

Introducción a la Edición Digital Académica.

María Gimena del Rio Riande y Clara Martínez Cantón.

Cita:

María Gimena del Rio Riande y Clara Martínez Cantón (Noviembre, 2014). *Introducción a la Edición Digital Académica. I Jornadas Nacionales de Humanidades Digitales: Culturas, Tecnologías, Saberes. Asociación Argentina de Humanidades Digitales, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/gimena.delrio.riande/67>

ARK: <https://n2t.net/ark:/13683/pdea/uaS>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.



Las Humanidades Digitales
desde **Argentina**

Tecnologías, Culturas, Saberes



Introducción a la Edición Digital Académica

Gimena del Rio (CONICET)
gdelrio@conicet.gov.ar

Clara Martínez Cantón (UNED)
cimartinez@flog.uned.es

Transcripción de un texto

Digitalización

Paz a España, paz segura.
Canten abiertos los campos
dichosos de Extremadura.
Lean los niños, las flores,
y entre las negras encinas
todos los trabajadores.
¡Lejos tanta noche oscura!
¡Para siempre en primavera,
le tierra de Extremadura!
Yo os traigo en este pregón,
el libro, fuente de gracia,
que ilumina el corazón.
Agua tranquila, agua pura,
agua que riegue la sangre
del alma de Extremadura.
Libros de ^{amor, luz,} ~~para~~, ensueño.
Alta vida y viento en calma
en todo el pueblo extremeño.

Transcripción

Paz a España, paz segura.
Canten abiertos los campos
dichosos de Extremadura. !
Lean los niños, las flores,
y entre las negras encinas
todos los trabajadores. !
¡Lejos tanta noche oscura!
¡Para siempre en primavera,
la tierra de Extremadura! !
Yo os traigo en este pregón,
el libro, fuente de gracia,
que ilumina el corazón. !
Agua tranquila, agua pura,
agua que riegue la sangre
del alma de Extremadura. !
Libros de amor, luz, ensueño.
Alta vida y viento en calma
en todo el pueblo extremeño

Codificación

```
<poema>
  <estrofa>
    <verso>Paz a <topon>España</topon>, paz segura.</verso>
    <verso>Canten abiertos los campos</verso>
    <verso>dichosos de <topon>Extremadura</topon>.</verso>
  </estrofa>
  <estrofa>
    <verso>Lean los niños, las flores,</verso>
    <verso>y entre las negras encinas</verso>
    <verso>todos los trabajadores.</verso>
  </estrofa>
  <estrofa>
    <verso>¡Lejos tanta noche oscura!</verso>
    <verso>¡Para siempre en primavera,</verso>
    <verso>la tierra de <topon>Extremadura</topon>!</verso>
  </estrofa>
  <estrofa>
    <verso>Yo os traigo en este pregón,</verso>
    <verso>el libro, fuente de gracia,</verso>
    <verso>que ilumina el corazón.</verso>
  </estrofa>
  <estrofa>
    <verso>Agua tranquila, agua pura,</verso>
    <verso>agua que riegue la sangre</verso>
    <verso>del alma de <topon>Extremadura</topon>.</verso>
  </estrofa>
  <estrofa>
    <verso>Libros de
      <sust><borr>luz</borr><adi>amor</adi></sust>,
      <sust><borr>para</borr><adi>luz</adi></sust>,
      ensueño.</verso>
    <verso>Alta vida y viento en calma</verso>
    <verso>en todo el pueblo extremeño.</verso>
  </estrofa>
</poema>
```

Lenguajes de marcado

- Se desarrollan para tratar con documentos (información predominantemente textual), hoy también para gestión de datos.
- Describen las partes lógicas (estructura lógica y semántica) de un tipo de documento (o conjunto de datos) particular, sin límite en el nivel de profundidad o detalle
- Son lenguajes porque tienen
 - Vocabulario: Para denominar a cada elemento estructural del documento de cualquier nivel de profundidad, a cada atributo de estos elementos, a cada valor de estos atributos (para cierto tipo de atributo), etc.
 - Sintaxis (o gramática): Para indicar qué construcciones estructurales están permitidas: qué debe o puede contener cada elemento, si un elemento o atributo es obligatorio, si un elemento puede o no repetirse, qué tipo de valor admite un atributo, etc.

¿Qué es XML?

- XML = Extensible Markup Language
- XML es un heredero, más sencillo y eficaz, de su antecesor SGML (Standard Generalized Markup Language)
- Se publicó en febrero de [1998](#) por el W3C Consortium.
- La documentación oficial está disponible en: <http://www.w3.org/XML/>
- Última versión: [XML 1.0 \(5ª edición\)](#)
- Tutoriales en líneas: [Introducción a XML elaborada por TEI](#)

¿Qué es TEI?



[Home](#) [Guidelines](#) [Activities](#) [Tools](#) [Membership](#) [Support](#) [About](#) [News](#)

[Home](#)

Entire site ▼ [Search](#)

TEI-C News

[Wiki page for P6 ideas](#)

Posted on: 2014-10-24

[Announcing TAPAS launch](#)

Posted on: 2014-10-24

[2014 TEI Conference, Evanston, Illinois](#)

Posted on: 2014-10-23

[Elections: Deadline VERY soon!](#)

Posted on: 2014-10-22

[REMINDER: TEI Board/Council voting deadline is October 24](#)

Posted on: 2014-10-20

[TEI Tweet Chat](#)

Posted on: 2014-10-06

Other News

[The iSchool at the University of Illinois is recruiting](#)

Posted on: 2014-10-29

[Job: Director of Academic IT Services, University of Oxford](#)

Posted on: 2014-10-29

TEI: Text Encoding Initiative

The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics. Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation. In addition to the Guidelines themselves, the Consortium provides a variety of [resources](#) and [training events](#) for learning TEI, information on [projects using the TEI](#), a [bibliography of TEI-related publications](#), and [software](#) developed for or adapted to the TEI.

The TEI Consortium is a nonprofit membership organization composed of academic institutions, research projects, and individual scholars from around the world. Members contribute financially to the Consortium and elect representatives to its Council and Board of Directors.

Want to become active in the TEI community? [Become a TEI Member](#), join a [special interest group](#), sign up for the [TEI-L mailing list](#), and come to our [annual conferences and members' meetings](#).

Last recorded change to this page: 2013-03-12 • For corrections or updates, contact web@tei-c.org

¿Qué es TEI?

- Text Encoding Initiative: Iniciativa de codificación de textos, especialmente en Humanidades, Ciencias Sociales y Lingüística.
- Se describe en las TEI *Guidelines*: el manual de uso
- Se expresa a través del lenguaje XML
- Tiene un carácter modular y es personalizable
- Es código abierto ([sourceforge](https://sourceforge.net/projects/tei/)), independiente de cualquier programa o plataforma.

Objetivos

- Promover la creación, el intercambio y la integración de los datos textuales informatizados.
- Editar textos de cualquier tipología, género y disciplina (Ciencias Humanas y Sociales), en cualquier lengua y de cualquier período cronológico.

TEI Consortium: <http://www.tei-c.org>

- [Guidelines](#)
- [Lista de proyectos](#) que utilizan TEI
- [Wiki](#)
- [Herramientas](#)
- [Lista de discusión](#)

Módulos de TEI

1. **Infraestructura TEI**
2. **TEI Header (encabezado) = header**
3. **Elementos disponibles en todos los docs TEI = core**
4. Estructura del texto por defecto = textstructure
5. Caracteres no estándares y Glyph = gaiji
6. **Verso = verse**
7. Drama = drama
8. Transcripción de discurso = spoken
9. Diccionarios = dictionaries
10. Descripción de manuscritos = msdescription
11. Representación de fuentes primarias= transcr

Un documento XML-TEI

- Tiene que estar bien formado, es decir estar expresado correctamente según la gramática de XML
- Tiene que ser válido
 - Para ello, necesita un modelo que fije la estructura básica y las posibilidades de anidación del documento. Las DTD o esquemas (RelaxNG, W3CSchema) establecen la sintaxis que debe tener el documento XML.

¿Qué significa bien formado?

1. Tiene que existir un único elemento raíz que contenga todo el documento
2. Cada subárbol debe estar correctamente anidado, es decir, no se pueden solapar
3. Los nombres de las etiquetas son *case/sensitive*:
<de> ≠ <De> ≠ <DE>
4. Una etiqueta de apertura <p> debe tener siempre una etiqueta de cierre </p> (salvo en los casos de etiquetas vacías <vacía/>)
5. No debe haber espacios en blanco tras < o antes de </>
<p>párrafo</p>
Y no < p>párrafo</ p> ni <p >párrafo</p >
6. El valor de los atributos ha de estar siempre entrecomillado
<p type="text">algo de texto</p>

¿Qué significa válido?

Un documento XML es válido cuando cumple las reglas establecidas en un **esquema externo**.

El esquema especifica:

1. El nombre del elemento raíz
2. El nombre de todos los elementos que se pueden utilizar
3. Los nombres, tipo de datos y (ocasionalmente) los valores por defecto de los atributos
4. Reglas acerca de cómo se pueden anidar los elementos
5. Y otras cosillas dependiendo del lenguaje del esquema

¿Para qué un esquema?

1. Asegura que en los documentos solo se utilizan los elementos, atributos y entidades predefinidas
2. Establece las reglas estructurales del tipo 'toda receta debe incluir una lista de ingredientes'
3. Asegura que la misma cosa se llame de la misma manera siempre

TEI Roma

TEI Roma: generating validators for the TEI Modules

You are currently working on **My TEI Extension**

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#) [Sanity Checker](#)

List of TEI Modules

	Module name	A short description	Changes
add	analysis	? Simple analytic mechanisms	
add	certainty	? Certainty and uncertainty	
add	core	? Elements common to all TEI documents	
add	corpus	? Corpus texts	
add	dictionaries	? Dictionaries	
add	drama	? Performance texts	
add	figures	? Tables, formulæ, notated music, and figures	
add	gaiji	? Character and glyph documentation	
add	header	? The TEI Header	
add	iso-fs	? Feature structures	
add	linking	? Linking, segmentation and alignment	
add	msdescription	? Manuscript Description	
add	namesdates	? Names and dates	
add	nets	? Graphs, networks, and trees	
add	spoken	? Transcribed Speech	
add	tagdocs	? Documentation of TEI modules	
add	textcrit	? Critical Apparatus	
add	textstructure	? Default text structure	
add	transcr	? Transcription of primary sources	
add	verse	? Verse structures	

List of selected Modules

[remove](#) [core](#)
[tei](#)
[remove](#) [header](#)
[remove](#) [textstructure](#)

TEI ROMA nos permite

- Generar y modificar un esquema
- Elegir la lengua de la documentación
- Elegir los módulos que nos convienen
- Añadir y eliminar elementos en una determinada clase
- Generar una documentación (en HTML, por ejemplo)
- Guardar nuestra personalización a través de un fichero ODD.
- Gracias a esta aplicación, podremos crear el fichero ODD = One Document Does it All

Ejercicio 1. Esquema en Roma



Ejercicio 2. Crear xml doc y asociar esquema



Semáforo verde

- Para crear un doc. En xml válido tengo que asociarle un esquema también válido. Solo así obtendré un documento bien formado

(LUZ ROJA vs. LUZ VERDE)

Etiquetas del doc (tags)

- Aislan una porción de texto con un significado semántico específico:
`<etiqueta>concepto</etiqueta>`
`<titulo>Int. a TEI</titulo>`
`<autor>Gimena del Rio</autor>`
`<fecha>17/11/2014</fecha>`
- La estructura consiste en la anidación sin límites de unos elementos en el interior de otros, construyendo así una estructura arborescente. Por eso, necesitamos siempre un solo elemento raíz del que cuelguen todos los otros.

Véase: [documento xml basico.xml](#)

Etiquetas: elementos, atributos y valores

```
<object type="value">contenido</object>
```

- **Elemento:** son las unidades de información semántica compuestas de una marca de apertura <elemento> y de cierre </elemento>. El nombre del elemento constituye su identificador general. Puede contener texto, otros elementos o estar vacío (<elemento/>).
- **Atributo:** añaden una característica adicional al elemento; su orden no es relevante; deben siempre llevar un valor.
- **Valor:** rellena el valor o contenido del atributo.
- Los atributos pueden ser generales, es decir, utilizables en todos los elementos TEI, como @xml:id, @xml:lang, o específicos de un determinado elemento.

Tipos de elementos

Un elemento puede contener, según se especifique en la DTD o esquema (si existe):

1. Otros elementos, subelementos o elementos hijos .

Ej.: `<text><body><lg><l>`

2. Texto.

Ej.: `<author>Prescod, Paul</author>`

3. Contenido mixto, o sea, una mezcla de elementos hijos y texto.

Ej.: `<titulo>Charles F. Goldfarb's<enfasis>XML</enfasis>
Handbook</titulo>`

4. Nada, si es un elemento vacío. Ej.: `<Cartel fuente="laentrega.jpg"/>`

Tipos de elementos

Un elemento puede ser según se especifique en la DTD o esquema (si existe):

- 1. Obligatorio**
- 2. Opcional**
- 3. Prohibido**

- Tipos de elementos según contenido
 - Vacíos
 - De texto
 - De otros elementos
 - Mixtos (texto y otros elementos)
- Según obligatoriedad
 - Obligatorios
 - Opcionales
 - No permitidos

Resultado: etiquetas core

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3   xsi:schemaLocation="http://www.tei-c.org/ns/1.0 metrica%20esquema%20jul14.xsd"
4   xmlns="http://www.tei-c.org/ns/1.0">
5   <teiHeader>
6     <fileDesc>
7       <titleStmt>
8         <title></title>
9       </titleStmt>
10      <publicationStmt>
11        <publisher></publisher>
12      </publicationStmt>
13      <sourceDesc>
14        <bibl>
15          <title></title>
16        </bibl>
17      </sourceDesc>
18    </fileDesc>
19  </teiHeader>
20  <text>
21    <body>
22      <lg>
23        <l></l> </lg>
24    </body>
25  </text>
26 </TEI>
```



Cancionero Popular de Jujuy / recogido y anotado por Juan Alfonso Carrizo

Carrizo, Juan Alfonso
Cancionero Popular de Jujuy



Autor: [Carrizo, Juan Alfonso](#)

Título: Cancionero Popular de Jujuy / recogido y anotado por Juan Alfonso Carrizo

Publicación: Alicante : Biblioteca Virtual Miguel de Cervantes, 2003

Publicación original: Tucumán, Miguel Violento, 1934.

Notas de reproducción original: Edición digital basada en la de Tucumán, Miguel Violento, 1934.

Portales: [Academia Argentina de Letras](#) | [Literatura Infantil y Juvenil](#)

Encabezamiento de materia:

[Poesía argentina -- Siglo 19º](#)

CDU:

821.134.2(82)-1"18"

Idioma: español

URI: <http://www.cervantesvirtual.com/nd/ark:/59851/bmcq23t8>

MÁS RECIENTES

Estudios sobre Salvador Rueda / Amparo Quiles Faz

Cinco reglas de nuestra moral antigua: una interpretación / Teodoro M. Kalaw

"La fortuna del mendigo" : farsa en dos actos / original de Rogelio San Luis

Debats Núm. 62-63, otoño 1998

Artecontexto, arte, cultura y nuevos medios Núm. 8, 2005

Artecontexto, arte, cultura y nuevos medios Núm. 9, 2006

Los orígenes de los liberalismos iberoamericanos (1750-1868) / Antonio-Filiu Franco Pérez

Scherzo Año II, núm. 10, diciembre 1986

Scherzo Año II, núm. 19, noviembre 1987

La historia del fuego / Remei Sipi ; Inmaculada Díaz Narbona (ed.)

1782-a

Arbolito naranjado,
¡bonito pago me has dado!,
después de quedar conmigo,
a la luna me has dejado.

http://www.cervantesvirtual.com/obra-visor-din/cancionero-popular-de-jujuy--0/html/ffea72c4-82b1-11df-acc7-002185ce6064_168.html#I_2265

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3   xsi:schemaLocation="http://www.tei-c.org/ns/1.0 metrica%20esquema%20jul14.xsd"
4   xmlns="http://www.tei-c.org/ns/1.0">
5   <teiHeader>
6     <fileDesc>
7       <titleStmt>
8         <title>Arbolito naranjado</title>
9       </titleStmt>
10      <publicationStmt>
11        <publisher>Clara</publisher>
12      </publicationStmt>
13      <sourceDesc>
14        <bibl>
15          <title>Cancionero popular de Jujuy</title>
16        </bibl>
17      </sourceDesc>
18    </fileDesc>
19  </teiHeader>
20  <text>
21    <body>
22      <lg>
23        <l>Arbolito naranjado</l>
24        <l>;bonito pago me has dado</l>
25        <l>después de quedar conmigo</l>
26        <l>a la luna me has dejado</l> </lg>
27    </body>
28  </text>
```