

[Inicio](#) | [Presentación](#) | [Instrucciones autores](#) | [Call for papers](#) | [Indizada en](#) | [Archivos](#)

Búsqueda (<http://temaria.net/simple.php?origen=1575-5886&idioma=es>)

TEI: un estándar para codificar textos en el ámbito de las humanidades digitales

[Versió catalana]

① RUBÉN ALCARAZ MARTÍNEZ
EINA, Centre Universitari de Disseny i Art de
Barcelona. Archivo

ralcaraz@eina.cat

① ELISABET VÁZQUEZ PUIG
Biblioteca Can Rajoler (Parets del Vallès)
vazquezpe@diba.cat

Opciones



PDF

<meta />

Metadatos

Resumen

Se describen las principales características del estándar TEI (Text Encoding Initiative), un lenguaje basado en XML para codificar cualquier tipo de estructura textual como, por ejemplo, novelas, obras teatrales, poemas, discursos o artículos científicos, entre otros. Se muestran sus diferentes campos de aplicación y se comentan diferentes ejemplos de proyectos representativos en el ámbito de las humanidades digitales y las colecciones patrimoniales en línea.

Resum

Es descriuen les característiques principals de l'estàndard TEI (Text Encoding Initiative), un llenguatge basat en l'XML per codificar qualsevol tipus d'estructura textual, com ara novel·les, obres teatrals, poemes, discursos o articles científics, entre d'altres. Es mostren els diferents camps d'aplicació i es comenten diferents exemples de projectes representatius en l'àmbit de les humanitats digitals i les col·leccions patrimonials en línia.

Abstract

In this paper we describe the main features of the mark-up language developed by the Text Encoding Initiative (TEI), now the standard protocol for digitally encoding all types of text structure, from novels and plays to poetry, speeches and scientific articles. We examine different areas of application and look at examples of digital humanities projects and online heritage collections that use TEI.

Palabras clave: TEI, XML, Bibliotecas digitales

Keywords: TEI, XML, Digital libraries

Recibido: 30/06/2016. Aceptado: 14/07/2016.

1 ¿Qué es la TEI?

La Text Encoding Initiative (TEI) se encarga de desarrollar un estándar para representar textos en formato digital, basado en el metalenguaje XML.

Los orígenes de la iniciativa se remontan al año 1987, en el marco de un proyecto de cooperación internacional bajo los auspicios de la [Association for Computers and the Humanities \(http://ach.org/\)](http://ach.org/), la [Association for Computational Linguistics \(http://www.aclweb.org/\)](http://www.aclweb.org/), y la Association for Literary and Linguistic Computing, actual [European Association for Digital Humanities \(http://eadh.org/\)](http://eadh.org/). El objetivo de este proyecto no era otro que el de encontrar un esquema de codificación común para estructuras textuales complejas, que permitiese reducir la diversidad de prácticas existentes en aquel momento por lo que respecta a la codificación de textos digitales.

Desde el año 2000, se trata de un proyecto mantenido por el [Consorcio TEI \(http://www.tei-c.org/\)](http://www.tei-c.org/), una organización sin ánimo de lucro integrada por instituciones académicas y por proyectos de

investigación y académicos individuales de todo el mundo.¹ Los objetivos de este consorcio son: desarrollar las directrices TEI, promover y difundir estas directrices, formar y divulgar en el uso del estándar, y crear y mantener una comunidad de investigadores en torno a la iniciativa.²

Para evitar confusiones, es necesario precisar que el término *TEI* se utiliza indistintamente para hacer referencia a: la Text Encoding Initiative o el Consorcio TEI, la documentación oficial conocida como las *TEI guidelines* y el mismo lenguaje. En este artículo, utilizaremos la forma femenina del término para referirnos a la iniciativa, y hablaremos del estándar TEI para referirnos al lenguaje en formato XML desarrollado y mantenido por el Consorcio TEI.

2 ¿En qué consiste la codificación de textos?

En el contexto que nos ocupa, Renear (2004) define la codificación de textos como "the practice of creating machine-readable texts to support humanities research". En este sentido, el estándar TEI ofrece un esquema que incluye y describe un conjunto de elementos que permiten marcar las principales características estructurales, interpretativas y conceptuales de diferentes tipologías de textos (literarios, periodísticos, científicos, etc.), con el objetivo de poder ser procesadas posteriormente (Baena Sánchez [*et al.*], 2014). La codificación de textos permite, pues, facilitar la recuperación de la información contenida en los documentos, a los cuales se podrá acceder de manera sistemática como si se tratase, salvando las distancias, de una base de datos. También facilita los análisis informatizados como, por ejemplo, los estudios estilométricos para clasificar textos o para determinar su autoría, los análisis de contenido o los estudios estadísticos, entre otros.

3 ¿Por qué XML?

Desde sus orígenes, el estándar TEI se ha creado bajo la premisa de ser un lenguaje flexible y extensible, capaz de representar cualquier tipo de texto y de adaptarse a las necesidades específicas de cada centro o proyecto de investigación.³ El uso, primero del estándar SGML (Standard Generalized Markup Language) y, a partir de 2002, de XML (Extensible Markup Language), aseguran esta premisa.

El metalenguaje XML es un estándar abierto e internacionalmente reconocido, que se caracteriza por su simplicidad, facilidad de uso y flexibilidad, y destaca por su capacidad para asegurar la interoperabilidad entre aplicaciones, plataformas y lenguajes informáticos (Allés Torrent, 2015).

XML no se presenta solo, sino que viene acompañado de todo un conjunto de tecnologías que lo complementan. Entre otras, son:

- **XSL** (Extensible Stylesheet Language). Un lenguaje que, junto con las CSS (Cascading Style Sheets), permite definir la presentación, visualización y transformación de los documentos XML en un medio específico como, por ejemplo, una página web.
- **XPath** (XML Path Language). Pensado para acceder a partes concretas de un documento XML.
- **XLink**. Orientado a la creación de relaciones internas y externas entre documentos XML.
- **XPointer**. Pensado para identificar y hacer referencias a puntos concretos o fragmentos que se encuentran dentro de un documento XML.
- **XQL** (XML Query Language). Lenguaje de consulta que facilita la extracción de datos desde documentos XML.

Las características propias de XML, combinadas con el resto de tecnologías que acabamos de describir, permiten que el estándar TEI presente una gran capacidad para adaptarse constantemente y evolucionar de acuerdo a nuevas necesidades y retos, y mantener, al mismo tiempo, sus principios fundamentales (Romary, 2009). Por otro lado, tecnologías como XSL, y la pareja XPath y XQL permiten garantizar la edición en diferentes formatos de los textos codificados, y facilitan la explotación de los datos codificados, respectivamente.

4 La infraestructura TEI

Las *TEI Guidelines for electronic text encoding and interchange* (TEI Consortium, 2016) son unas pautas que definen y documentan el estándar TEI. Estas directrices se organizan en diferentes capítulos en los cuales se aborda el sistema de clases y de atributos, y los diferentes módulos que conforman el esquema.

4.1 Los módulos

El esquema de codificación TEI se encuentra formado por 21 módulos independientes en los que se declaran los diferentes elementos y atributos XML disponibles. Esta manera de estructurar el esquema, hace que su uso sea altamente flexible y extensible, y que sea extremadamente fácil combinar y añadir diferentes elementos para confeccionar un esquema adecuado a los requerimientos de cada tipo de proyecto.

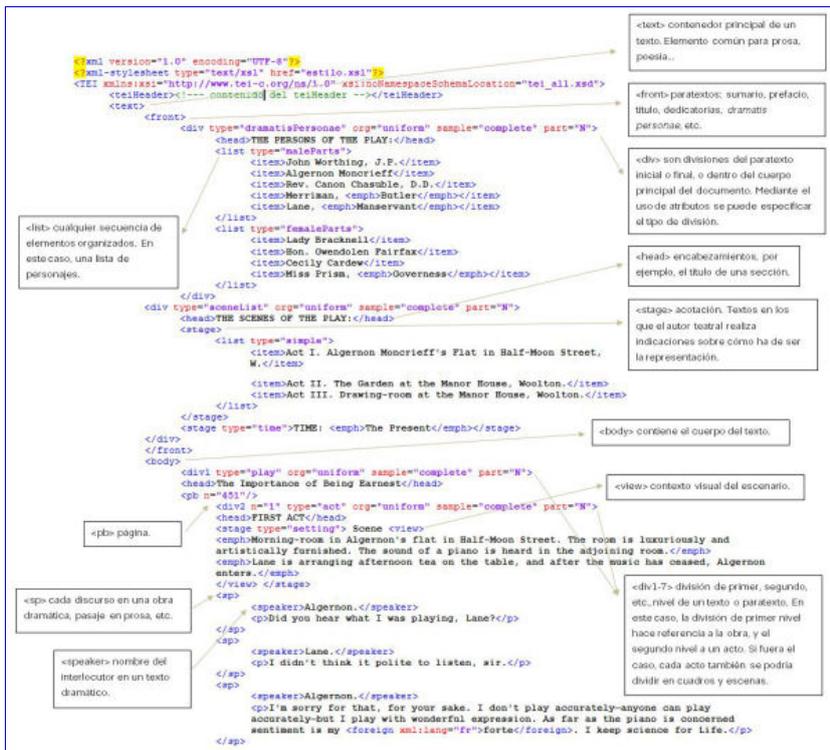
La combinación de módulos es totalmente libre. Sin embargo, existen 4 módulos obligatorios que deben estar presentes en cualquier combinación: *tei*, *core*, *header* y *textStructure*.

La lista completa de módulos es la siguiente:

- **analysis**: pensado para asociar análisis simples e interpretaciones a elementos textuales. Por ejemplo, para indicar el predicado de una frase, el prefijo o sufijo de una palabra, o su categoría gramatical.
- **certainty**: permite indicar que aspectos concretos del texto son problemáticos o inciertos. Esto se

- manifiesta en forma de notas o de aclaraciones en el texto.
- core:** se trata de un conjunto de elementos básicos disponibles para la codificación de cualquier tipo de texto como, por ejemplo, párrafos, listas, texto enfatizado, referencias bibliográficas, etc.
- corpus:** pensado para crear corpus lingüísticos.
- dictionaries:** en él encontramos los elementos necesarios para codificar cualquier tipo de recurso léxico como, por ejemplo, diccionarios o glosarios.
- drama:** diseñado para codificar obras teatrales, guiones cinematográficos o radiofónicos, entre otros.
- figures:** orientado a describir gráficos, tablas, imágenes, obras de arte, fórmulas matemáticas, etc., que forman parte de los documentos.
- gaiji:** para codificar glifos y otros caracteres especiales, tipos de escrituras (vertical, horizontal, de derecha a izquierda, etc.), etc.
- header:** proporciona metadatos descriptivos sobre el recurso codificado (información bibliográfica, menciones de responsabilidad, información sobre el proyecto e investigadores y centros participantes, información de aspectos no bibliográficos como, por ejemplo, materias o códigos de clasificación, el historial de revisión de la codificación, etc.).
- iso-fs:** permite representar las interrelaciones entre piezas de información, proporcionando su instanciación en el marcado, un metalenguaje para representaciones genéricas de los análisis e interpretaciones realizados.
- linking:** ofrece diferentes elementos que permiten representar relaciones entre partes de los documentos a partir de identificadores y enlaces.
- msdescription:** define elementos pensados para proporcionar información descriptiva detallada sobre cualquier tipo de texto manuscrito. Inicialmente, fue concebido para satisfacer las necesidades de los catalogadores y académicos que trabajaban con manuscritos medievales de tradición europea, aunque actualmente se puede extender su uso a otras tradiciones y materiales.
- namesdates:** un módulo pensado para codificar fechas, nombres y otros identificadores de personas, lugares y organizaciones.
- nets:** permite codificar representaciones gráficas utilizadas para visualizar las relaciones que se establecen entre unidades de información. Se prevé el uso de grafos y mapas conceptuales.
- spoken:** orientado a describir transcripciones de cualquier tipo de discurso oral.
- tagdocs:** ofrece un conjunto de elementos que pueden ser utilizados para documentar los elementos XML. Muy pensado para los proyectos en los cuales se personaliza o modifica el núcleo o alguno de los módulos del estándar.
- tei:** contiene las declaraciones de todos los tipos de datos y las declaraciones iniciales de las clases de atributos, modelos y las macros utilizadas por el resto de módulos del esquema.
- textcrit:** permite crear aparatos críticos propios de la crítica textual.
- textstructure:** un módulo obligatorio que contiene la estructura de alto nivel por defecto de cualquier documento TEI.
- transcr:** pensado para representar fuentes primarias, como manuscritos, cartas o diarios.
- verse:** diseñado para codificar textos escritos total o parcialmente en verso.

A continuación, se muestra un ejemplo comentado del marcado de un pequeño fragmento de la obra de Oscar Wilde, *The Importance of Being Earnest*.⁴



(<http://bid.uib.edu/sites/bid.uib.edu/files/37/Alcaraz1bg.jpg>)

Figura 1. Ejemplo de codificación de un texto dramático

4.2 El sistema de clases y atributos

Las declaraciones de los 500 elementos que forman el esquema de la TEI también incluyen la asignación a una o más clases de elementos que sirven para expresar diferentes tipos de relaciones entre éstos. Cuando dos o más elementos comparten uno o varios atributos, se organizan bajo la misma clase de atributo (*attribute class*), mientras que si lo que tienen en común es la parte del documento en la que pueden aparecer, entonces los encontraremos dentro de una misma clase de modelo (*model class*). Por ejemplo, todos los elementos que pueden aparecer dentro de un elemento <div> los encontraremos en la clase "model.divPart". Las clases se organizan jerárquicamente, y dan lugar a las llamadas *subclases* y *superclases*. Los elementos que comparten una misma rama del árbol heredan la posibilidad de aparecer en aquellas partes del documento donde también lo puede hacer cualquiera de los elementos de sus superclases superiores en la jerarquía.

<sp>		Principal C Elementos
<sp> (discurso) Un discurso individual de un texto dramático, o un pasaje en prosa o en verso presentado como tal. [3.12.2 Core Tags for Drama 3.12 Passages of Verse or Drama 7.2.2 Speeches and Speakers]		
Módulo	core — Elements Available in All TEI Documents	
Atributos	att.global (@xml:id, @n, @xml:lang, @xml:base, @xml:space) (att.global.rendition (@rend, @style, @rendition)) (att.global.linking (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) (att.global.analytic (@ana)) (att.global.facs (@facs)) (att.global.change (@change)) (att.global.responsibility (@cert, @resp)) att.ascribed (@who)	
Miembro de	model.divPart	
Contenido en	core: item note q quote said stage drama: castList epilogue performance prologue set spGrp view figures: cell figure header: change handNote licence scriptNote typeNote msdescription: accMat acquisition additions collation condition custEvent decoNote filiation foliation layout musicNotation origin provenance signatures source summary support surrogates tagdocs: specGrp textcrit: lem rdg textstructure: argument body div div1 div2 div3 div4 div5 div6 div7 epigraph postscript transcr: metamark	
Puede contener	analysis: interp interpGrp span spanGrp certainty: certainty precision respons core: cb cit gap gb index l lb lg list milestone note p pb q quote said speaker stage drama: camera caption move sound tech view figures: figure notatedMusic table iso-fs: fLib fs fvLib linking: ab alt altGrp anchor join joinGrp link linkGrp timeline namesdates: listEvent listNym listOrg listPerson listPlace spoken: incident kinesic pause shift vocal writing textcrit: app listApp listWit witDetail textstructure: floatingText transcr: addSpan damageSpan delSpan fw listTranspose metamark space substJoin	

Figura 2. Fragmento de la descripción del elemento <sp> (*speech*) en la versión en línea de las directrices TEI

5 Personalización del esquema de la TEI

El estándar TEI pretende cubrir un amplio abanico de usos. Para poder cumplir con esta premisa, resulta totalmente necesario disponer de un lenguaje flexible y extensible que permita personalizaciones según las necesidades de cada proyecto. Esto es posible gracias a la estructura basada en módulos y clases presentada anteriormente, y al uso mismo de XML.

En la web de la TEI, encontramos diferentes ejemplos de estas personalizaciones, basadas en la combinación de algunos módulos "oficiales".⁵ Por ejemplo, la TEI Lite incorpora los elementos básicos para codificar documentos simples; la TEI Drama se orienta a describir obras teatrales, o el jTEI Article, es una personalización enfocada a la creación de artículos de revista, y utilizada en el *Journal of the Text Encoding Initiative* (<http://journal.tei-c.org/>) mantenido por el mismo Consorcio TEI. Otras líneas de investigación han derivado en desarrollos como el *CBML* (<http://dcl.slis.indiana.edu/cbml/>) (Comic Book Markup Language) pensado para codificar cómics (Walsh, 2012).

Para facilitar la personalización del esquema, la TEI ofrece una herramienta llamada *Roma* (<http://www.tei-c.org/Roma/>), un asistente que paso a paso y a través de sencillos formularios, nos permite generar nuestro propio esquema. Las personalizaciones del estándar TEI se conocen informalmente como las *ODD* (*One Document Does it All*).

Como acabamos de comentar, el uso del asistente es extremadamente sencillo. Sólo debemos decidir con qué conjunto de elementos queremos comenzar a trabajar: una versión mínima con los cuatro módulos obligatorios, la TEI Lite, la TEI All con todos los módulos o, incluso, una versión personalizada que podemos cargar desde nuestro ordenador.

TEI Roma: generating customizations for the TEI

TEI Roma is a tool for working with TEI customizations. A TEI customization is a document from which you can generate a system you want to use, along with customized HTML or PDF documentation of it. The schema generated can be expressed in several languages.

You can make or modify your TEI customization in several different ways:

- Build up: create a new customization by adding elements and modules to the smallest recommended schema
- Reduce: create a new customization by removing elements and modules from the largest possible schema
- Create a new customization starting from a template
- Use or modify an existing TEI-defined customization
- Upload a customization Ningún archivo seleccionado

Community-maintained customizations can be downloaded from [the TEI website](#)

Start

Figura 3. Pantalla de inicio de la herramienta Roma

A continuación, el sistema nos pedirá una serie de metadatos: el título del esquema, nombre del fichero, espacio de nombres, nombre del autor, etc.

TEI Roma: generating validators for the TEI

Set your parameters

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

Set your parameters

Title

Filename

Namespace for new elements

Prefix for TEI pattern names in schema

Language English, Deutsch, Italiano, Español, Français, Portugues, Russian, Svenska, 日本語, 中文

Author name

Description

Save

Figura 4. Metadatos necesarios para identificar el nuevo esquema

La pestaña "Módulos" nos permite añadir y eliminar módulos de nuestro esquema. También es posible acceder a cada uno de los módulos para incluir o excluir del esquema alguno de los diferentes elementos que lo forman.

TEI Roma: generating validators for the TEI

Modules

New Customize Language **Modules** Add Elements Change Classes Schema Documentation Save Customization

List of TEI Modules			List of selected Modules
Module name	A short description	Changes	
add analysis	Simple analytic mechanisms		remove tei
add certainty	Certainty and uncertainty		remove core
add core	Elements common to all TEI documents	changed	remove header
add corpus	Corpus texts		remove textstructure
add dictionaries	Dictionaries		remove figures
add drama	Performance texts		remove linking
add figures	Tables, formulæ, notated music, and figures		remove analysis
add gajji	Character and glyph documentation		remove tagdocs
add header	The TEI Header	changed	
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		

Figura 5. Selección de los módulos para el nuevo esquema

Una vez seleccionados los módulos y elementos disponibles, podemos crear algunos nuevos, en el caso de que los necesitemos. Para hacerlo, debemos asociar cada nuevo elemento a una clase de modelo y a una de atributo de las existentes. Dicho esto, es importante destacar que la incorporación de un exceso de nuevos elementos, comportará trabajar al margen del estándar, práctica que se debe limitar para asegurar la interoperabilidad entre proyectos.

TEI Roma: generating validators for the TEI

Add Element

New Customize Language Modules **Add Elements** Change Classes Schema Documentation Save Customization

[go back to list](#)

Defining a new element:

Name:

Namespace:

Description:

Model classes:

- model.addrPart
- model.addressLike
- model.applicationLike
- model.biblLike
- model.biblPart
- model.castitemPart
- model.certLike
- model.choicePart
- model.common
- model.correspActionPart
- model.correspContextPart
- model.correspDescPart
- model.desclike
- model.dimLike
- model.div1Like
- model.div3Like
- model.div4Like
- model.div5Like
- model.div7Like
- model.divBottom
- model.divBottomPart
- model.divLike
- model.divPart
- model.divPart spoken

Figura 6. Creación de nuevos elementos para el esquema

Con nuestro esquema personalizado ya finalizado, podemos descargar cualquiera de los sistemas de validación disponibles: Relax NG, Schematron, XML Schema o DTD (Document Type Definition).

TEI Roma: generating validators for the TEI

Time to give you a schema

New Customize Language Modules Add Elements Change Classes **Schema** Documentation Save Customization

Creating a schema

Which format do you prefer?

Figura 7. Descarga del sistema de validación para el esquema XML personalizado

Finalmente, podemos descargar en diferentes formatos (HTML, Microsoft Word, LaTeX, etc.) la documentación de nuestro esquema con la descripción de cada uno de los elementos y atributos que lo conforman.

6 Campos de aplicación y estudio de casos

El estándar TEI extiende sus usos a los campos de las humanidades, las ciencias sociales y la lingüística, y destaca principalmente su utilización en centros de investigación, bibliotecas y archivos en la edición de textos para desplegar colecciones digitales y ediciones críticas; la descripción de manuscritos y cartas medievales; la creación de corpus lingüísticos y diccionarios; el mantenimiento de datos estructurados para crear bases de datos para almacenar listas de personajes, cronologías o colecciones de objetos, o, incluso, para usos editoriales.

A continuación, se describen dos proyectos representativos en estas áreas.

6.1 ReMetCa: repertorio métrico digital de la poesía medieval castellana

ReMetCa (<http://www.remetca.uned.es/>) es un repertorio digital que recoge poemas medievales con el objetivo de construir una herramienta para clasificar, definir y delimitar el corpus poético castellano (González, 2013). El proyecto, financiado por la UNED, nace en el año 2011 a partir de la necesidad de crear un instrumento de estudio centrado en la métrica de las primeras manifestaciones líricas de la edad media. De esta manera, los estudios sobre la métrica medieval castellana se suman a los de otras tradiciones literarias que ya tienen, desde hace años, repertorios métricos digitales. Este es el caso de Galicia (**MedDB** (<http://www.cirp.es/meddb>)), Francia (**Nouveau Naetebus** (<http://nouveauaetebus.elte.hu/>)) o los Países Bajos (**Dutch Song DB** (<http://www.liederenbank.nl/index.php?lan=en>)). De hecho, hace ya algunos años, se empezó a proyectar un metabuscador que, bajo el nombre de *Megarep*, pretende unificar los diferentes repertorios métricos nacionales con el objetivo de crear una herramienta a nivel europeo que permita consultar, desde un único portal, todos estos recursos (González; Seláf, 2013).

Por lo que respecta al apartado técnico, ReMetCa utiliza el estándar TEI para codificar los versos. En el caso concreto de este proyecto, se ha desarrollado más ampliamente el módulo "verse", con etiquetas específicas para describir los fenómenos métricos (González [et al.], 2014). El uso del estándar TEI les ha permitido desarrollar un potente instrumento para llevar a cabo un análisis detallado de los versos, así como todo tipo de búsquedas que facilitarán posteriores estudios métricos y literarios comparativos.

Finalmente, hay que destacar que en este proyecto, el estándar TEI funciona conjuntamente con un sistema de gestión de bases de datos MySQL encargado del almacenaje y recuperación de la información.

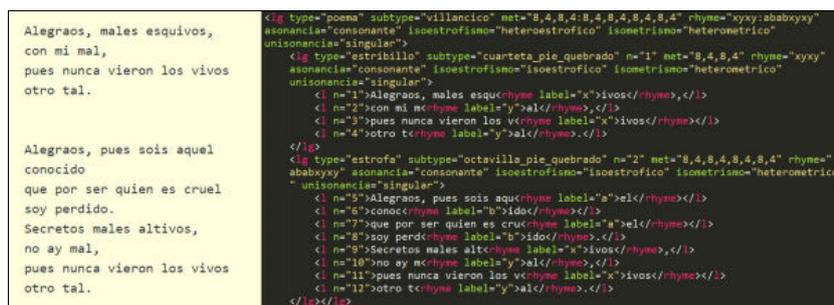


Figura 8. Fragmento de uno de los poemas disponibles en ReMetCA. En la parte derecha se puede observar el texto codificado con el estándar TEI

6.2 The World of Dante

The World of Dante (<http://www.worldofdante.org/>) es un proyecto del **Institute for Advanced Technologies in the Humanities** (<http://www.iath.virginia.edu/>) de la **University of Virginia** (<http://www.iath.virginia.edu/>), nacido con la finalidad de ofrecer un conjunto de herramientas multimedia e interactivas para profundizar en el estudio de *La Divina comedia* de Dante Alighieri.

Además de mapas interactivos, líneas de tiempo y galerías de imágenes relacionadas con la obra, el portal ofrece una edición del poema etiquetada con el estándar TEI que permite su lectura en un entorno hipertexto (Parker, 2001). La identificación y marcado de los diferentes elementos concretos dentro de la obra como, por ejemplo, los personajes, lugares, criaturas, divinidades o las estructuras arquitectónicas, permiten recuperar rápidamente esta información. De esta manera, si el lector o investigador selecciona cualquiera de estos temas, aparecerán destacados tipográficamente en el texto. Esta información resulta de gran interés para analizar detalladamente el poema o para realizar estudios minuciosos sobre determinadas figuras concretas. Por ejemplo, se puede consultar en qué versos del poema se nombra al personaje de Alejandro Magno o qué palabras, términos y expresiones utiliza Dante para nombrar y describir el Flegetonte, uno de los ríos del infierno.

The screenshot shows the 'THE WORLD OF DANTE' website interface. At the top, there are navigation tabs for 'INFERNO', 'PURGATORIO', and 'PARADISO'. Below these are sub-tabs for 'Gallery', 'Maps', 'Music', 'Timeline', 'Search', 'Teacher Resources', and 'About'. The main content area displays search results for 'INFERNO'. The results are organized into a table with columns for line numbers, categories (PE), and text. A sidebar on the right lists categories like People, Places, Creatures, Deities, Structures, Images, and Music.

Line Number	Category	Text
12.106	PE	Here they lament their ruthless crimes; here are both Alexander and the fierce Dionysius, who brought such years of grief to Sicily.
12.109	PE	That brow with hair so black is Ezzelino; that other there, the blonde one, is Obizzo of Este, he who was indeed undone.
12.112	PE	within the world above, by his fierce son." Then I turned to the poet, and he said: "Now let him be your first guide, me your second."
12.115		A little farther on, the Centaur stopped above a group that seemed to rise above the boiling blood as far up as their throats.
12.118	PE	He pointed out one shade, alone, apart, and said: "Within God's bosom, he impaled the heart that still drips blood upon the Thames."
12.121		Then I caught sight of some who kept their heads and even their full chests above the tide; among them-many whom I recognized.
12.124		And so the blood grew always shallower until it only scorched the feet, and here we found a place where we could ford the ditch.

Right sidebar categories:

- People: (PE)
- 12.17 Teseo
- 12.20 Ariadne
- 12.38 Cristo
- 12.66 Deianira
- 12.71 Achille
- 12.83 Virgilio
- 12.88 Beatrice
- 12.107 Alessandro Magn
- 12.107 Dionisio
- 12.110 Ezzelino III
- 12.111 Obizzo II
- 12.112 Azzo VIII
- 12.118 Guido di Montfort
- 12.120 Arrigo II
- 12.134 Attila
- 12.135 Pierro
- 12.135 Sesto
- 12.137 Rinieri da Cornet
- 12.137 Rinieri dei Pazzi
- +Places: (PL)
- +Creatures: (C)
- +Deities: (D)
- +Structures: (S)
- +Images: (I)
- +Music: (M)

Figura 9. El marcado de los personajes, lugares, criaturas, etc. permite recuperar esta información rápidamente

Además de la navegación por los diferentes cánticos de la *Divina comedia*, The World of Dante ofrece un buscador que permite recuperar los diferentes cánticos a partir de las categorías antes citadas. Gracias al marcado con TEI, cada figura concreta incorpora información adicional que permitirá afinar las búsquedas. Por ejemplo, en el caso de las personas se puede buscar por género, por el origen del personaje (histórico, mitológico o literario) o, incluso, por su afiliación política. La inclusión de estas categorías permite realizar búsquedas afinadísimas, que facilitan estudios muy concretos y detallados del texto literario.

7 Aplicación en los repositorios y bibliotecas digitales, y rol de los profesionales de la información

A pesar de ser estudiadas como dos realidades diferentes por algunos autores (Siemens [et al.], 2011), es evidente que muchos de los proyectos originados en el ámbito de la biblioteconomía y la documentación pueden etiquetarse como proyectos de humanidades digitales. La relación entre estas dos disciplinas no se limita al campo de actuación, sino que el mismo perfil y competencias propios de los profesionales de la información pueden resultar muy interesantes para cualquier instituto o centro de investigación en humanidades digitales, tanto en tareas de soporte a la investigación, como al diseño e implementación de los proyectos que derivan de éstas (Rodríguez Yunta, 2013).

Por lo que respecta a la aplicación del estándar TEI en las bibliotecas digitales, el mismo Consorcio TEI mantiene un documento de trabajo que recoge un conjunto de buenas prácticas para codificar textos en este tipo de proyectos (Hawkins; Dalmau; Bauman, 2011). Estas directrices fueron creadas para utilizarse en proyectos de digitalización de grandes colecciones bibliotecarias, pero pueden ser útiles para cualquier trabajo de digitalización y codificación de documentos. Entre otras informaciones de interés, incluyen un mapeo entre los elementos de la cabecera de un documento TEI y las etiquetas MARC correspondientes.

Un buen ejemplo de la aplicación de estas recomendaciones lo encontramos en la incorporación del estándar TEI en los flujos de trabajo para producir libros digitales de la Biblioteca Virtual Miguel de Cervantes. Esta práctica permite generar libros digitales en diferentes formatos a partir del documento XML codificado con TEI (Bia Platas; Sánchez Quero, 2001).

La estrecha relación entre la cabecera de un documento TEI y los datos de un registro MARC también es protagonista de proyectos como los de la Biblioteca de la University of Michigan, en la que se crearon diferentes *scripts* en el lenguaje de programación Perl para extraer de manera automatizada los valores de las etiquetas MARC y traspasarlos a los elementos TEI correspondientes (Marko; Powell, 2001).

Otro ejemplo del uso del estándar TEI en una biblioteca digital lo encontramos en la University of Virginia Library. En este caso, se han integrado en el catálogo general de la biblioteca algunas **colecciones digitales marcadas con TEI** (http://search.lib.virginia.edu/catalog?f%5bdigital_collection_facet%5d%5b%5d=UVa+Text+Collection&facet.limit=500&id=digital_collection_facet&sort=date_received_facet+desc&width=490). En el marco de este proyecto, el Digital Production Group de la biblioteca desarrolló unas directrices propias⁶ que son la base para el trabajo de codificación de estos recursos.

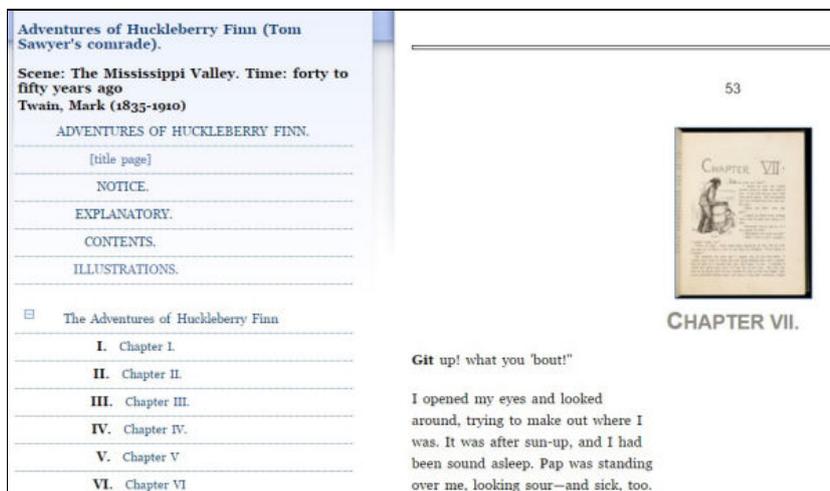


Figura10. Edición digital de la obra *Adventures of Huckleberry Finn* disponible al catálogo VIRGO de la University of Virginia Library

También destaca el trabajo que muchas bibliotecas académicas están realizando en la formación de la comunidad universitaria en el uso del estándar TEI. Encontramos un buen ejemplo en la University of California, en la que son los mismos bibliotecarios los que enseñan a profesores y estudiantes el estándar TEI, para que éstos puedan desarrollar proyectos de investigación en el ámbito de las humanidades digitales (Green, 2014). A demanda de los interesados, los bibliotecarios especialistas elaboran pequeños talleres y cursos a medida. Fruto de esta colaboración, se han originado proyectos dentro de la universidad como la [UCLA Encyclopedia of Egyptology \(http://uee.ucla.edu/\)](http://uee.ucla.edu/).

8 Conclusiones

A pesar de que el estándar TEI hace casi tres décadas que se aplica, el apogeo de las humanidades digitales en los últimos años ha renovado el interés en este y otros estándares entre estudiosos de las humanidades y profesionales de la información y la documentación (Zhang; Liu; Mathews, 2015). No obstante, actualmente son pocos los proyectos de digitalización en España que han optado por enriquecer sus objetos digitales con un marcado como el propuesto por el Consorcio TEI. En este sentido, tal y como explica Rodríguez Yunta (2014), en nuestro territorio se ha antepuesto la cantidad por encima de la calidad, acumulándose en los repositorios y bibliotecas digitales, documentos en formato PDF sin estructurar o, simplemente, archivos de imagen sin, ni siquiera, una versión derivada de un proceso de reconocimiento óptico de caracteres (OCR). Proyectos como [TESORO \(http://www.bib.uc3m.es/~nogales/xml/tesoro/\)](http://www.bib.uc3m.es/~nogales/xml/tesoro/) (Edición electrónica del Teatro Español del Siglo de Oro para la difusión del español y la formación a distancia) o la [Biblioteca Virtual Miguel de Cervantes \(http://www.cervantesvirtual.com/\)](http://www.cervantesvirtual.com/), son la excepción a la práctica más extendida. La utilización de estos sistemas de marcado aumentaría el uso efectivo de los recursos disponibles en los repositorios de bibliotecas y archivos, al facilitar su procesamiento automatizado, una de las principales líneas de trabajo e investigación en el área de las humanidades digitales (Schmidt, 2012).

Bibliografía

- Allés Torrent, Susanna (2015). *Introducción a la edición digital de textos: TEI-XML*. <<http://susannalles.github.io/Web-TEI/1.1.html> (<http://susannalles.github.io/Web-TEI/1.1.html>) >. [Consulta: 25/06/2016].
- Bia Platas, Alejandro; Sánchez Quero, Manuel (2001). "Diseño de un procedimiento de marcado para la automatización del procesamiento de textos digitales usando XML y TEI". En: De-la-Fuente, Pablo; Pérez, Adoración (ed.). *II Jornadas bibliotecas digitales (Jbidi)*, Almagro, 19 y 20 de noviembre. Alicante: Biblioteca Virtual Miguel de Cervantes, p. 153–165. <<http://www.biblioteca.org.ar/libros/142321.pdf> (<http://www.biblioteca.org.ar/libros/142321.pdf>) >. [Consulta: 27/09/2016].
- Baena Sánchez, Francisco [et al.] (2014). "Codificación y representación cartográfica de noticias: aplicación de las humanidades digitales al estudio del periodismo de la Edad moderna". *El profesional de la información*, vol. 23, nº 5 (septiembre-octubre), p. 519–526.
- González, Elena (2013). "Actualidad de las humanidades digitales y un ejemplo de ensamblaje poético en red: ReMetCa". *Cuadernos hispanoamericanos*, nº 761 (noviembre), p. 53–67.
- González, Elena [et al.] (2014). "Una propuesta de integración del sistema de formularios de bases de datos MySQL con etiquetado TEI: ReMetCa, Repertorio digital de la métrica medieval castellana". En: López Poza, Sagrario; Pena Sueiro, Nieves (ed.). *Humanidades digitales: desafíos, logros y perspectivas de futuro*. A Coruña: SIELAE-JANUS. <<http://www.janusdigital.es/anexo.htm?id=5> (<http://www.janusdigital.es/anexo.htm?id=5>) >. [Consulta: 25/06/2016].
- González, Elena; Seláf, Levente (2013). "Megarep: a comprehensive research tool in medieval and renaissance poetic and metrical repertoires". En: Soriano, Lourdes [et al.] (ed.). *Humanitats a la xarxa*:

món medieval = *Humanities on the web: the medieval world*. Bern [etc.]: Peter Lang, p. 333–344.

Green, Harriett E. (2014). "Facilitating communities of practice in digital humanities: librarian collaborations for research and training in text encoding". *Library quarterly: information, community, policy*, vol. 84, no. 2, p. 219–234.

Hawkins, Kevin; Dalmau, Michelle; Bauman, Syd (ed.) (2011). *Best practices for TEI in libraries*. Version 3.0 (October). <<http://www.tei-c.org/SIG/Libraries/teinlibraries/main-driver.htm> (<http://www.tei-c.org/SIG/Libraries/teinlibraries/main-driver.html>) >. [Consulta: 27/06/2016].

Marko, Lynn; Powell, Christina (2001). "Descriptive metadata strategy for TEI headers: a University of Michigan Library case study". *OCLC systems & services: international digital library perspectives*, vol. 17, no. 3, p. 117–21.

Nogales Flores, J. Tomás [et al.](2003). "Una experiencia de aplicación de XML y TEI a obras teatrales del Siglo de Oro español". En: *8as Jornadas españolas de documentación. Los sistemas de información en las organizaciones: eficacia y transparencia*. Barcelona, 6–8 de febrero. Barcelona: FESABID, p. 395–404. <<http://hdl.handle.net/10016/906> (<http://hdl.handle.net/10016/906>) >. [Consulta: 26/06/2016].

Parker, Deborah (2001). "The World of Dante: a hypermedia archive for the study of the Inferno". *Literary and linguistic computing: Journal of the Association for Literary and Linguistic Computing*, vol. 16, no. 3, p. 287–297. <<http://llc.oxfordjournals.org/content/16/3/287.abstract> (<http://llc.oxfordjournals.org/content/16/3/287.abstract>) >. [Consulta: 26/06/2016].

Renear, Allen H. (2004). "Text Encoding". En: Schreibman, Susan; Siemens, Ray; Unsworth, John (ed.). *A companion to digital humanities*. Oxford: Blackwell, p. 218–239. <<http://www.digitalhumanities.org/companion/> (<http://www.digitalhumanities.org/companion/>) >. [Consulta: 26/06/2016].

Rodríguez Yunta, Luis (2013). "Humanidades digitales, ¿una mera etiqueta o un campo por el que deben apostar las ciencias de la documentación?". *Anuario ThinkEPI*, vol. 7, p. 37–43. <<http://hdl.handle.net/10261/77511> (<http://hdl.handle.net/10261/77511>) >. [Consulta: 26/06/2016].

— (2014). "Ciberinfraestructura para las humanidades digitales: una oportunidad de desarrollo tecnológico para la biblioteca académica". *El profesional de la información*, vol. 23, n.º 5 (septiembre-octubre), p. 453–462.

Romary, Laurent (2009). "Questions & answers for TEI newcomers". *Jahrbuch für Computerphilologie*, nr. 10. <<http://arxiv.org/abs/0812.3563> (<http://arxiv.org/abs/0812.3563>) >. [Consulta: 24/06/2016].

Schmidt, Desmond (2012). "The role of markup in the digital humanities". *Historical social research*, vol. 37, no. 3, p. 125–146. <<http://nbn-resolving.de/urn:nbn:de:0168-ssoar-378369> (<http://nbn-resolving.de/urn:nbn:de:0168-ssoar-378369>) >. [Consulta: 27/06/2016].

Siemens, Lynne [et al.](2011). "A tale of two cities: implications of the similarities and differences in collaborative approaches within the digital libraries and digital humanities communities". *Literary and linguistic computing*, vol. 26, no. 3, p. 335–348.

TEI Consortium (ed.) (2016). *TEI P5: guidelines for electronic text encoding and interchange*. Version 3.0.0. [S.l.]: Text Encoding Initiative Consortium. <<http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> (<http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>) >. [Consulta: 25/06/2016].

Walsh, John A. (2012). "Comic Book Markup Language: an introduction and rationale". *Digital humanities quarterly*, vol. 6, no. 1. <<http://www.digitalhumanities.org/dhq/vol/6/1/000117/000117.html> (<http://www.digitalhumanities.org/dhq/vol/6/1/000117/000117.html>) >. [Consulta: 25/06/2016].

Zhang, Ying; Liu, Shu; Mathews, Emilee (2015). "Convergence of digital humanities and digital libraries". *Library management*, vol. 36, no. 4–5, p. 362–377.

Notas

¹ La lista completa de miembros se puede consultar en: <<http://members.tei-c.org/Directory> (<http://members.tei-c.org/Directory>) >. [Consulta: 30/06/2016].

² Es posible encontrar más información en el web de la TEI, en la sección "TEI: Goals and Mission". <<http://www.tei-c.org/About/mission.xml> (<http://www.tei-c.org/About/mission.xml>) >. [Consulta: 30/06/2016].

³ Esta voluntad ya se pone de manifiesto en el prólogo de la primera versión publicada de las directrices del estándar en el año 1994.

⁴ Comentarios de los autores del artículo a partir de los ejemplos disponibles en: <<http://teibyexample.org/> (<http://teibyexample.org/>) >. [Consulta: 30/06/2016].

⁵ <<http://www.tei-c.org/Guidelines/Customization/> (<http://www.tei-c.org/Guidelines/Customization/>) >. [Consulta: 30/06/2016].

⁶ <<http://dcs.library.virginia.edu/digital-stewardship-services/tei-encoding-guidelines/> (<http://dcs.library.virginia.edu/digital-stewardship-services/tei-encoding-guidelines/>) >.

[Consulta: 30/06/2016].

Cita recomendada

Alcaraz Martínez, Rubén; Vázquez Puig, Elisabet (2016). "TEI : un estándar para codificar textos en el ámbito de las humanidades digitales". *BID: textos universitaris de biblioteconomia i documentació*, núm. 37 (deseembre) . <<http://bid.ub.edu/es/37/alcaraz.htm>>. [Consulta: 15-12-2016].

