

Integración de múltiples marcados en documentos XML

Luis Jesús Arévalo Rosado, Antonio Polo Márquez,
Juan Carlos Manzano Pérez, Miryam Salas Sánchez

*Departamento de Informática. Universidad de Extremadura*¹
e-mail: { [ljarevalo](mailto:ljarevalo@unex.es), [polo](mailto:polo@unex.es), [juancman](mailto:juancman@unex.es), [miryam](mailto:miryam@unex.es) }@unex.es

Resumen. La regla de anidamiento de etiquetas de marcado en XML provoca, en muchas ocasiones, que distintas interpretaciones de un mismo texto no puedan representarse en un mismo documento. Esta situación conduce a la duplicación del texto en diferentes ficheros, cada uno con el marcado apropiado a cada interpretación. Esto origina, no sólo problemas de replicación, sino la dificultad de no disponer de una única representación que integre todos los marcados. En este trabajo se presenta una técnica para resolver este problema que denominaremos *metamarcado múltiple*. Esta técnica consiste en almacenar la información de cada marca, sea de inicio o final, en un solo elemento. Para ilustrarla, se presenta un problema de múltiple marcado incompatible y se muestran las técnicas para integrar todos los marcados en un único fichero y cómo obtener a partir de él cada marcado independiente mediante transformaciones XSL. Esta solución permite una integración de conocimiento múltiple del documento, que se puede aplicar en problemas como versionado y evolución de documentos.

Palabras claves: Metamarcado, XML, marcado múltiple, anidamiento de etiquetas

1 Introducción

El estándar XML [1] se ha consolidado como una herramienta para el intercambio y representación de la información debido a la flexibilidad que aporta, tanto para la representación de información estructurada como semi-estructurada, así como por las facilidades de manipulación y conversión que presentan los documentos basados en dicho estándar. Esta tecnología se complementa con otras herramientas [1] que proporcionan nuevas funcionalidades a los documentos, como pueden ser las hojas de transformación XSL, consultas sobre la información almacenada mediante XQuery y XPath, vinculación de documentos con XLink y XPointer y la aparición de nuevos lenguajes de definición de datos como XSD o RDF.

La especificación de documentos XML obliga a un anidamiento correcto de las etiquetas. En determinadas ocasiones no se puede cumplir dicho anidamiento. Esta situación se produce, por ejemplo, cuando distintas interpretaciones de un mismo texto desean ser representadas dentro de un único documento. Si cada interpretación se re-

¹ Trabajo financiado por la Junta de Extremadura (Consejería de Educación, Ciencia y Tecnología) y el Fondo Social Europeo (Proyecto IPR00A057)

fleja mediante un marcado específico, la combinación de todos estos marcados en el mismo documento puede provocar anidamientos incorrectos de etiquetas. Actualmente esto se resuelve duplicando el texto en diferentes ficheros, cada uno de ellos con el marcado apropiado a cada interpretación. Esta solución origina duplicados del contenido y además no permite disponer de una única representación que integre todos los marcados en el mismo documento.

Este artículo comienza mostrando un ejemplo que ilustra el problema expuesto anteriormente. En segundo lugar se revisan las soluciones actuales. A continuación se presenta una técnica que permite transformar un documento XML en otro equivalente. Esta técnica se basa en el uso del metamarcado [3] y consiste en almacenar el estado de cada una de las marcas del documento origen en un solo elemento. En cuarto lugar se muestra que, mediante el uso de esta técnica, se pueden realizar marcados múltiples a un texto ajustándose a la especificación XML. Finalmente se presentan las conclusiones y trabajos futuros.

2 Motivación

Imaginemos que deseamos realizar una adaptación cinematográfica de la novela “El ingenioso hidalgo Don Quijote de la Mancha”, de Cervantes y que disponemos de una adaptación de los diálogos de la obra en formato XML. Un pequeño fragmento de dicha adaptación se muestra en la figura 1.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<DOCUMENTO>
<TÍTULO>¿Molinos o Gigantes?</TÍTULO>
<INTERVENCIÓN personaje="Sancho">¿Qué gigantes?
</INTERVENCIÓN>
<INTERVENCIÓN personaje="Quijote">Aquellos que allí ves de los brazos largos que
los suelen tener algunos de casi dos leguas.
</INTERVENCIÓN>
<INTERVENCIÓN personaje="Sancho">Mire vuestra merced, que aquellos que allí se
parecen no son gigantes, sino molinos de viento y lo que en ellos parecen brazos son
las aspas, que, volteadas del viento, hacen andar la piedra del molino.
</INTERVENCIÓN>
<INTERVENCIÓN personaje="Quijote">Bien parece que no estás cursado en esto de
las aventuras: ellos son gigantes; y si tienes miedo, quítate de ahí, y ponte en oración
en el espacio que yo voy a entrar con ellos en fiera y desigual batalla.
</INTERVENCIÓN>
.....
</DOCUMENTO>
```

Figura 1. Adaptación de la novela “El Quijote” en formato XML

Este diálogo debe completarse con nueva información para ser adaptado sin problemas al cine. Por ejemplo, es necesario incluir información acerca de la música que deberá sonar en cada secuencia de la película, detallar los movimientos de los actores, seleccionar las cámaras que filmarán cada momento exacto, etc. Se pretende que toda esta nueva información pueda ser incluida en el documento sin que el contenido de éste se vea afectado. Para el desarrollo de este artículo nos centraremos en el texto y los planos de cámara.

El problema con el que nos encontramos es sincronizar el texto con los planos que deben mostrarse en cada momento. Este problema podría ser resuelto mediante el etiquetado del documento con marcas de planos de forma que, en cada instante y a medida que avanzásemos por el documento, nos indicasen el plano exacto que debe visualizarse. Un ejemplo de este marcado se muestra en la figura 2, donde el significado de cada etiqueta “Plano” indica que esa toma dura mientras se interpreta el texto que anida y cada etiqueta “intervención” indica que personaje interpreta dicho texto..

```

.....
<PLANO id="toma10-g1" descripción="Plano general">
<INTERVENCIÓN personaje="Sancho">¿Qué gigantes?
</INTERVENCIÓN>
<INTERVENCIÓN personaje="Quijote">Aquellos que allí ves de los<PLANO>
<PLANO id="toma24" descripción="Visión de los gigantes">brazos largos que los
suelen tener algunos de casi dos leguas
</INTERVENCIÓN>
<INTERVENCIÓN personaje="Sancho">Mire vuestra merced, que aquellos que allí se
parecen no son gigantes, sino molinos de viento<PLANO><PLANO id="toma48" des-
cripción="Panorámica de los molinos">y lo que en ellos parecen brazos son las as-
pas, que, volteadas del viento, hacen andar la piedra del molino
</INTERVENCIÓN>
<INTERVENCIÓN personaje="Quijote"> Bien parece que no estás cursado en esto de
las aventuras: <PLANO><PLANO id="toma24-g2" descripción="Plano general">
ellos son gigantes; y si tienes miedo, quítate de ahí, y ponte en oración en el espacio
que yo voy a entrar con ellos en fiera y desigual batalla.
</INTERVENCIÓN></PLANO>
.....

```

Figura 2. Documento XML con múltiple marcado

La definición de estas nuevas etiquetas no debe afectar al contenido del documento y debe permitir la recuperación del texto, bien por las marcas de los planos o por las etiquetas de diálogo, tal y como se refleja en la figura 3.

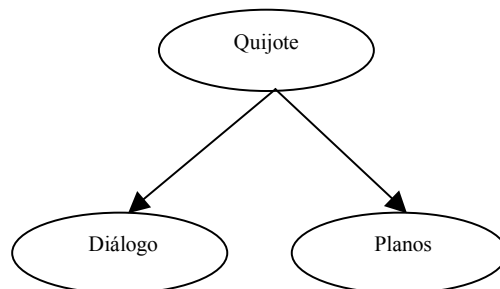


Figura 3. Distintas interpretaciones del documento “El Quijote”

Sin embargo el etiquetado de la figura 2 produce un documento XML que no está bien formado debido al anidamiento incorrecto de sus etiquetas. Una posible solución a este problema sería duplicar el texto en diferentes ficheros, cada uno con el marcado apropiado a cada interpretación. Esto origina un problema de replicación y no permite disponer de una única representación que integre todos los marcados en un mismo do-

cumento. A continuación se resumen brevemente algunas propuestas para resolver este problema.

3 Estudios anteriores

Los documentos XML se encuentran formados por un conjunto de etiquetas anidadas unas con otras. En [5] se definen los distintos tipos de anidamientos que se pueden producir entre dos marcas, centrándose en aquellas situaciones en las cuales se produce solapamiento. La necesidad de poder almacenar en un mismo documento distintos marcados a un texto ha sido estudiado por el grupo de trabajo que definió TEI (Text Encoding Initiative) [2, 3] que está basado en SGML[4] y que propone las siguientes soluciones mediante TEI:

La primera propuesta consiste en la utilización de la propiedad `CONCUR` de SGML[4]. Ésta se basa en la distinción de distintos subconjuntos de etiquetas para cada interpretación, y cada vez que se reconoce el documento sólo se considera una de ellas.

La segunda propuesta consiste en la fragmentación del contenido de cada etiqueta introduciendo nuevas etiquetas que consigan que el documento esté bien formado en aquellos casos en los que se produzca solapamientos. De forma que considerando un tipo de marcado se puede reconstruir el contenido de la etiqueta inicial. Mediante esta solución se han desarrollado varios proyectos [9, 10].

La tercera solución se basa en la definición de elementos vacíos. La idea consiste en marcar los límites entre los elementos en una estructura no anidada mediante etiquetas vacías. En este caso las marcas no encierran ningún contenido de texto, solamente marcan hitos o puntos en el documento. Una implementación en XML de esta solución se desarrolla en este trabajo.

Existen diversos proyectos basados en dicha propuesta. Así `Mecs/TexMecs` [6] definen un nuevo lenguaje sintáctico desarrollando con él un sistema para el soporte de múltiple marcado basado en elementos vacíos. La principal desventaja de esta solución se encuentra que solamente dispone del desarrollo teórico. En el caso de su implementación supondría que no existe ninguna relación entre el lenguaje creado y XML, imposibilitando una aplicación común.

`BUVH` [7] realiza una implementación basada en elementos vacíos en XML. La solución que propone consiste en la definición de *elementos atómicos* (etiqueta) donde almacena la información de cada una de las interpretaciones. Este elemento atómico realiza un marcado a nivel de palabra incluyendo en cada etiqueta como atributos el valor XPath de cada una de las etiquetas del documento origen (interpretación). La principal ventaja que aporta es la posibilidad de realizar consultas basadas en varias interpretaciones debido a que incorpora en el elemento atómico información de todas las interpretaciones.

Existen otros proyectos `Encoding Old Norse-Icelandic Primary`[8] y `Women Writers Project`[10] que utilizan los elementos vacíos como herramientas para solucionar la definición de múltiples interpretaciones mediante el lenguaje de marcado TEI.

Las propuestas basadas en elementos vacíos fueron ideadas para su utilización con lenguajes de marcado literario como TEI, donde el solapamiento de etiquetas es muy

frecuente. Principalmente han sido utilizadas por lingüistas para etiquetar el texto con información del análisis sintáctico, morfológico y semántico. Un ejemplo claro de su uso sería incluir en un documento etiquetas que representen el análisis del texto (sujeto, verbo, predicado, ...) frente a las indicaciones de su representación impresa (página, línea, verso, ...).

BUVH [7] es la única propuesta basada en XML y su implementación sigue enfocada al almacenamiento de las distintas estructuras sintácticas del texto. El inconveniente de esta solución se produce porque el elemento atómico se define a nivel de palabra implicando un incremento desorbitado de la longitud del documento metamarcado. Además, en la solución que se propone en BUHV, los atributos de cada una de las interpretaciones no son representados en el documento base, produciéndose una pérdida de información. En el resto del artículo proponemos una solución que resuelve estos inconvenientes.

4 Metamarcado

La solución que proponemos para el problema presentado anteriormente se basa en sustituir cada etiqueta de inicio o fin por un elemento que la representa, que denominaremos *metamarca*. Si tomamos un documento de texto inicial (T) y le añadimos un conjunto de etiquetas (E) conforme a la especificación XML, se consigue un documento XML (D) que representa una interpretación del texto inicial (T) y que expresaremos mediante la simbología $D = E + T$.

Todo documento XML se puede transformar en otro equivalente convirtiendo cada etiqueta en una metamarca. Esta transformación se denominará metamarcado y se define mediante la siguiente función M.

$$M : \chi \rightarrow \chi' = M(\chi) \subseteq \chi$$

$$\forall D \in \chi, D = E + T \mapsto M(D) = D' = E' + T$$

donde:

- χ es la clase de todos los documentos XML bien formados.
- **E** es el conjunto de marcas de inicio y final que definen el esquema original del documento.
- **E'** son las metamarcas obtenidas a partir del proceso de transformación y que siguen la definición de un esquema, que denominaremos *esquema base*.

Este esquema base común para documentos metamarcados se muestra en la figura 4. En él, existen dos elementos que representan las marcas de apertura y de cierre de un documento XML:

- oe: Elemento utilizado para representar una etiqueta de inicio.
- ce: Elemento utilizado para representar una etiqueta de cierre.

Para cada una de estas metamarcas sólo es necesario almacenar el nombre de la marca (atributo *nm*) para completar la información de una etiqueta de apertura o cierre. Estas metamarcas se definen como elementos vacíos en el documento excepto cuando una etiqueta de inicio tiene atributos, en cuyo caso el elemento “oe” puede in-

cluir varios elementos "at", donde cada uno recoge la información de un atributo utilizado en la marca de apertura original, indicando su nombre (atributo *nm*) y el valor asignado (atributo *value*).

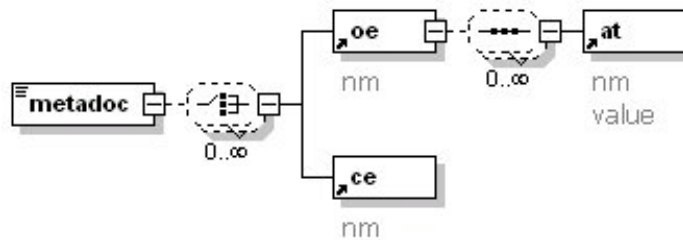


Figura 4. Esquema base de un documento metamarcado

La representación del documento de la figura 1 mediante metamarcado se muestra en la figura 5.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<MM:metadoc xmlns:MM=http://www.unex.es/2002/MMBeta
  xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
  xsi:schemaLocation="http://www.unex.es/2002/MMBeta metaschema.xsd">
  <MM:oe nm="DOCUMENTO"/>
  <MM:oe nm="TÍTULO"/> ¿Molinos o Gigantes? <MM:ce nm="TÍTULO"/>
  <MM:oe nm="INTERVENCIÓN">
    <MM:at nm="personaje" value="Sancho"/>
  </MM:oe> ¿Qué gigantes?
  <MM:ce nm="INTERVENCIÓN"/>
  <MM:oe nm="INTERVENCIÓN">
    <MM:at nm="personaje" value="Quijote"/>
  </MM:oe> Aquellos que allí ves de los brazos largos que los suelen tener algunos de
  casi dos leguas
  <MM:ce nm="INTERVENCIÓN"/>
  <MM:oe nm="INTERVENCIÓN">
    <MM:at nm="personaje" value="Sancho"/>
  </MM:oe> Mire vuestra merced, que aquellos que allí se parecen no son gigantes, sino
  molinos de viento y lo que en ellos parecen brazos son las aspas, que, volteadas del
  viento, hacen andar la piedra del molino
  <MM:ce nm="INTERVENCIÓN"/>
  <MM:oe nm="INTERVENCIÓN">
    <MM:at nm="personaje" value="Quijote"/>
  </MM:oe> Bien parece que no estás cursado en esto de las aventuras: ellos son gigan-
  tes; y si tienes miedo, quítate de ahí, y ponte en oración en el espacio que yo voy a en-
  trar con ellos en fiera y desigual batalla.
  <MM:ce nm="INTERVENCIÓN"/>
  .....
  <MM:ce nm="DOCUMENTO"/>
</MM:metadoc>
```

Figura 5. Documento Metamarcado

Si observamos la figura 5 se puede comprobar que ninguno de los elementos contiene texto del documento original debido a que las etiquetas están definidas como vacías. Esto permite añadir cualquier marca al documento metamarcado sin que exista

conflicto de anidamiento. La ubicación de cada metamarca en el documento debe ser la misma que la de la etiqueta a la que sustituye en el documento original. A partir del documento metamarcado se puede obtener el documento original mediante la definición de una función inversa M^{-1} que sustituya cada elemento por la marca que representa, de modo que:

$$M^{-1} : \mathcal{X}' \rightarrow \mathcal{X}$$

$$\forall D' \in \mathcal{X}' \rightarrow M^{-1}(D') = E + T = D$$

5 Definición de múltiples marcados o dimensiones

Como se ha indicado anteriormente, la característica principal de un documento metamarcado es que la adición de cualquier marca no va a producir conflictos de anidamiento. Esto puede ser utilizado para combinar en un único fichero distintas interpretaciones D_i sobre el mismo texto T , representadas por diferentes marcados E_i de la forma $D_i = T + E_i$. Para ello aplicamos la función de metamarcado a cada interpretación y combinamos todas las metamarcas en un solo documento obteniendo un documento D' representado $D' = E' + T$. La definición de “n” interpretaciones de T sería:

$$T \left\{ \begin{array}{l} D_1 = T + E_1 \\ D_2 = T + E_2 \\ \vdots \\ D_n = T + E_n \end{array} \right\} \xrightarrow{M} \left\{ \begin{array}{l} T + E'_1 = D'_1 \\ T + E'_2 = D'_2 \\ \vdots \\ T + E'_n = D'_n \end{array} \right\} T + \cup_{1 \leq i \leq n} E'_i = T + E' = D' \xrightarrow{M^{-1} / \text{dim}_i} T + E_i = D_i$$

Mediante la función de fusión “ \cup ” las metamarcas E'_i procedentes de cada interpretación D_i se incluirán en un único documento D' . Cada una de las interpretaciones se distinguirán de forma única mediante un identificador que denominaremos **dimensión**. Por tanto, para almacenar la información de los distintos etiquetados es necesario modificar el esquema base representado en la figura 4. Un nuevo atributo, denominado “dim”, ha sido añadido a las tres metaetiquetas anteriormente expuestas (oe, ce, at). Este atributo representará para qué dimensión o dimensiones es válida esta definición de etiqueta. Además incorporaremos una nueva metamarca llamada “dimension” que almacene la información de cada dimensión definida sobre el texto. Este elemento contendrá un atributo identificador e información adicional sobre la dimensión, como un nombre descriptivo y, en caso de estar disponible, el esquema original que la define.

En la figura 6 aparece un fragmento del documento metamarcado que incorpora las dos dimensiones (diálogo y planos) tratadas en nuestro ejemplo. Los dos valores que puede tomar el atributo “dim” son “planos” ó “diálogo”. Cuando la etiqueta “dim” no se encuentre definida representa validez para todas las dimensiones.

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<MM:metadoc xmlns:MM="http://www.unex.es"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.unex.es metaschema.xsd">
  <MM:dimension id="v1" desc="Diálogo" doc="dialogo.xsd"/>
  <MM:dimension id="v2" desc="Planos" doc="planos.xsd"/>
  <MM:oe nm="DOCUMENTO"/>
  <MM:oe nm="TÍTULO"/> ¿Molinos o Gigantes? <MM:ce nm="TÍTULO"/>
  <MM:oe nm="PLANO" dim="v2">
    <MM:at nm="id" dim="v2" value="toma10"/>
    <MM:at nm="descripción" dim="v2" value="Plano general"/>
  </MM:oe>
  <MM:oe nm="INTERVENCIÓN" dim="v1">
    <MM:at nm="personaje" dim="v1" value="Sancho"/>
  </MM:oe> ¿Qué gigantes?
  <MM:ce nm="INTERVENCIÓN" dim="v1"/>
  <MM:oe nm="INTERVENCIÓN" dim="v1">
    <MM:at nm="personaje" value="Quijote" dim="v1"/>
  </MM:oe> Aquellos que allí ves de los
  <MM:ce nm="PLANO" dim="v2"/>
  <MM:oe nm="PLANO" dim="v2">
    <MM:at nm="id" dim="v2" value="toma24"/>
    <MM:at nm="descripción" dim="v2" value="Visión de los gigantes"/>
  </MM:oe> brazos largos que los suelen tener algunos de casi dos leguas
  <MM:ce nm="INTERVENCIÓN" dim="v1"/>
  <MM:oe nm="INTERVENCIÓN" dim="v1">
    <MM:at nm="personaje" value="Sancho" dim="v1"/>
  </MM:oe> Mire vuestra merced, que aquellos que allí se parecen no son gigantes, sino
  molinos de viento
  <MM:ce nm="PLANO" dim="v2"/>
  <MM:oe nm="PLANO" dim="v2">
    <MM:at nm="id" dim="v2" value="toma48"/>
    <MM:at nm="descripción" dim="v2" value="Panorámica de los molinos"/>
  </MM:oe> y lo que en ellos parecen brazos son las aspas, que, volteadas del vien-to,
  hacen andar la piedra del molino
  <MM:ce nm="INTERVENCIÓN" dim="v1"/>
  .....
  <MM:ce nm="DOCUMENTO"/>
</MM:metadoc>

```

Figura 6. Documento metamarcado con distintas dimensiones

5.1 Validación y recuperación de documentos metamarcados

La validación del documento metamarcado consistirá, por una parte, en comprobar que es válido con respecto a la definición del esquema base y, por otra, que para cada dimensión esté bien formado y sea válido con respecto al esquema original. La validación con respecto al esquema base es trivial, pues mediante cualquier procesador de documentos XML podrá comprobarse.

El principal problema es poder validar los esquemas originales, o lo que es lo mismo, comprobar que las metamarcas de cada dimensión “i” definen correctamente la interpretación. Para conseguir esta validación se eliminarán todas las metamarcas que

no sean de la dimensión “i” y se comprobará que dicho documento D_i está bien formado y es válido con respecto a su esquema original.

Para la recuperación de las distintas marcas solamente es necesario definir una hoja de transformación XSL que será la encargada de generar cada uno de los documentos con sus correspondientes marcas. El funcionamiento de esta hoja de transformación consiste en recorrer el documento metamarcado y recuperar todo el texto y sólo aquellas etiquetas que pertenezcan a la dimensión solicitada. Por ejemplo en la figura 6 se podría seleccionar la dimensión v1 y se obtendría la interpretación del texto sólo con las marcas del diálogo. Gráficamente el proceso completo de validación y recuperación de un documento metamarcado se puede visualizar en la figura 7.

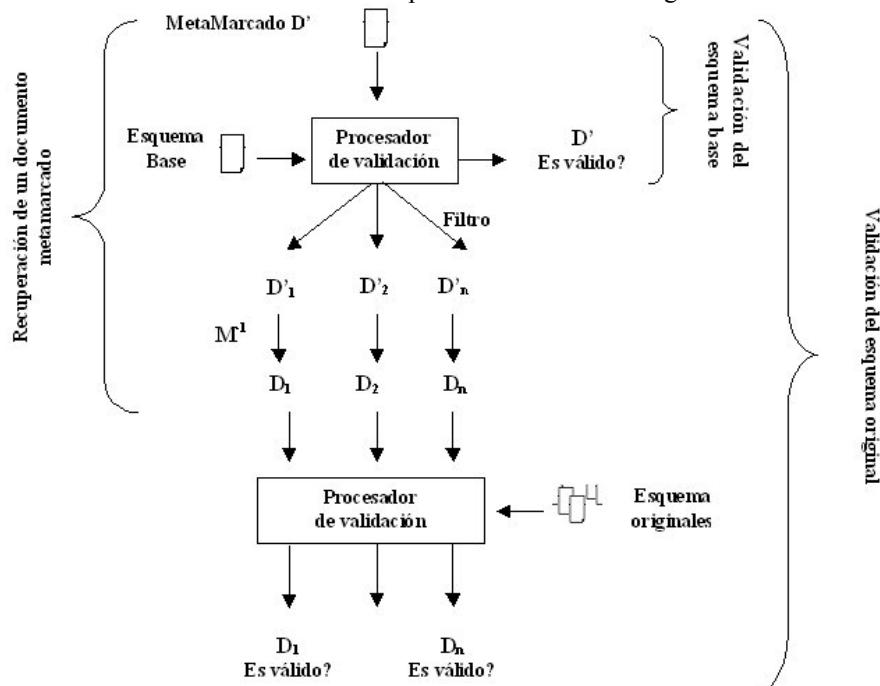


Figura 7. Validación y recuperación de las distintas interpretaciones de un documento metamarcado

6 Conclusiones y trabajos futuros

La utilización de múltiples marcas en un documento XML puede ser de gran importancia en el campo de las bibliotecas digitales. Por ejemplo, en aquellas situaciones en las cuales cada miembro de un grupo de trabajo colaborativo realiza distintas anotaciones sobre un mismo documento y posteriormente se desea recuperar cada una de sus interpretaciones. Sin embargo la tecnología XML no permite la integración de múltiples marcas en un documento XML.

En este trabajo se ha presentado una técnica basada en XML para la integración de distintos marcados de un documento sin tener que duplicar el texto con cada una de

sus interpretaciones. Esta técnica consiste en representar un documento XML como otro equivalente. Mediante esta sencilla acción se puede añadir cualquier marca al documento con la garantía de que su anidamiento es correcto.

Las ventajas de utilizar la técnica del metamarcado para definir diferentes interpretaciones en un mismo documento son:

- El almacenamiento de las distintas dimensiones se realiza en un único documento y no es necesario duplicar el texto para conseguir este objetivo.
- Para obtener los distintos marcados del documento solamente es necesario definir una hoja de transformación XSL que convierta el documento metamarcado en cada una de sus interpretaciones.
- Mediante la definición de una hoja de transformación XSL todos los documentos XML del sistema son convertidos en documentos metamarcados. Además es necesario una herramienta adicional que permita combinar en un documento las distintas dimensiones.
- Estos documentos metamarcados solamente necesitan un visualizador de documentos XML que tenga soporte para hojas de transformación XSL, para obtener cada una de las interpretaciones.

Las siguientes pautas de investigación son:

- Establecer una medida de rendimiento del uso de esta tecnología sobre documentos con un volumen elevado de dimensiones y de información.
- Aplicar esta técnica en un sistema de información real. Actualmente se está integrando en un sistema de información de Museos [11].
- Debido a que esta solución permite una integración de conocimiento múltiple del documento, éste puede ser aplicado en problemas como evolución y versionado de documentos.

Bibliografía

1. XML y tecnología relacionada. Disponible en <http://www.w3.org>
2. TEI. <http://www.tei-c.org>
3. TEI. Múltiple Hierarchies. <http://www.tei-c.org/P4X/NH.html>
4. SGML. <http://www.w3.org/MarkUp/SGML/>
5. Durand, David G. Palimpsest: Change-Oriented Concurrency Control for the Support of Collaborative Applications, Phd Thesis. Boston University. 1999.
6. MECS - A Multi-Element Code System. Huitfeldt, Claus. Año 1998. TexMECS. An experimental markup meta-language for complex documents. Claus Huitfeldt. 2001
7. Concurrent markup for XML documents. Patrick Durusau, Matthew Brook O'Donnell. XML Europe 2002. Barcelona. 2002.
8. Old Norse-Icelandic Primary Sources Project. Encoding old-norse-Icelandic primary sources using TEI-conformant SGML/XML: A handbook. M. J. Driscoll., 2001.
9. Hyperlink semantics for standoff markup of read-only documents. Henry S. Thompson, David McKelvie. 1997.
10. Women Writer Project. Introduction to encoding. A tutorial for new encoders. Carole E. Mah . <http://www.wwp.brown.edu/encoding/training/intro/intro.html>. 1999.
11. MUSEO: Hacia un Entorno Experimental de Sistemas de Información. Antonio Polo, Miryam Salas, Juan Carlos Manzano, Luis Arévalo, Elena Jurado. Valladolid. 2000.