

Descubrimiento de Información en Textos Tarea del Tema 3 Estándares de Anotaciones

José Alberto Benítez Andrades

71454586A

Descubrimiento de Información en Textos

Máster en Lenguajes y Sistemas Informáticos - Tecnologías del Lenguaje en la Web

UNED

7 de enero de 2011

Tarea obligatoria del Tema 3. Estándares de anotaciones

El fichero Practica_Tema_3_ejemplo.xml que encontrarás en el repositorio de materiales contiene un documento XML con ejemplares de un tipo de documento administrativo: Órdenes Forales. La realización de la práctica consiste en:

1. Comprobar si dicho documento XML es un documento conforme con TEI Lite o con un subconjunto conforme con TEI Lite. Explicar por qué lo es o no.
2. En caso de que no lo sea, realizar las adiciones y sustituciones de elementos y atributos necesarias para que lo sea. Explica el porqué de cada adición y sustitución.
3. Diseña una DTD que sólo contemple los elementos y atributos XML que hayas utilizado.

Para editar documentos XML puedes ayudarte de algún editor de XML. Hay numerosos editores en el mercado, la mayoría de ellos de pago. Si utilizas el editor Emacs, bien en Linux o en Windows, hay un paquete TEI-Emacs que te puedes descargar de forma gratuita de <http://sourceforge.net/projects/tei/files/tei-emacs/>. En caso de que no utilices Emacs, hay software comercial que permite descargas de prueba de 30 días. Por ejemplo, el editor XML <oXygen/>:

http://www.oxygenxml.com/download_oxygenxml_editor.html.

En <http://wdvl.internet.com/Software/XML/editors.html> podéis encontrar una lista de editores XML.

1. Comprobar si dicho documento XML es un documento conforme con TEI Lite o con un subconjunto conforme con TEI Lite. Explicar por qué lo es o no.

El documento Practica_Tema_3_ejemplo.xml no está conforme con TEI Lite por los siguientes motivos:

- <corpus> no es una etiqueta adecuada para comenzar un corpus teniendo en cuenta la normativa TEI Lite.
- <ITEM> no es una etiqueta adecuada que muestre un objeto de tipo TEI según la normativa TEI Lite.
- <encabezado> <tipo> e <idioma> no es la manera correcta de etiquetar una cabecera según TEI Lite.
- En los elementos <rs> el atributo *id* no es el correcto.
- Todos los valores de los atributos deben estar entre comillas dobles .
Por ejemplo: <rs type="law" key="LES2">
- Los elementos <seg#10> y derivados, están mal nombrados.

2. En caso de que no lo sea, realizar las adiciones y sustituciones de elementos y atributos necesarias para que lo sea. Explica el porqué de cada adición y sustitución.

Las sustituciones que he realizado sobre el fichero de ejemplo han sido las siguientes:

1. El elemento `<corpus>` ha sido sustituido por el elemento `<teiCorpus>`
2. El elemento `<ITEM>` ha sido sustituido por el elemento `<TEI.2>`
3. El elemento `<encabezado>` ha sido sustituido por el elemento `<teiHeader>`
4. El elemento `<tipo>` ha sido sustituido por el atributo `type` dentro del elemento `<teiHeader>`. Ej: `<teiHeader type="loquesea">`
5. El elemento `<idioma>` ha sido sustituido por el atributo `lang` en el elemento `<teiHeader>`. Ej: `<teiHeader type="loquesea" lang="es">`
6. En los elementos `<item>` he modificado `n=X` por `n="X"`, es decir, en el valor de los atributos he puesto comillas. Todos los atributos que tenían un valor asignado sin unas comillas previas, han sido modificados con comillas, por ejemplo: `<rs type=law id=LES2>` se ha modificado a `<rs type="law" id="LES2">`
7. A la hora de asignar un id a los elementos de tipo `<rs>`, he modificado el atributo `id` por el atributo `key` de la siguiente manera:
`<rs type="law" id="LES2">` por `<rs type="law" key="LES2">`
8. En los elementos de tipo `seg` he modificado su manera de numerarlos, donde había #, las he quitado, por ejemplo: `<seg#9>` ahora es `<seg9>`.

3. Diseña una DTD que sólo contemple los elementos y atributos XML que hayas utilizado.

El documento DTD generado contiene el siguiente formato:

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT text ((body))>
<!ELEMENT teiHeader (#PCDATA)>
<!ATTLIST teiHeader
    lang CDATA #FIXED "es"
    type (T3 | T2) #REQUIRED
>
<!ELEMENT teiCorpus ((TEI.2+))>
<!ELEMENT seg9 ((p+))>
<!ELEMENT seg13 (#PCDATA | colon)*>
<!ELEMENT seg10 ((p))>
<!ELEMENT s (#PCDATA | rs | num | date | dateline | docAuthor | colon | category | keyent |
name | seg13)*>
<!ELEMENT rs (#PCDATA)>
<!ATTLIST rs
```

7 de enero de 2011

```

    key (UES5 | UES4 | UES3 | UES2 | UES1 | TLES6 | TLES5 | TLES4 | TLES3 | TLES2 |
    TLES1 | PES6 | PES5 | PES4 | PES3 | PES2 | PES1 | PBES1 | OES9 | OES8 | OES7 | OES6 |
    OES5 | OES4 | OES3 | OES2 | OES15 | OES14 | OES13 | OES12 | OES11 | OES10 | OES1 |
    NES6 | NES5 | NES4 | NES3 | NES2 | NES1 | LES9 | LES8 | LES7 | LES6 | LES5 | LES4 |
    LES3 | LES2 | LES17 | LES16 | LES15 | LES14 | LES13 | LES12 | LES11 | LES10 | LES1)
#REQUIRED
    type (uncat | title | publication | place | organization | name | law) #REQUIRED
>
<!ELEMENT p ((docAuthor | s))>
<!ELEMENT num (#PCDATA)>
<!ATTLIST num
    num (7 | 69 | 60394 | 6 | 58 | 5 | 387 | 3092 | 30 | 20 | 2 | 193 | 18394 | 109 | 10094 | 1 |
    09894) #REQUIRED
>
<!ELEMENT name (#PCDATA | rs)*>
<!ELEMENT list ((item+))>
<!ELEMENT keyent (#PCDATA | classCode | date)*>
<!ELEMENT item ((p))>
<!ATTLIST item
    n (4 | 3 | 2 | 1) #REQUIRED
>
<!ELEMENT docAuthor (#PCDATA | rs | s)*>
<!ELEMENT div3 ((p))>
<!ELEMENT div2 (((p, seg9, seg10) | (list, seg9, seg10?)))>
<!ELEMENT div1 ((p))>
<!ELEMENT dateline (#PCDATA | rs | date)*>
<!ELEMENT date (#PCDATA | rs)*>
<!ATTLIST date
    date CDATA #IMPLIED
>
<!ELEMENT colon (#PCDATA)>
<!ELEMENT classCode (#PCDATA)>
<!ELEMENT category ((rs))>
<!ELEMENT body ((div1, div2, div3))>
<!ELEMENT TEI.2 ((teiHeader, text))>

```

Y para finalizar, el XML quedó de la siguiente forma:

```

<teiCorpus>
  <TEI.2>
    <teiHeader type="T3" lang="es">
    </teiHeader>
    <text>
      <body>
        <div1>
          <p> <s> Mediante <category> <rs type="law" key="LES1"> Orden Foral </rs>
          </category> número <keyent> <classCode> 3607/94, </classCode> de <date> 9 de <rs
          type="uncat" key="UES1"> Noviembre </rs> </date> </keyent> <name> del <rs type="title"
          key="TLES1"> Diputado Foral de Medio Ambiente y Acción Territorial </rs>, </name> <seg13>
          ha adoptado la resolución cuya parte dispositiva es la siguiente <colon> : </colon> </seg13>
          </s>
          </p>
        </div1>
      <div2>
        <list>
          <item n="1">

```

7 de enero de 2011

<p> <s> Primero <colon> : </colon> Revocar el nombramiento provisional otorgado mediante <rs type="law" key="LES2"> Orden Foral </rs> número <num num="60394"> 603/94 </num>, de

<date> 22 de febrero </date>, a favor de doña <rs type="name" key="NES1"> Ana Fernández Gutierrez-Crespo </rs>, para el puesto de <rs type="organization" key="OES1"> Tesorería del Ayuntamiento de Getxo </rs> por incapacidad laboral transitoria de su titular, por haber fallecido

este último . </s> </p>

</item>

<item n="2">

<p> <s> Segundo <colon> : </colon> Conferir nombramiento provisional en favor de Dña . <rs type="name" key="NES2"> Ana Fernández Gutierrez-Crespo </rs>, para el puesto de <rs type="organization" key="OES2"> Tesorería </rs> vacante en el <rs type="organization" key="OES3"> Ayuntamiento de Getxo </rs> . </s> </p>

</item>

<item n="3">

<p> <s> Tercero <colon> : </colon> El nombramiento podrá ser revocado en cualquier momento por el <rs type="title" key="TLES2"> Diputado Foral de Medio Ambiente y Acción Territorial </rs>, a propuesta de la <rs type="organization" key="OES4"> Corporación Local </rs> interesada, con audiencia del

funcionario o a instancia de éste, previo informe de la <rs type="organization" key="OES5"> Corporación </rs> . </s> </p>

</item>

<item n="4">

<p> <s> Cuarto <colon> : </colon> La presente <rs type="law" key="LES3"> Orden Foral </rs> se notificará a la interesada, al <rs type="organization" key="OES6"> Ayuntamiento de Getxo </rs> y al <rs type="organization" key="OES7"> Ministerio para las Administraciones Públicas </rs>, ordenándose su publicación en el « <rs type="publication" key="PBES1"> Boletín Oficial de Bizkaia </rs> ». </s> </p>

</item>

</list>

<seg9>

<p> <s> La anterior resolución es definitiva, contra la misma podrá interponerse recurso contencioso

administrativo en el plazo de dos meses contados desde el día siguiente a aquel en que tenga lugar la

notificación del presente escrito sin perjuicio de que los interesados puedan ejercitar cualquier otro

recurso que estimen pertinente de acuerdo con la legislación vigente . </s> </p>

<p> <s> Todo ello de conformidad con lo que disponen el artículo <num num="69"> 69 </num> de la <rs type="law" key="LES4"> Norma Foral </rs> <num num="387"> 3/87 </num>, de <date> 13 de febrero </date> ,

sobre <rs type="uncat" key="UES2"> Elección </rs>, <rs type="organization" key="OES8"> Organización </rs>, <rs type="law" key="LES5"> Régimen y Funcionamiento de las Instituciones Forales del Territorio Histórico de Bizkaia </rs>, y su modificación por <rs type="law" key="LES6"> Norma Foral </rs> <num num="193"> 1/93 </num>, de <date> 18 de febrero </date>, el artículo <num num="58"> 58 </num> y

siguientes de la <rs type="law" key="LES7"> Ley de la Jurisdicción Contencioso Administrativa </rs> de <date date="27/12/1956"> 27 de diciembre de 1956 </date> y el artículo <num num="109"> 109 </num> de la <rs type="law" key="LES8"> Ley </rs> <num num="3092"> 30/92 </num>, de <date> 26 de noviembre </date>, de <rs type="law" key="LES9"> Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común </rs> . </s> </p>

</seg9>

</div2>

<div3>

<p> <s> <dateline> <rs type="place" key="PES1"> Bilbao </rs>, <date date="9/11/1994"> 9 de noviembre de 1994 </date>. </dateline> - <docAuthor> El <rs type="title" key="TLES3">

7 de enero de 2011

Director General de Medio Ambiente y Acción Territorial </rs> , <rs type="name" key="NES3">
Ander Salaberria Amesti </rs>

```

    </docAuthor> </s> </p>
  </div3>
</body>
</text>
</TEI.2>
<TEI.2>
  <teiHeader type="T2" lang="es">
    </teiHeader>
    <text> <body> <div1>
      <p> <s> <category> <rs type="law" key="LES10"> Orden Foral </rs> </category> número
<keyent> <classCode> 795/1994 </classCode> de <date> 28 de octubre. </date> </keyent>
</s> </p>
    </div1>
    <div2>
      <p> <s> Aprobar definitivamente la <rs type="organization" key="OES9"> Modificación
Puntual de las Normas Subskeyiarias Municipales </rs> en la U.A.
      número <num num="20"> 20 </num>, para la ejecución de viviendas de <rs
type="organization" key="OES10"> Protección Oficial </rs>, en el municipio de <rs type="place"
key="PES2"> Gautegiz de Arteaga </rs>
      debiendose de incorporar la U.A. <num num="20"> -20 a </num> la calificación de <rs
type="uncat" key="UES3"> Suelo Urbano de Media Denskeyad </rs>, con un
      máximo de <num num="5"> 5 </num> viviendas . </s> </p>
    <seg9>
      <p> <s> Contra dicha <rs type="law" key="LES11"> Orden Foral </rs>, que agota la vía
administrativa podrá interponerse
      recurso contencioso-administrativo ante la <rs type="organization" key="OES11"> Sala
de lo Contencioso-Administrativo del Tribunal Superior de Justicia del País Vasco </rs>, en el
plazo de dos meses, contado desde el día siguiente a esta notificación,
      sin perjuicio de la utilización de otros medios de defensa que estime oportunos . </s>
</p>
    </seg9>
    <seg10>
      <p> <s> Durante el referkeyo plazo el expediente BHI-<num num="10094"> 100/94-
</num> P05- A quedará de manifiesto para su exámen en las
      dependencias de <rs type="place" key="PES3"> Bilbao calle Alameda Rekalde </rs>,
<num num="30"> 30 </num>, <num num="5"> 5.a </num> y <num num="6"> 6.a </num>
plantas . </s> </p>
    </seg10>
  </div2>
  <div3>
    <p> <docAuthor> <s> El <rs type="title" key="TLES4"> Diputado Foral de Urbanismo
</rs> . </s> <s> - <rs type="name" key="NES4"> Pedro Hernández González </rs>

    </s> </docAuthor> </p>
  </div3>
</body>
</text>
</TEI.2>
<TEI.2>
  <teiHeader type="T2" lang="es">
    </teiHeader>
    <text> <body> <div1>
      <p> <s> <category> <rs type="law" key="LES12"> Orden Foral </rs> </category> número
<keyent> <classCode> 818/1994 </classCode> de <date> 3 de noviembre. </date> </keyent>
</s> </p>
    </div1>

```

7 de enero de 2011

<div2>

<p> <s> Aprobar definitivamente la <rs type="law" key="LES13"> Modificación de las Normas Subskeyiarias de Planeamiento de Leioa </rs> en cuanto a los límites y ordenación del área <num num="5"> 5. </num> </s> </p>

<seg9>

<p> <s> Contra dicha <rs type="law" key="LES14"> Orden Foral </rs>, que agota la vía administrativa, podrá interponerse recurso contencioso-administrativo ante la <rs type="organization" key="OES12"> Sala de lo Contencioso-Administrativo del Tribunal Superior de Justicia del País Vasco </rs>, en el plazo de dos meses, contado desde el día siguiente a esta notificación, sin perjuicio de la utilización de otros medios de defensa que estime oportunos . </s> </p>

</seg9>

<seg10>

<p> <s> Durante el referkeyo plazo el expediente BHI-<num num="09894"> 098/94-</num> P05- A quedará de manifiesto para su exámen en las dependencias de <rs type="place" key="PES4"> Bilbao calle Alameda Rekalde </rs>, <num num="30"> 30 </num>, <num num="5"> 5.a </num> y <num num="6"> 6.a </num> plantas . </s> </p>

</seg10>

</div2>

<div3>

<p> <docAuthor> <s> El <rs type="title" key="TLES5"> Diputado Foral de Urbanismo </rs> . </s> <s> - <rs type="name" key="NES5"> Pedro Hernández González </rs>

</s> </docAuthor> </p>

</div3>

</body>

</text>

</TEI.2>

<TEI.2>

<teiHeader type="T2" lang="es">

</teiHeader>

<text> <body> <div1>

<p> <s> <category> <rs type="law" key="LES15"> Orden Foral </rs> </category> número <keyent> <classCode> 819/1994 </classCode> de <date> 3 de noviembre. </date> </keyent> </s> </p>

</div1>

<div2>

<list>

<item n="1">

<p> <s> <num num="1"> 1. </num> Aprobar definitivamente la <rs type="law" key="LES16"> Modificación Puntual de las Normas Subskeyiarias </rs> del término municipal de <rs type="place" key="PES5"> Atxondo </rs>, en la <rs type="uncat" key="UES4"> Unkeyad de Actuación </rs> UA-<num num="7"> 7 </num> debiendo corregirse los errores expresados en el fundamento tercero . </s> </p>

</item>

<item n="2">

<p> <s> <num num="2"> 2. </num> El <rs type="organization" key="OES13"> Ayuntamiento de Atxondo </rs>, una vez corregkeyo el documento remitirá tres ejemplares del <rs type="uncat" key="UES5"> Texto Refundkeyo </rs> a este <rs type="organization" key="OES14"> Departamento Foral </rs> para proceder a su autenticación . </s> </p>

</item>

</list>

<seg9>

<p> <s> Contra dicha <rs type="law" key="LES17"> Orden Foral </rs>, que agota la vía administrativa, podrá interponerse recurso contencioso-administrativo ante la <rs type="organization" key="OES15"> Sala de lo Contencioso-Administrativo del Tribunal Superior

7 de enero de 2011

de Justicia del País Vasco </rs>, en el plazo de dos meses, contado desde el día siguiente a esta notificación, sin perjuicio de la utilización de otros medios de defensa que estime oportunos . </s> </p>

</seg9>

<seg10>

<p> <s> Durante el referkeyo plazo el expediente BHI- <num num="18394"> 183/94- </num> P05- A quedará de manifiesto para su exámen en las dependencias de <rs type="place" key="PES6"> Bilbao calle Alameda Rekalde </rs>, <num num="30"> 30 </num>, <num num="5"> 5.a </num> y <num num="6"> 6.a </num> plantas . </s> </p>

</seg10>

</div2>

<div3>

<p> <docAuthor> <s> El <rs type="title" key="TLES6"> Diputado Foral de Urbanismo </rs> . </s> <s> - <rs type="name" key="NES6"> Pedro Hernández González </rs>

</s> </docAuthor> </p>

</div3>

</body>

</text>

</TEI.2>

</teiCorpus>