutando de consultas en SPARQL sobre el punto de entrada que este almacén de datos proporciona, puede obtenerse información sobre películas. Sin embargo, y aunque LinkedMDB pretende ser para la Web de Datos lo que es hoy en día IMBD en la Web de Documentos, todavía no tiene entradas para numerosas películas.
- Implementar el método descrito en [5] que utiliza el patrón de consumo "Crawling". Consiste en crear un almacén de datos propio utilizando la herramienta Jena TDB ⁴, al que se va añadiendo la información recolectada por el Crawler de Linked Data DSpider⁵.

³ http://www.linkedmdb.org/

⁴ http://openjena.org/TDB/

⁵ http://code.google.com/p/ldspider/

La desventaja de este método es que tiene un alto coste computacional. Además, el proceso de recolección es lento y debe ser repetido periódicamente para asegurar que la información no está desactualizada.

 Acceder al mashup semántico SIG.MA. La ventaja de esta alternativa es que es posible acceder a información relevante, procedente de muy diversas fuentes semánticas, sin realizar costosos y lentos procesos de recolección. SIG.MA realiza frecuentes actualizaciones en sus datos por lo que la información sobre películas que se obtiene es lo suficientemente actual.

2) Funcionamiento del Componente LinkedData Movies

El componente "LinkedData Movies" ha sido codificado en el Java y lleva a cabo las siguientes acciones para extraer información sobre películas de la Web de Datos:

a) Conseguir descripciones en formato RDF. El mecanismo básico para acceder a Linked Data en la Web, es resolver URIs HTTP y recolectar así el RDF asociado. En el caso de SIG.MA, es necesario realizar una petición HTTP a "http://sig.ma/search?q=moviename", dónde "moviename" es una cadena que representa el nombre de la película. El Código 1 muestra cómo es posible recuperar documentos RDF utilizando la librería org.apache.commons.httpclient:

Código 1. SOLICITUD DE INFORMACIÓN EN FORMATO RDF

```
import org.apache.commons.httpclient.*;
   //Get Method
   HttpClient client = new HttpClient();
   HttpMethod method = new GetMethod(urlsigma + filmName);
   method.addRequestHeader("Accept", "application/rdf+xml");
   int responseCode = client.executeMethod(method);
   //Write RDF to FILE
   InputStream is = method.getResponseBodyAsStream();
   OutputStream os = new FileOutputStream(rdfFile);
   byte[] buffer = new byte[4096];
   for (int n; (n = is.read(buffer)) != -1;)
        os.write(buffer, 0, n);
```

b) Extraer información de los Ficheros RDF mediante consultas SPARQL. El fichero RDF que contiene la información sobre la película buscada está ya disponible en el sistema. Ahora se procede a extraer datos de él realizando consultas SPARQL. Para ello se ha utilizado la librería Jena ARQ⁶, disponible para Java.

```
Código 2. EJECUCIÓN DE CONSULTAS EN SPARQL
```

```
import com.hp.hpl.jena.query.*;
Model m;
m = ModelFactory.createMemModelMaker().createModel("");
model.read(in,null);
//Execute the Query
Query query = QueryFactory.create(stquery);
QueryExecution qe;
qe = QueryExecutionFactory.create(query, m);
ResultSet results = ge.execSelect();
```

El Código 2 ejecuta aquellas consultas SPARQL que vengan expresadas en la variable "stquery". En la Figura 4 se recoge un ejemplo de consulta que permite averiguar el director de una película accediendo a la propiedad "director" del vocabulario de SIG.MA, (http://sig.ma/property/). De forma similar, también puede extraerse información sobre el idioma de la película, su duración, actores principales, etc.

```
6 http://jena.sourceforge.net/ARQ/
```

```
PREFIX sigma: http://sig.ma/property/
PREFIX rdfs: http://www.w3.org/2000/01/rdf-schema#
SELECT Zdirector ?name
WHERE {
    ?film sigma:director ?director.
    ?director rdfs:label ?name.

| name
```

Figura 4. Consulta en SPARQL que averigua el director de la película

<http://dbpedia.org/resource/Ridley_Scott> | "Sir Ridley Scott"

c) Acceso a otros Datasets. La filosofia Linked Data se basa en navegar a través del conocimiento global. Por ello, si la información que el portal SIG.MA ofrece es insuficiente, es posible acceder a documentos alternativos siguiendo los enlaces RDF existentes en el documento. Por ejemplo, en la Figura 4 la URI correspondiente al director Ridley Scott hace referencia a un documento en el dataset DBpedia. Si se repite con esta URI el mismo proceso anterior, pueden recuperarse datos adicionales, como la fecha de nacimiento del director.

Analizando el procedimiento completo, quedan de manifiesto las ventajas derivadas de consumir información que se encuentra publicada en la Web de Datos respecto a los accesos tradicionales a fuentes de datos no estructuradas. El uso de URIs y del protocolo HTTP proporcionan un acceso a los distintos almacenes más universal. Así mismo, el hecho de que los datos se encuentren en formato RDF y estructurados según ciertos vocabularios (como el de SIG.MA) facilita notablemente la forma en que la información es interpretada.

B. Publicación de Datos con Linked Data

Como se ha señalado en el apartado 2.b, la única forma posible de acceder a la información que OntoTV almacena en su base de conocimiento es implementar una lógica específica de intercambio de mensajes fijada de antemano para el sistema. En esta sección se explican los cambios realizados en el sistema OntoTV para dar soporte a la publicación de datos sobre contenidos televisivos en Linked Data, de forma que cualquier agente en la Web de Datos pueda aprovecharlos.

1) Ontología de Dominio Televisivo

El primer paso para publicar datos utilizando la metodología Linked Data es elegir un vocabulario del dominio en base al cual poder representar la información televisiva. Anteriormente, OntoTV utilizaba la ontología propuesta por AVATAR [8]. Sin embargo para esta investigación ha sido sustituida por la ontología de Programas de la BBC (British Broadcasting Corporation). Esta organización ha creado este vocabulario haciendo uso de su amplia experiencia a la hora de ofrecer su parrilla televisiva en la Web con tecnologías semánticas. Estos antecedentes han llevado a considerar esta ontología como la más adecuada para representar programas y canales de televisión de una forma más estándar y por tanto más acorde a la filosofía Linked Data.

La ontología proporciona un vocabulario sencillo que incluye varias clases pertenecientes al dominio de contenidos televisivos y organismos de emisión. En la Figura 5 se aprecia el recuadro "Content" (Programme, Brand, Series, Episode), que contiene clases para describir tipos contenidos televisivos. En el recuadro "Medium" destacan la clase Channel para representar los medios de transmisión disponibles, y la clase

Broadcaster u organismo emisor. En el recuadro "Publishing" es importante la clase Version, que representa las distintas apariciones de un programa concreto en un canal y a una hora determinadas. Las clases dentro del recuadro "Temporal Annotations" no han sido utilizadas en esta investigación.

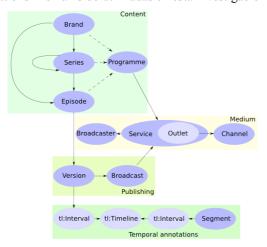


Figura 5. BBC Programmes (www.bbc.co.uk/ontologies/programmes)

2) Generando de Datos en RDF

En este apartado se describen los pasos que se llevan a cabo para generar las instancias de canales y programas de televisión en la ontología de la BBC, a partir de la información XMLTV procedente del Módulo de Recolección de Datos. El elemento software encargado de ejecutar esta transformación es XMLTV2OWL. Como puede comprobarse en la Figura 2, este componente está dentro del "Módulo de Recolección", pero sin embargo participa de forma determinante en el proceso de conseguir que esta información quede disponible en la Web de Datos, porque es el que genera las instancias RDF que posteriormente son publicadas. A continuación se describen más detalladamente cada una de las fases de este proceso:

- a) Paso 1. Cada elemento de tipo "<channel>" incluido en el fichero XMLTV es transformado en una instancia de la clase "Service" de la ontología de la BBC. Antes de añadir este nuevo individuo en la base de conocimiento, se comprueba que no existen colisiones por el hecho de que ese canal ya haya sido insertado en el sistema con anterioridad.
 - b) Paso2. Por cada elemento XMLTV "cramme>":
 - Se crea una instancia de tipo "Programme" en la ontología de datos televisivos, transformando los campos del formato XMLTV a sus correspondientes propiedades en la ontología de la BBC. Antes de insertar la instancia en la ontología se comprueba que ese programa no haya sido introducido anteriormente.
 - Se crea una instancia de la clase "Versión" en la ontología de la BBC que toma los atributos de programa "start" y "stop", y que se asocia la instancia de programa utilizada en el punto anterior. Antes de introducirla en la base de conocimiento, se comprueban nuevamente posibles colisiones entre individuos. Si existen coincidencias, se mantiene el que ha sido recolectado más recientemente.

3) Enlazando con otros Datasets

La filosofía Linked Data hace especial hincapié en la necesidad de establecer enlaces entre datos que estén relacionados [9], de forma que sea posible navegar por el conocimiento. Por ello, en OntoTV se lleva a cabo una fase consistente en tomar la información RDF disponible en la ontología de contenidos televisivos del sistema, y realizar sobre ella operaciones de alineamiento con instancias de otros almacenes, para que los datos televisivos de OntoTV queden enlazados con otros en la Web de Datos:

- Enlaces con el dataset de DBpedia: DPpedia es considerado el núcleo de la Web de Datos. Contiene información acerca de cualquier dominio, por lo que se ha convertido en un dataset de referencia. La detección de instancias coincidentes se ha realizado aplicando funciones de similitud léxica sobre los valores de los atributos de las clases "Service" y Programme", como por ejemplo "productor", "director", "actor", etc.
- Enlaces al dataset Geonames. Ciertos atributos de la ontología hacen referencia a lugares. En estos casos se comprueba si existen instancias geográficamente equivalentes en el dataset Geonames, que contiene una enorme cantidad de información territorial.
- Enlaces a LinkedMDB. Aunque actualmente tiene pocos registros, está llamado a ser el dataset de referencia sobre contenidos cinematográficos en la Web de Datos, por lo que se intentará identificar alineamientos de instancias "director", "actor" y "película" con esta fuente. Para detectarlas se aplicarán funciones de similitud textual en títulos.

El módulo de "Publicación en LD" (ver Figura 7) es el encargado de realizar esta búsqueda de instancias coincidentes en otros datasets. Como se observa en la Figura 6, los enlaces encontrados se expresan en forma de tripletas <owl:sameas>.

```
<rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"</pre>
  <owl:sameAs>http://dbpedia.org/page/RTVE</owl:sameAs>
</bbc_ont:Channel>
<bbc_ont:Programme rdf:about="http://purl.org/ontology/po/#Blade_Runne</pre>
  <bbc_ont:long_synopsis rdf:datatype="http://www.w3.org/2001/XMLSchen</pre>
 >Película de Ciencia Ficción que ha marcado un antes y un después er
  <bbc_ont:format>
    <bbc_ont:Format rdf:about="http://purl.org/ontology/po/#43-9"/>
 </bbc ont:format>
  <bbc ont:actor>
    <foaf:Agent rdf:about="http://purl.org/ontology/po/#Harrison_Ford"
    <owl:sameAs>http://dbpedia.org/page/Harrison_Ford</owl:sameAs>
    <owl:sameAs>http://data.linkedmdb.org/page/actor/755</owl:sameAs>
   <bbc_ont:place>
    <bc_ont:Place rdf:about="http://purl.org/ontology/po/#Place_11"/>
    <owl:sameAs>http://www.geonames.org/countries/US/united-states.rdf
 </bbc_ont:place>
```

Figura 6. Instancias de la Ontologia de la BBC en OntoTV

Para finalizar, es preciso indicar la existencia de ciertas herramientas de publicación de datos como Virtuoso Openlink⁷, que partiendo de las tripletas RDF disponibles proporcionan opciones de navegación en modo HTML y ejecución de consultas en SPARQL, aunque este aspecto no ha sido abordado en esta investigación.

⁷ http://virtuoso.openlinksw.com/

C. Esquema de Onto TV Resultante de Aplicar Linked Data

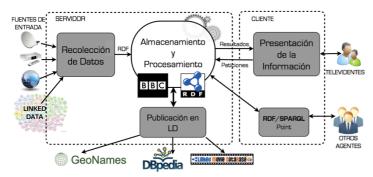


Figura 7. BBC Programmes (www.bbc.co.uk/ontologies/programmes)

En la Figura 6 se muestran los cambios producidos en OntoTV tras la aplicación de metodologías Linked Data. Por un lado, el módulo de recolección añade una nueva fuente: la propia Web de Datos. Por otro, el módulo de almacenamiento ahora contiene información en RDF correctamente enlazada con otros almacenes y representada según la ontología de la BBC. En cuanto a la presentación, no sólo los terminales cliente tienen acceso a los datos, sino que otros agentes que consuman Linked Data también pueden beneficiarse de ellos.

IV. CONCLUSIONES

En la actualidad los usuarios deben realizar un considerable esfuerzo por encontrar, acceder y comparar la gran variedad de información sobre contenidos televisivos disponible. El sistema OntoTV fue diseñado para dar solución a este problema, utilizando técnicas de recolección y almacenamiento basadas en ontologías. Sin embargo, las fuentes de datos a las que accede tienen naturaleza no estructurada, por lo que las funciones de recolección deben ser implementadas independientemente para cada una de ellas. Además, sólo los clientes que son compatibles con OntoTV pueden disponer de esta información televisiva recolectada.

En este trabajo se describe cómo se han aplicado principios de Linked Data al sistema OntoTV para solucionar estos inconvenientes. Por un lado, se ha conseguido acceder a recursos Linked Data para enriquecer la información disponible sobre películas, y por otro se ha diseñado un mecanismo para la publicación de canales y contenidos de televisión.

Referente al **consumo** de datos, se ha comprobado que los datos Linked Data recolectados han mejorado la escasa información enviada por los proveedores. El hecho de que las fuentes de datos consultadas utilicen principios Linked Data, conlleva la posibilidad de acceder a una información más estructurada y que presenta enlaces semánticos entre conceptos que no están presentes en los clásicos enlaces HTML. De esta forma, resulta sencillo extraer la información útil, no sólo para el caso de descripciones de películas, sino también para otros tipos de contenidos. También, la decisión de acceder a una fuente que ya integra automáticamente muchas otras, como es el caso del portal SIG.MA, ha propocionado ventajas sobre las estrategias de crawling en local o las consultas al punto de acceso SPARQL de algunos datasets, como LinkedMDB. La

información procede de diversas fuentes, por lo que es posible encontrar descripciones de películas para casi cualquier título.

En cuanto a la **publicación** de información, ahora los datos pueden ser utilizados no sólo por clientes de OntoTV, sino por cualquier otro agente que consuma datos Linked Data, sin necesidad de tener que implementar una lógica de intercambio de mensajes particular ni interpretar formatos específicos como XMLTV. También, la decisión de utilizar una ontología que está ya ampliamente consensuada en el dominio televisivo como la de la BBC, ha resultado muy adecuada para que la información recolectada por el sistema en OntoTV quede disponible en la Web de Datos de la forma más global posible.

A pesar de las mejoras obtenidas, es necesario seguir mejorando el proceso de transformación de los datos XMLTV recolectados a instancias de la ontología, incorporando mejores mecanismos de protección frente a errores, detectando colisiones al insertar instancias de un mismo programa en el modelo ontológico, y perfeccionando los algoritmos de alineamiento de información con DBpedia y LinkedMDB, que hasta ahora sólo emplean comparaciones léxicas simples.

En conclusión, la aplicación de metodologías Linked Data resulta muy beneficiosa para mejorar el rendimiento de sistemas que, como OntoTV, consumen y publican datos. Con estas nuevas estrategias de gestión de información sobre contenidos televisivos, los usuarios tendrán acceso a una información más precisa, completa y útil.

AGRADECIMIENTOS

Este trabajo está patrocinado por la beca de investigación PD10006 de la Junta de Extremadura, por el proyecto TIN2011-27340 del Ministerio de Ciencia e Innovación, y por el Fondo Europeo de Desarrollo Regional (FEDER).

REFERENCES

- Redondo-Garcia, J.L, Valiente-Rocha, P., Lozano-Tello, A.: Ontologybased system for content management in Digital Television. CISTI, pp. 277–283, (2010).
- [2] Berners-Lee, T.: Linked Data. International Journal on Semantic Web and Information Systems, vol. 4, no. 2, W3C (2006).
- [3] Redondo-Garcia, J.L., Lozano-Tello, A.: Recolección de Datos sobre Contenidos Televisivos en el Sistema OntoTV. CISTI, Accepted in press (2011).
- [4] Gomez-Perez, A., Corcho, O., and Fernandez-Lopez, M.: Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web. Springer-Verlag, New York (2004).
- [5] Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool (2011).
- [6] Schopman, B., Brickley, D., Aroyo, L., van Aart, C., Buser, V., Siebes, R.: NoTube: making the Web part of personalised TV. Proceedings of the WebSci10 (2010).
- [7] Buser, V.: NoTube: experimenting with Linked Data to improve user experience, Summer School on Multimedia Semantics, Amsterdam, 3 September 2010
- [8] Y. Blanco-Fernández, et al, "Exploiting synergies between semantic reasoning and personalization strategies in intelligent recommender systems" Journal of Systems and Software, vol.81, pp. 2371-2385, 2008.
- [9] Volz, J., Bizer, T., Gaedke, M., Kobilarov, G.: Discovering and Maintaining Links on the Web of Data. International Semantic Web Conference (ISWC2009), Westfields, USA, pp. 650—665, (2009).