



Texts, Languages and Communities

24th annual Conference and Members' Meeting of the Text Encoding Initiative Consortium

Textos, lenguas y comunidades Vigésimo cuarto congreso de la Text Encoding Initiative

Gimena del Río Riande, ed.



Text
Encoding
Initiative



Texts, Languages and Communities: TEI 2024. Book of Abstracts

Textos, lenguas y comunidades: TEI 2024. Libro de resúmenes

7-11 October / 7-11 de octubre, Universidad del Salvador, Ciudad
de Buenos Aires, Argentina

Texts, languages and communities : 24th Annual Conference and Members' Meeting of the Text Encoding Initiative Consortium = Textos, lenguas y comunidades : vigésimo cuarto congreso de la Text Encoding Initiative Consortium / Gustavo Fernández Riva. [et al.] ; Editado por Gimena Del Río Riande. - 1a edición bilingüe - Ciudad Autónoma de Buenos Aires : Universidad del Salvador, 2024.

Libro digital, PDF

Archivo Digital: descarga y online
Edición bilingüe : Español ; Inglés.
ISBN 978-950-592-316-8

1. Ciencias Sociales y Humanidades. 2. Tecnología Informática. 3. Metadatos. I. Fernández Riva, Gustavo II. Del Río Riande, Gimena, ed.
CDD 306.46

Workshops / Talleres

Sistemas de HTR, ediciones digitales y acceso al patrimonio textual: taller sobre Transkribus (READ Coop) y EVT (Edition Visualization Technology)

Stefano Bazzaco

Universidad de Verona (Italia)

Los sistemas de HTR están actualmente al centro de los intereses de los humanistas porque permiten acrecentar el acceso al patrimonio textual y su procesamiento con herramientas computacionales. Alcanzando resultados de transcripción automática con documentos impresos que se acercan a un margen de error del 1%, estas aplicaciones permiten la exportación del texto en distintos formatos, entre los cuales XML. Por consiguiente, es posible establecer un flujo de trabajo que, partiendo de las digitalizaciones del documento fuente, lleve a la conversión del contenido en texto electrónico, su posterior maquetación según el estándar TEI y su publicación web con aplicaciones de visualización.

El taller que se propone se inserta en el contexto de los estudios sobre transcripción automática y edición digital y tiene como objetivo introducir los participantes al uso de la plataforma Transkribus (READ Coop), actualmente el sistema de HTR más fiable y que cuenta con la comunidad de usuarios más extendida. Entre las finalidades del taller subsiste la idea de que el uso consciente de Transkribus pueda proporcionar informaciones muy valiosas para el filólogo digital, impulsando la creación de ediciones digitales académicas de documentos impresos de forma simplificada.

HTR systems, digital editions and access to textual heritage: workshop on Transkribus (READ Coop) and EVT (Edition Visualization Technology)

Stefano Bazzaco

University of Verona (Italy)

HTR systems are currently the focus of interest for humanists as they enable increasing access to textual heritage and processing with computational tools. These applications facilitate the process of exporting text in different formats including XML and reach automatic transcription results of printed documents with a 1% margin of error. Consequently, it is possible to establish a workflow that, based on the digitalization of the source document, leads to the conversion of the content into electronic text, the

subsequent layout according to the TEI standard and the web publication with visualization applications.

This workshop is inserted in the context of studies concerning automatic transcription and digital editing, and its objective is to introduce participants to the use of the Transkribus (READ Coop) platform, which, at present, is considered to be the most reliable HTR system and is home to the largest user community. One of the purposes of the workshop is to show that making sensible use of Transkribus can provide valuable information for the digital philologist, thus, fostering the creation of scholarly digital editions of printed documents in a simplified way.

Introducción a la codificación de música utilizando el formato de la Music Encoding Initiative (MEI)

Martha Eladia María Thomae Elias

Universidade NOVA de Lisboa (Portugal)

La comunidad de la Music Encoding Initiative (MEI) se inspiró en la de la Text Encoding Initiative (TEI). La comunidad MEI trabaja en desarrollar las mejores prácticas para la codificación de diversos documentos musicales, respondiendo a las distintas necesidades de bibliotecas, archivos musicales, musicólogos, etc. El formato MEI, al igual que el TEI, se basa en etiquetas XML. Dichas etiquetas definen los distintos elementos de codificación musical por medio de sus atributos.

En este taller, la audiencia será introducida al formato MEI, su estructura básica y las etiquetas utilizadas para la codificación de elementos musicales como notas, silencios, compases, claves, etc. También codificaremos juntos nuestro primer archivo MEI usando la webapp de mei-friend (<https://mei-friend.mdw.ac.at/>). Mei-friend puede ser utilizado en cualquier explorador de internet (con la excepción de Safari). La aplicación tiene diferentes paneles, incluido un panel de edición donde el usuario puede ingresar el código MEI y un panel de visualización donde el usuario puede ver la música desplegada por Verovio (herramienta de visualización de MEI).

Además de codificar, la audiencia será introducida a varias de las herramientas desarrolladas alrededor de MEI, algunos de los usos de este formato, y lugares donde encontrar más documentación sobre MEI para poder explorar en el futuro (las Directrices o “Guidelines” de MEI y tutoriales disponibles en español).

Introduction to music encoding using the Music Encoding Initiative (MEI) format

Martha Eladia María Thomae Elias

University NOVA of Lisbon (Portugal)

The Music Encoding Initiative community (MEI) was inspired by the Text Encoding Initiative (TEI). The MEI community works to develop best practices for the encoding of various musical documents, responding to the different needs of libraries, music archives, musicologists, etc. The MEI format, as well as TEI, is based on XML tags which define the different musical encoding elements by their attributes.

In this workshop, the audience will be introduced to the MEI format, its basic structure and the tags used for encoding musical elements such as notes, rests, bars, clefs, etc. We will also encode our first MEI file together using the mei-friend webapp. (<https://mei-friend.mdw.ac.at/>). Mei-friend can be accessed using any internet browser (with the exception of Safari). The application has different panels, including an editing panel, where the user can enter the MEI code, and a display panel, where the user can view the music displayed by Verovio (MEI display tool).

In addition to encoding, the audience will be introduced to several tools developed around MEI, some of the uses of this format, and places where more documentation on MEI can be found so as to explore in the future (the MEI Guidelines and tutorials available in Spanish).

Producing Semantic Digital Editions Using LEAF Commons Tools

Diane Jakacki¹, James Cummings², Gabriel Calarco³

¹Bucknell University (United States of America), ²Newcastle University (United Kingdom), ³Universidad de Buenos Aires (Argentine Republic)

This half-day in-person workshop introduces textual scholars and practitioners to the LEAF Commons suite of tools, a set of web-based, easy-to-use tools that support text encoding, named entity recognition, web annotation, text analysis and publication without users having to learn complex encoding languages, and the easy movement from one interoperable tool to the other depending on users' needs. These freely available tools support digital scholarly workflows for the collaborative production and publication of scholarly and documentary texts, editions, and collections on the web, without the need for software installation, while promoting best practices for text encoding, annotation, and metadata standards. The LEAF Commons suite enables the use of individual tools for specific purposes, as well as supporting an end-to-end workflow beginning with outputs of optical character or handwritten text recognition

systems, transcriptions, or born-digital texts and ends with publication on the web, allowing it to serve a wide range of research and pedagogical uses.

LEAF stands for the Linked Editing Academic Framework, a collaborative software suite that provides both a comprehensive virtual research environment but also a set of 'Commons' modular tools for text editing and publication. The LEAF Commons tools constitute an accessible, low-barrier, no-cost infrastructure for the production of online texts, editions, or collections, whether for teaching or for undertaking research and collaboration on a sustainable basis. The Commons makes LEAF tools freely available in the browser, enabling collaboration and publication through Github, in addition to local storage. Promoting the reuse of data in keeping with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, LEAF uses open-source software, open-access platforms, and open international standards for best practices in text-encoding (TEI-XML) and web annotation (RDF). LEAF Commons offers communities of researchers, teachers, and students the opportunity to take part in digital knowledge production and open collaboration. The workshop will end with an open discussion about pursuing such forms of open knowledge production and collaboration.

The LEAF Commons tools introduced will include:

- LEAF-Writer: an open-source, open-access Extensible Markup Language (XML) editor that runs in a web browser and offers scholars and their students a rich textual editing experience without the need to download, install, and configure proprietary software, pay ongoing subscription fees, or learn complex coding languages. This user-friendly editing environment incorporates Text Encoding Initiatives (TEI) and Resource Description Framework (RDF) standards, meaning that texts edited in LEAF-Writer are interoperable with other texts produced by the scholarly editing community and with other materials produced for the Semantic Web. It also incorporates Named Entity Recognition and reconciliation with, or linking to, linked open data identifiers through the incorporation of the NERVE tool. LEAF-Writer is particularly valuable for pedagogical purposes, allowing instructors to teach students best practices for encoding texts without also having to teach students how to code. LEAF-Writer is designed to help bridge the gap by providing access to all who want to engage in new and important forms of textual production, analysis, and discovery.
- DToC: (the Dynamic Table of Contexts) provides an online interactive reading and publication environment for digital scholarly texts where the two conventional overviews provided in print editions - the table of contents and the index - have been dynamically merged to provide an interactive online e-reading experience that leverages the power of XML markup. Users can build a DToC edition from

one or more TEI-XML files, then curate and label the underlying elements and attributes in order to understand where named entities, topics, and concepts can be traced within the edition. Editions can be stored using URLs and shared with readers as published or teaching texts.

- LEAF-TE: (The LEAF Turning Engine) is a web interface that enables users to easily and automatically transform documents between formats. It converts HTR/OCR output (from various sources including Trankribus) to TEI-XML for importing into LEAF-Writer or other editors.) It converts TEI-XML to HTML, Markdown, and plain text for exporting to web publishing and text analysis environments, including the Dynamic Table of Contexts.
- NERVE: (the Named Entity Relationship and Vetting Environment) is an application that performs Named Entity Recognition (NER) on machine-readable texts, allowing users to identify candidate entities in a document, review, and correct the results. NERVE suggests relevant Uniform Resource Identifiers (URIs) for entities, so users can reconcile data to an authority such as Wikidata or the Virtual International Authority File to provide the basis for Linked Open Data (LOD) Web Annotations. Users can export their reconciled data in TEI-XML or HTML formats to an online repository or to their desktop. NERVE will be able to be used from within LEAF-Writer, or as a stand-alone tool and will be demonstrated during the workshop.

Creación de ediciones digitales semánticas con las herramientas de LEAF Commons

Diane Jakacki¹, James Cummings², Gabriel Calarco³

¹Universidad de Bucknell (Estados Unidos de América), ²Universidad de Newcastle (Reino Unido), ³Universidad de Buenos Aires (República Argentina)

Este taller presencial de medio día presenta a los académicos y profesionales del texto el conjunto de herramientas LEAF Commons, un conjunto de herramientas basadas en la web y fáciles de usar que admiten la codificación de texto, el reconocimiento de entidades nombradas, la anotación web, el análisis y la publicación de texto sin que los usuarios tengan que aprender lenguajes de codificación complejos, y el fácil movimiento de una herramienta interoperable a otra dependiendo de las necesidades de los usuarios. Estas herramientas, disponibles de manera gratuita, apoyan los flujos de trabajo académicos digitales para la producción y publicación colaborativa de textos, ediciones y colecciones académicas y documentales en la web, sin necesidad de instalación de software, al tiempo que promueven las mejores prácticas para la codificación de texto, la anotación y los estándares de metadatos. La suite LEAF Commons permite el uso de herramientas individuales para propósitos específicos,

además de respaldar un flujo de trabajo de extremo a extremo que comienza con salidas de caracteres ópticos o sistemas de reconocimiento de texto escrito a mano, transcripciones o textos digitales y finaliza con la publicación en el sitio web, lo que le permite servir para una amplia gama de usos pedagógicos y de investigación.

LEAF significa Linked Editing Academic Framework, un paquete de software colaborativo que proporciona un entorno de investigación virtual integral pero también un conjunto de herramientas modulares "Comunes" para la edición y publicación de textos. Las herramientas LEAF Commons constituyen una infraestructura accesible, sin barreras y sin costo para la producción de textos, ediciones o colecciones en línea, ya sea para la enseñanza o para realizar investigaciones y colaboraciones de forma sostenible. The Commons hace que las herramientas LEAF estén disponibles gratuitamente en el navegador, lo que permite la colaboración y publicación a través de Github, además del almacenamiento local. Para promover la reutilización de datos de acuerdo con los principios de datos FAIR (encontrables, accesibles, interoperables y reutilizables), LEAF utiliza software de código abierto, plataformas de acceso abierto y estándares internacionales abiertos para las mejores prácticas en codificación de texto (TEI-XML) y anotación web (RDF). LEAF Commons ofrece a comunidades de investigadores, profesores y estudiantes la oportunidad de participar en la producción de conocimiento digital y la colaboración abierta. El taller finalizará con un debate abierto sobre la búsqueda de formas abiertas de producción y colaboración de conocimiento.

Las herramientas LEAF Commons introducidas incluirán:

- LEAF-Writer: un editor de lenguaje de marcado extensible (XML) de código abierto y acceso abierto que se ejecuta en un navegador web y ofrece a los académicos y a sus estudiantes una rica experiencia de edición de texto sin la necesidad de descargar, instalar y configurar software propietario, ni pagar tarifas de suscripción continuas o aprender lenguajes de codificación complejos. Este entorno de edición fácil de usar incorpora los estándares Text Encoding Initiatives (TEI) y Resource Description Framework (RDF), lo que significa que los textos editados en LEAF-Writer son interoperables con otros textos producidos por la comunidad de edición académica y con otros materiales producidos para la Web Semántica. . También incorpora el reconocimiento de entidades nombradas y la conciliación o vinculación con identificadores de datos abiertos vinculados mediante la incorporación de la herramienta NERVE. LEAF-Writer es particularmente valioso para fines pedagógicos, ya que permite a los instructores enseñar a los estudiantes las mejores prácticas para codificar textos sin tener que enseñarles también a codificar. LEAF-Writer está diseñado para ayudar a cerrar la brecha brindando acceso a todos los que quieran

participar en formas nuevas e importantes de producción, análisis y descubrimiento textual.

- DToC: (la tabla dinámica de contextos) proporciona un entorno interactivo de lectura y publicación en línea para textos académicos digitales donde las dos descripciones generales convencionales proporcionadas en las ediciones impresas (la tabla de contenido y el índice) se han fusionado dinámicamente para proporcionar una experiencia de lectura interactiva en línea que aprovecha el poder del marcado XML. Los usuarios pueden crear una edición DToC a partir de uno o más archivos TEI-XML, luego seleccionar y etiquetar los elementos y atributos subyacentes para comprender dónde se pueden rastrear entidades, temas y conceptos con nombre dentro de la edición. Las ediciones se pueden almacenar mediante URL y compartir con los lectores como textos publicados o didácticos.
- LEAF-TE: (The LEAF Turning Engine) es una interfaz web que permite a los usuarios transformar documentos entre formatos de forma fácil y automática. Convierte la salida HTR/OCR (de varias fuentes, incluido Trankribus) a TEI-XML para importar a LEAF-Writer u otros editores). Convierte TEI-XML a HTML, Markdown y texto sin formato para exportar a entornos de publicación web y análisis de texto, incluida la Tabla Dinámica de Contextos.
- NERVE: (el entorno de verificación y relación de entidades nombradas) es una aplicación que realiza el reconocimiento de entidades nombradas (NER) en textos legibles por máquina, lo que permite a los usuarios identificar entidades candidatas en un documento, revisar y corregir los resultados. NERVE sugiere Identificadores uniformes de recursos (URI) relevantes para entidades, de modo que los usuarios puedan conciliar datos con una autoridad como Wikidata o el Archivo de autoridad internacional virtual para proporcionar la base para las anotaciones web de datos abiertos vinculados (LOD). Los usuarios pueden exportar sus datos conciliados en formatos TEI-XML o HTML a un repositorio en línea o a su escritorio. NERVE podrá utilizarse desde LEAF-Writer o como herramienta independiente y se demostrará durante el taller.

Introduction to publishing TEI with static sites and front-end technologies

Raffaele Viglianti

University of Maryland (United States of America)

This half-day workshop will introduce strategies for handling TEI when publishing with static site generators and front-end technologies. The workshop will focus on

isomorphic approaches to publishing TEI on the web or, in other words, publishing TEI with little or no transformation, or with a structure-preserving mapping that allows working with the output as if it were the initial data source.

In particular, attendees will be introduced to CETELcean as a way of publishing TEI with minimal (or no) transformation (Cayless and Vigiante 2018). The main focus of the workshop will be learning how this approach can be used in conjunction with static site generators and will work on examples in “vanilla” JavaScript, React, and Gatsby. This workshop is aimed at attendees who already have some experience with programming (including XSLT) and the command line; however, all are welcome and will be supported as much as possible throughout the workshop.

A version of this workshop was previously given at the Text Encoding Initiative conference in 2022 and 2023. The 2023 workshop had an attendance of about 25 individuals, including grad students, faculty at various stages of their career, and research software developers. This version of the workshop will be lightly but significantly updated with a new template and new examples. Instruction will be in English with bilingual slides English and Spanish.

Digital humanities projects that result in the creation of digital output –typically a website– digital editions are prone to what Smithies et al. call the “digital entropy of software and digital infrastructure” (2019). Static sites have become a common choice for archiving legacy projects that risk going offline (Smithies et al., 2019; Summers, 2016) because they only require the absolute minimum from hosting infrastructure: a server to distribute documents at a given address. The sites themselves, once created, require no active maintenance and can be easily moved and transferred like any other collection of files. The Endings project at the University of Victoria, British Columbia (<https://endings.uvic.ca/>), for example, recommends static sites as a viable strategy for ensuring the longevity of Digital Humanities project publications. The Endings Principles for Digital Longevity include, among other strategies, the reduction of both software complexity and dependency on infrastructure. On the other hand, static sites cannot support features that would require an active server, such as large scale text search and user management; these features, therefore, are removed when projects are archived into static sites. Deriving static sites from an end-of-life project is the clear choice when access to infrastructure becomes limited. But this workshop addresses the question: What does it take to adopt static sites from the start?

After an introduction on static sites and the motivations for using them, the workshop will cover the following topics:

- Exploring isomorphic approaches: CETELcean and Verovio in “vanilla” JavaScript.
- Extend CETELcean with Web Components.

- Extend CETELcean with React.
- Using the static site generator Gatsby and the TEI plugins developed for the Scholarly Editing journal (gatsby-transformer-ceteicean and gatsby-theme-ceteicean).

In order to account for multiple levels of expertise, we may break into multiple groups for attendee-led collaborative work:

- Optional collaborative work: experiment with other frameworks that attendees may be familiar with (e.g. Vue, Angular, etc).
- Optional collaborative work: keeping within a single stack: applying optional transformations to XML documents in NodeJS (using JavaScript / TypeScript or XSLT via SaxonJS).

Introducción a la publicación de archivos TEI con sitios estáticos y tecnologías front-end

Raffaele Viglianti

Universidad de Maryland (Estados Unidos de América)

Este taller de medio día presentará estrategias para utilizar TEI en la publicación de archivos con generadores de sitios estáticos y tecnologías front-end. El taller se centrará en enfoques isomórficos para publicar archivos TEI en la web o, en otras palabras, publicar archivos TEI con poca o ninguna transformación, o con un mapeo que preserve la estructura y permita trabajar con el resultado como si fuera la fuente de datos inicial.

Particularmente, se presentará CETELcean, una librería abierta escrita en el lenguaje de programación JavaScript, que permite publicar archivos TEI con una transformación mínima o casi nula (Cayless y Viglianti, 2018). El objetivo principal del taller será aprender cómo se puede utilizar esta herramienta junto con generadores de sitios estáticos y se trabajará con ejemplos en Vanilla JavaScript, React y Gatsby. Si bien este taller está dirigido a asistentes que ya tengan cierta experiencia con la programación (incluido XSLT) y la línea de comandos, todos son bienvenidos y serán asistidos a lo largo del mismo.

Se impartió una versión de este taller durante las conferencias TEI llevadas a cabo en 2022 y en 2023. Asistieron al taller realizado en el año 2023 unas 25 personas, incluidos estudiantes de posgrado, profesores en diversas etapas de su carrera y desarrolladores de software de investigación. Esta versión del taller se actualizará de

manera leve pero significativa haciendo uso de una nueva plantilla y brindando nuevos ejemplos. La instrucción será en inglés y se utilizarán diapositivas en inglés y español.

Los proyectos de humanidades digitales que dan como resultado la creación de resultados digitales (normalmente un sitio web), son propensos a lo que Smithies et al. denominan “la entropía digital del software y la infraestructura digital” (2019). Los sitios estáticos se han convertido en una opción común para archivar proyectos heredados que corren el riesgo de quedar fuera de línea (Smithies et al., 2019; Summers, 2016) ya que solo requieren el mínimo absoluto de la infraestructura de alojamiento: un servidor para distribuir documentos en una dirección determinada. Una vez que los sitios han sido creados, no requieren mantenimiento activo y pueden moverse y transferirse fácilmente como cualquier otra colección de archivos. Por ejemplo, el proyecto Endings de la Universidad de Victoria, Columbia Británica (<https://endings.uvic.ca/>), recomienda los sitios estáticos como estrategia viable para garantizar la longevidad de las publicaciones del proyecto de Humanidades Digitales. Los Principios Endings para la longevidad digital incluyen, entre otras estrategias, la reducción tanto de la complejidad del software como de la dependencia de la infraestructura. Por otro lado, los sitios estáticos no pueden admitir funciones que requerirían un servidor activo, tales como la búsqueda de texto a gran escala y la gestión de usuarios. En consecuencia, se eliminan dichas funciones cuando los proyectos se archivan en sitios estáticos. Cuando el acceso a la infraestructura se vuelve limitado, derivar sitios estáticos de un proyecto al final de su vida útil es la opción clara. No obstante, este taller aborda la siguiente pregunta: ¿Qué se necesita para adoptar sitios estáticos desde el principio?

Después de realizar una introducción sobre los sitios estáticos y las razones para usarlos, el taller abordará los siguientes temas:

- Exploración de enfoques isomórficos: CETELcean y Verovio en Vanilla JavaScript
- Extensión CETELcean con Componentes Web.
- Extensión CETELcean con React.
- Uso del generador de sitios estáticos Gatsby y de los complementos TEI desarrollados para la revista Scholarly Editing (gatsby-transformer-ceteicean y gatsby-theme-ceteicean).
- Teniendo en cuenta los múltiples niveles de especialización, es probable que nos dividamos en varios grupos para realizar un trabajo colaborativo dirigido por los asistentes.
- Trabajo colaborativo opcional: experimentar con otros marcos de trabajo con los que los asistentes puedan estar familiarizados (por ejemplo, Vue, Angular, etc.).

- Trabajo colaborativo opcional: mantenerse dentro de una sola pila: aplicar transformaciones opcionales a documentos XML en NodeJS (usando JavaScript/TypeScript o XSLT a través de SaxonJS).

TEI Publisher: ediciones sofisticadas sin necesidad de programar

Helena Bermúdez Sabel¹, Magdalena Turska²

¹Jinntec (Germany), ²e-editiones (Switzerland)

TEI Publisher fue inicialmente concebido como una herramienta con la que cubrir la brecha entre los archivos TEI con la codificación de las fuentes y su publicación como edición académica digital. TEI Publisher es un software libre y de código abierto cuyo desarrollo es respaldado y coordinado por la sociedad internacional sin ánimo de lucro e-editiones.

La piedra angular de TEI Publisher es la implementación del Modelo de Procesamiento de TEI (TEI Processing model), el cual define de manera descriptiva cómo los diferentes elementos de un documento TEI deben ser presentados para su publicación. Sin embargo, el desarrollo de una edición digital va más allá de la mera transformación de elementos, por lo que TEI Publisher permite generar fácilmente la interfaz de la edición (navegación, paginación, presentación de diferentes versiones incluida la edición facsimilar, y un largo etcétera). En su versión más reciente, la 9.0.0, TEI Publisher ofrece numerosas funcionalidades que facilitan cada una de las etapas de la elaboración de una edición digital: aparte de incluir nuevos componentes que permiten una presentación del texto crítico altamente sofisticada (sin necesidad de programar), TEI Publisher asiste al equipo editorial en la creación de las fuentes TEI, en la anotación editorial, analítica y/o semántica, así como en la exploración de la edición a través de un motor de búsqueda facetada y de texto integral, y de registros de entidades.

El objetivo de este taller es mostrar cómo podemos utilizar TEI Publisher para generar ediciones digitales que respondan a las necesidades informativas del equipo editorial. Enseñaremos cómo construir una interfaz adaptada a cada tipo de edición, con el fin de que esta transmita el discurso histórico-filológico deseado. Haremos hincapié en las nuevas funcionalidades que facilitan la anotación editorial, analítica y semántica (esta última adaptada al español). Los puntos más relevantes del programa son:

- Introducción al Modelo de Procesamiento de TEI: cómo definir la transformación deseada para cada elemento desde el ODD.

- Adaptación de los diferentes componentes que permiten una anotación automatizada para la creación de registros.
- Detección y anotación automática de entidades nombradas en español.
- Adaptación de la búsqueda facetada y de texto integral.

TEI Publisher: sophisticated editions without the need to program

Helena Bermúdez Sabel¹, Magdalena Turska²

¹Jinntec (Germany), ²e-editions (Switzerland)

TEI Publisher was originally conceived as a tool to bridge the gap between TEI files with source encoding and their publication as a scholarly digital edition. TEI Publisher is free-open source software whose development is supported and coordinated by the international non-profit association e-editiones.org.

Implementing the TEI Processing model is the cornerstone of TEI Publishers as it descriptively defines how the different elements of a TEI document should be presented for publication. However, since developing a digital edition is much more than merely transforming elements, TEI Publisher makes it possible to easily generate the edition interface (navigation, pagination, presentation of different versions including the facsimile edition, and a long etcetera). TEI Publisher 9.0.0, the current major version, offers numerous features that facilitate each of the stages involved in the preparation of a digital edition: apart from including new elements to make a refined presentation of the critical text (no programming required), TEI Publisher assists the editorial team in the creation of TEI sources, in editorial, analytical and/or semantic annotation, as well as in the exploration of the edition through a faceted and full-text search engine, and entity records.

The aim of this workshop is to show how TEI Publisher can be used to create digital editions that respond to the information needs of the editorial team. We will teach how to build an interface adapted to each type of edition so that it transmits the desired historical-philological discourse. We will emphasize the new features that facilitate editorial, analytical and semantic annotation (the latter has been adapted to Spanish).

Key points discussed in this workshop:

- Introduction to the TEI Processing Model: how to define the desired transformation for each element from the ODD format.
- Adaptation of the different components that allow automated annotation for the creation of records.
- Automatic detection and annotation of named entities in Spanish.

- Adaptation of faceted and full text search.

Navigating and Processing Data from the TEI with XSLT

Elisa Beshero-Bondar¹, Martina Scholger², Gustavo Fernández Riva³

¹Penn State Erie (United States of America), ²University of Graz (Austria), ³Heidelberg University (Germany)

Knowing how to locate and explore data in your encoding can help to learn how to work with TEI and XML generally. This workshop is designed for people who have some experience with TEI and seek to learn how to work with XML markup for analysis and research. Participants will gain a working, practical knowledge of the query language XPath and the transformation language XSLT, and learn how these can help to reduce reliance on software, packages and plugins that may become obsolete without warning. Further, XSLT's functional programming can serve as a way of articulating research questions around a document data model expressed in XML.

The emphasis of our workshop is "pull-processing": that is, extracting data and metadata from markup documents for analysis, as opposed to providing the reading view of a digital scholarly edition. Markup in documents supplies structures and contexts that are especially useful for processing data, beyond what we can do with so-called "plain text". We will demonstrate some basic XPath navigation and calculation functions, and then show how XPath is applied in XSLT templates to address specific nodes that hold data of interest for visualization.

We will process TEI documents composed in Spanish and in languages represented by our workshop members' projects, to show that the code we write is transferable to multiple projects across language and cultural borders. Workshop instructors will collaborate and seek advice from the conference organizers on preparing Spanish-language source materials and documentation to establish an international foundation for this workshop.

Participants will learn how to "pull" data from TEI and output text formats required for simple online tools, where the structure of the output data is transferable to many different online calculation programs and amenable to statistical processing. During the workshop we will produce some simple structured documents for storing, sharing, and visualizing data: HTML lists and tables as well as plain text tabulated data (CSV or TSV files), and (if we have time) simple SVG bar or line graphs.

We hope to process some participant-supplied XML before, during, and after the workshop. We will carefully document the XSLT that we supply during the workshop to assist participants with revising and adapting the code to their own projects.

Navegar y procesar datos de la TEI con XSLT

Elisa Beshero-Bondar¹, Martina Scholger², Gustavo Fernández Riva³

¹Penn State Erie, (Estados Unidos de América), ²Universidad de Graz (Austria),

³Universidad de Heidelberg (Alemania)

Saber cómo localizar y explorar datos en su labor de codificación puede ayudarle a aprender a trabajar con TEI y XML en general. Este taller está diseñado para personas que tienen alguna experiencia con TEI y buscan aprender a trabajar con marcado XML para análisis e investigación. Los participantes obtendrán conocimientos prácticos del lenguaje de consulta XPath y del lenguaje de transformación XSLT, y aprenderán de qué manera estos pueden ayudar a reducir la dependencia de software, paquetes y complementos (plugins) que tal vez queden obsoletos sin previo aviso. Asimismo, la programación funcional de XSLT puede servir para articular preguntas de investigación en torno a un modelo de datos de documento expresado en XML.

El énfasis de nuestro taller está puesto en el "procesamiento pull": es decir, extraer datos y metadatos de documentos marcados para su análisis, en lugar de proporcionar la vista de lectura de una edición digital. El marcado de documentos proporciona estructuras y contextos que son especialmente útiles para procesar datos, más allá de lo que podemos hacer con el llamado "texto plano". Expondremos algunas funciones básicas de cálculo y navegación XPath, y luego mostraremos cómo se aplica XPath en plantillas XSLT para abordar nodos específicos que contienen datos de interés para visualización.

Con el objetivo de demostrar que el código que escribimos es transferible a múltiples proyectos a través de fronteras lingüísticas y culturales, procesaremos documentos TEI compuestos en español y en idiomas representados por los proyectos de los miembros de nuestro taller. Los instructores del taller colaborarán y buscarán asesoramiento por parte de los organizadores de la conferencia sobre la preparación de material y documentación en español para establecer una base internacional para este taller.

Los participantes aprenderán cómo "extraer" datos de TEI y generar formatos de texto necesarios para herramientas en línea simples, donde la estructura de los datos de salida es transferible a muchos programas de cálculo en línea diferentes y susceptible al procesamiento estadístico. Durante el taller elaboraremos algunos documentos simples estructurados para almacenar, compartir y visualizar datos: listas y tablas

HTML, así como datos tabulados de texto sin formato (archivos CSV o TSV) y, si disponemos de tiempo, gráficos lineales o barras SVG simples.

Esperamos procesar algún XML proporcionado por los participantes antes, durante y después del taller. Documentaremos cuidadosamente el XSLT que brindamos durante el taller para ayudar a los participantes en la revisión y adaptación del código a sus propios proyectos.

Long Papers / Presentaciones extensas

“I don’t even see the code”: The Dynamic Table of Contexts – Optimizing our Encoded Texts for Web-based Reading and Analysis Environments

Diane Jakacki¹, Susan Brown²

¹Bucknell University (United States of America), ²University of Guelph (Canada)

Editors of scholarly digital objects pour untold hours into our encoding. We structure, annotate, and interpret our texts with intricate and meaningful TEI elements; we constrain our schemas; we curate complex taxonomies; and we meticulously reference other authorities and scholarly sources from within our code. Many of us even use the encoding process to undertake our editorial analysis. Yet when we transform our TEI files into publishable formats so that others can read our work, much of that painstaking analysis seems lost.

Certainly we can present structure and surface annotations –even gesture to external authorities through web-optimized content management systems and publishing tools. But the ability to really leverage the analysis within our web-published texts can still elude us, and it is difficult for readers to follow our intrinsically-encoded analysis unless they download our TEI files and know how to read XML-TEI code.

One way in which we can engage readers in our editorial process is by producing an index. Unlike a structurally-oriented table of contents, an index provides a conceptual navigation mechanism that guides the reader in ways that the author or editor wants them to read the text. In doing so, we can reveal the analysis – and the encoding – underneath.

The Dynamic Table of Contexts (DToC) publishing environment was devised to do just this –to bring together the conceptual index with the web-presented encoded text. Developed as part of the Implementing New Knowledge Environments (INKE) project, DToC’s design team was led by Susan Brown, Stan Ruecker, and Stéfan Sinclair. The

tool has been housed within the Canadian Writing Research Collaboratory (CWRC) platform, with Voyant's visual analysis sensibilities peering through. Until now, DToC has been accessible mainly through CWRC. However, DToC has now been upgraded and made more widely available to editing audiences, both within the Linked Editing Academic Framework (LEAF) virtual research environment and the modular LEAF Commons tool suite. This means that anyone with a GitHub account can produce their own DToC editions, leveraging GitHub's versioning and collaboration affordances. In this way DToC will see new life and offer new opportunities for reader engagement by accommodating complex scholarly editing projects that weave together the analytical, indexical work, and encoded text. It should not come as a surprise, considering its interleaving with Voyant, that DToC also offers editors the opportunity to integrate iterative text analysis into their encoding process – 'seeing' patterns within texts drawn out by both tagging and full text search (Fig. 1), and then continuing to enhance their encoding on the fly, strengthening arguments through encoding; furthermore, they can then invite their readers to interact with the texts by curating their own tagsets and pursue (and share) their own analysis.

“Ni siquiera veo el código”. La tabla dinámica de contextos: optimizar nuestros textos codificados para entornos de lectura y análisis basados en la web

Diane Jakacki¹, Susan Brown²

¹Universidad Bucknell (Estados Unidos de América), ²Universidad de Guelph (Canadá)

Los editores de objetos digitales académicos dedicamos incontables horas a la codificación. Estructuramos, anotamos e interpretamos nuestros textos con elementos TEI complejos y significativos; restringimos nuestros esquemas; seleccionamos taxonomías complejas; y hacemos referencia de manera meticulosa a otras autoridades y fuentes académicas dentro de nuestro código. Muchos de nosotros incluso utilizamos el proceso de codificación para realizar nuestro análisis editorial. Sin embargo, cuando transformamos nuestros archivos TEI en formatos publicables para que otros puedan leer nuestro trabajo, gran parte de ese minucioso análisis parece perderse.

Ciertamente, podemos presentar anotaciones estructurales y superficiales, e incluso apuntar a autoridades externas a través de sistemas de gestión de contenido y herramientas de publicación optimizados para la web. Pero la capacidad de aprovechar realmente el análisis dentro de nuestros textos publicados en la web aún puede eludirnos, y es difícil para los lectores seguir nuestro análisis intrínsecamente codificado a menos que descarguen nuestros archivos TEI y sepan leer el código XML-TEI.

Una manera de involucrar a los lectores en nuestro proceso editorial es elaborar un índice. A diferencia de una tabla de contenido orientada estructuralmente, un índice proporciona un mecanismo de navegación conceptual que guía al lector en la forma en que el autor o editor quiere que lea el texto. Al hacerlo, podemos revelar el análisis (y la codificación) que se encuentra debajo.

El entorno de publicación de la Tabla Dinámica de Contextos (DToC) fue ideado para hacer precisamente esto: unir el índice conceptual al texto codificado presentado en la web. Desarrollado como parte del proyecto Implementación de nuevos entornos de conocimiento (INKE), el equipo de diseño de DToC estuvo dirigido por Susan Brown, Stan Ruecker y Stéfan Sinclair. La herramienta se ha alojado dentro de la plataforma Canadian Writing Research Collaboratory (CWRC), junto con las capacidades de análisis visual de Voyant. Hasta ahora, se podía acceder a DToC principalmente a través de CWRC. Sin embargo, ahora DToC se ha actualizado y está más disponible para las audiencias de edición, tanto dentro del entorno de investigación virtual LEAF (marco académico de edición vinculada) como en el conjunto de herramientas modulares LEAF Commons. Esto significa que cualquier persona con una cuenta de GitHub puede producir sus propias ediciones DToC, aprovechando las funciones de colaboración y control de versiones de GitHub. De esta manera, DToC cobrará nueva vida y ofrecerá nuevas oportunidades para la participación del lector al incorporar proyectos complejos de edición filológica que entrelazan el trabajo analítico, indexado y el texto codificado. No debería sorprender, dada su intercalación con Voyant, que DToC también ofrezca a los editores la oportunidad de integrar el análisis de texto iterativo en su proceso de codificación, "ver" patrones dentro de los textos extraídos tanto mediante etiquetado como mediante búsqueda de texto completo, y luego continuar mejorando su codificación al instante, fortaleciendo los argumentos a través de la codificación; asimismo, pueden invitar a sus lectores a interactuar con los textos seleccionando sus propios conjuntos de etiquetas y realizar (y compartir) su propio análisis.

El lenguaje de los libros: analizando los inventarios de libros en Navarra en los siglos XVI y XVII

Alexandra Elizabeth Wingate

Universidad de Indiana Bloomington (Estados Unidos de América)

Para los historiadores del libro, los inventarios que documentaban el contenido de las bibliotecas privadas y las tiendas de librerías e impresores son entre las fuentes más importantes para entender cuáles libros circulaban en un periodo y lugar determinado.

Estos inventarios tienen entradas que enumeran los libros de una colección, y el proceso normal para los historiadores del libro es tratar de descifrar las descripciones de los libros hechas por los escribanos, profesionales de libros, u otros individuos para identificar un libro en específico. A pesar de ser fuentes muy cruciales, todavía no entendemos muy bien cómo se hacían los inventarios y especialmente si existían normas de descripción debido a que la información dada en una entrada de un inventario puede variar a sobremanera. Hay evidencia que había por lo menos algunas normas ya que unas obras y unos autores populares se referían por nombres estándares. Por ejemplo, en un pleito entre profesionales de libro en Navarra en España en 1665, la viuda de un librero declaró que las obras de Pedro de Barbosa “ordinariamente se rotulan Petri Barbosa opera” (AGN, n. 189680, fol. 580r), insinuando que había una forma regular de referirse a este autor y sus obras. No obstante, la aplicación de TEI a estos inventarios nos ofrece un método para entender mejor el lenguaje bibliográfico formado por estas normas.

En esta ponencia, voy a demostrar la utilidad de TEI para el estudio de los inventarios de libros a través de un corpus de 15 inventarios de bibliotecas privadas de clérigos y 3 inventarios de libreros de Navarra en los siglos XVI y XVII transcritos en TEI. Primero, argumento que el uso de TEI permite que los académicos compartan sus transcripciones más fácilmente no solo como documentos a través de ediciones digitales sino también como datos de los inventarios, especialmente en comparación con los métodos tradicionales para diseminar los inventarios transcritos. Segundo, afirmo que el uso de TEI nos permite empezar a desvelar el lenguaje bibliográfico ya mencionado porque con el marcado descriptivo de TEI es posible especificar los diferentes tipos de información en una entrada y cómo los individuos se referían a diferentes obras y autores. Mi análisis preliminar de estas entradas como datos estructurados revela que existía un lenguaje bibliográfico más estandarizado entre los profesionales del libro en cuanto a la estructura de las entradas y las referencias a ciertas obras y autores. Sin embargo, los inventarios hechos por escribanos u otros individuos no manifiestan la misma consistencia en su estructura, pese a que sí usan más frecuentemente las denominaciones normativas para obras y autores. Fundamentalmente, esta presentación evidencia el papel crucial de TEI para apoyar una comprensión mayor de un tipo de fuente común en la historia del libro.

The language of books: analyzing book inventories in Navarra in the 16th and 17th centuries

Alexandra Elizabeth Wingate

Indiana University Bloomington (United States of America)

Book historians consider inventories documenting the contents of private libraries, bookshops and printing houses to be among the most important sources in order to see the books that circulated in a given period and place. The entries in these inventories list the books in a collection, and trying to decipher descriptions of books made by scribes, book professionals, or other individuals so as to identify a specific book is a common process for book historians. Despite being extremely important sources, we still do not fully understand how inventories were made and, especially, if there were descriptive standards because the information given in an inventory entry can vary greatly. There is evidence that there were at least some norms since popular works and authors were referred to by standard names. For instance, in 1665, the widow of a bookseller declared in a lawsuit between book professionals in Navarra, Spain that the works of Pedro de Barbosa “are frequently labeled Petri Barbosae opera” (AGN, n. 189680, fol. 580r), suggesting that there was a regular way of referring to this author and his works. However, using TEI in these inventories provides a method which enables us to better understand the bibliographic language shaped by these standards.

In this paper, I am going to show how useful TEI is when studying book inventories by means of a corpus of 15 inventories of private clerical libraries and 3 inventories of Navarrese booksellers in the 16th and 17th centuries transcribed in TEI. First, I would argue that using TEI enables scholars to share their transcripts easily not only as documents through digital editions, but also as inventory data, especially compared to traditional methods of disseminating transcribed inventories. Second, I strongly believe using TEI helps us unravel the bibliographic language already mentioned because it is possible to specify the different types of information in an entry and how individuals referred to different works and authors through the use of TEI descriptive markup language. My preliminary analysis of these entries as structured data reveals that a more standardized bibliographic language existed among book professionals in terms of the structure of the entries and references to certain works and authors. However, inventories made by scribes or other individuals do not show the same consistency in their structure despite using normative designations for works and authors more frequently. Fundamentally, this presentation evidences the crucial role TEI plays in supporting greater understanding of a common source type in book history.

Encoding Edgeworth: TEI Development for Correspondence and Entity Indices

Meredith L. Hale

University of Tennessee (United States of America)

The Maria Edgeworth Letters Project (MELP) aims to make an open-access digital archive containing the complete correspondence of the Anglo-Irish Regency author Maria Edgeworth (1768-1849) and her circle. TEI is central to this project as it is used to encode the text of the letters, create indices of named persons, places, and works, and ensure interoperability of the project long term. Letters in both English and French are present in the corpus. To date, over 200 letters have been encoded and made available on our beta site –<https://melp.dh.tamu.edu/>. In the next three years, this inter-institutional collaborative project plans to encode an additional 500 letters and provide minimal metadata for all remaining digitized letters using elements with the `teiHeader`.

This presentation will cover the workflows currently being used to achieve this work as well as challenges the team is currently working through. For current workflows, the majority of encoding is completed by graduate assistants who are aiming to advance their skills in the digital humanities. These students use a TEI template to guide their letter encoding while spreadsheet data entry is the foundation for the indices. Adding entities to the indices involves in depth authority research using VIAF, GeoNames, Wikidata, and the Library of Congress Name Authority File (LCNAF). Separate TEI files for persons, places, and works are subsequently generated from the spreadsheets using GREL in OpenRefine. GitHub is key to managing the TEI contributions of all project contributors for version control purposes. As many in the humanities have not used GitHub previously, training is essential.

Possible solutions for future work will also be featured. As the team continues to add entities to our indices, opportunities to further capitalize upon the content negotiable authorities we have added needs to be explored. At present on our beta site the links are simply displayed when an established entity is present in a letter. In order to ensure the validity of our TEI, a goal is to set up automated testing in GitHub. Special considerations when encoding the French correspondence also need to be established and incorporated into existing documentation. Solutions for the idiosyncrasies we uncover within the letters while encoding, such as letters within letters, notes added by librarians, and the inclusion of unique diagrams or figures, also need to be continually developed.

Codificando a Edgeworth: desarrollo de marcado TEI para correspondencia e índices de entidades

Meredith L. Hale

Universidad de Tennessee (Knoxville, Estados Unidos de América)

El proyecto Maria Edgeworth Letters Project (MELP) tiene como objetivo crear un archivo digital de acceso abierto que contenga la correspondencia completa de la autora de la Regencia angloirlandesa Maria Edgeworth (1768-1849) y su círculo. La TEI es fundamental para este proyecto, ya que se utiliza para codificar el texto de las cartas, crear índices de personas, lugares y obras nombrados, y garantizar la interoperabilidad del proyecto a largo plazo. El corpus contiene tanto las cartas en inglés como en francés. Hasta la fecha, se han codificado y puesto a disposición más de 200 cartas en nuestro sitio beta: <https://melp.dh.tamu.edu/>. En los próximos tres años, este proyecto de colaboración interinstitucional planea codificar 500 cartas adicionales y proporcionar metadatos mínimos para todas las cartas digitalizadas restantes utilizando elementos con el encabezado TEI.

Esta presentación cubrirá los flujos de trabajo que se utilizan hoy en día para lograr esta labor, así como los desafíos en los que el equipo está trabajando en la actualidad. La mayor parte de la codificación para los flujos de trabajo actuales la realizan asistentes graduados que buscan mejorar sus habilidades en humanidades digitales. Estos estudiantes utilizan una plantilla TEI para guiar el proceso de codificación de cartas, mientras que la entrada de datos en hojas de cálculo es la base para los índices. Agregar entidades a los índices implica una investigación de autoridad en profundidad utilizando VIAF, GeoNames, Wikidata y el Archivo de autoridad de nombres de la Biblioteca del Congreso (LCNAF). Posteriormente, se generan archivos TEI separados para personas, lugares y obras a partir de las hojas de cálculo utilizando GREL en OpenRefine. GitHub es clave para gestionar las contribuciones TEI de todos los contribuyentes al proyecto con fines de control de versiones. La capacitación es esencial dado que muchos profesionales de las humanidades no han utilizado GitHub anteriormente.

Asimismo, se presentarán posibles soluciones para trabajos futuros. A medida que el equipo continúa incorporando entidades a nuestros índices, se vuelve necesario analizar oportunidades para capitalizar aún más las autoridades de contenido negociable que hemos agregado. En la actualidad, se muestran simplemente los enlaces en nuestro sitio beta cuando una entidad establecida está presente en una carta. Configurar pruebas automatizadas en GitHub tiene como finalidad garantizar la validez de nuestra TEI. También es necesario establecer e incorporar consideraciones especiales a la documentación existente al momento de codificar la correspondencia francesa. Asimismo, es importante continuar desarrollando soluciones para las idiosincrasias que descubrimos dentro de las cartas mientras codificamos, tales como cartas dentro de cartas, notas agregadas por los bibliotecarios, y la inclusión de diagramas o figuras únicos.

Empowering Text Encoding with Large Language Models: Benefits and Challenges

Martina Scholger, C. Pollin

University of Graz (Austria)

This contribution will discuss how Large Language Models (LLMs) can be used to support and enhance text encoding with the standard of the Text Encoding Initiative, demonstrating an exemplary workflow –from model creation to data extraction to data analysis to presentation– in the context of a letter edition.

Potenciar la codificación de textos con grandes modelos de lenguaje: beneficios y desafíos

Martina Scholger, C. Pollin

Universidad de Graz (Austria)

Este artículo analizará cómo los grandes modelos de lenguaje (LLM) pueden utilizarse para respaldar y mejorar la codificación de textos con el estándar TEI Text, demostrando un flujo de trabajo ejemplar, desde la creación del modelo hasta la extracción de datos, el análisis de datos y la presentación, en el contexto de la edición de una carta.

Enhancing Technical Knowledge Acquisition with RAG Systems: the TEI use case

Mohamed Khemakhem¹, Houcem Eddine Rekik², Oussama Bouaziz¹

¹MandaNetwork, ²ENSTA Paris (France)

In an era dominated by an explosion of technical documentation across diverse domains, the need for effective knowledge acquisition mechanisms has become paramount. The assimilation of the Text Encoding Initiative (TEI) guidelines, for instance, presents challenges for organizations and individuals seeking to effectively adopt its encoding principles. Retrieval-Augmented Generation (RAG) systems emerge as a promising paradigm to address this challenge, seamlessly integrating information retrieval with natural language generation to facilitate the acquisition of technical knowledge from large documentation material. In this publication, we explore how RAG

systems can mitigate these challenges while maximizing the benefits of TEI adoption, particularly in the context of learning and implementation.

Challenges in learning and adopting TEI guidelines revolve around the complexity of the markup language and the diverse skill levels of users. Mastering TEI requires familiarity with its intricate syntax, encoding conventions, and domain-specific applications, posing a steep learning curve for novices. Furthermore, the extensive volume of its published guidelines poses another challenge, even for experienced users, in efficiently retrieving relevant information.

RAG systems provide a novel approach to technical knowledge acquisition by seamlessly integrating the power of Large Language Models (LLMs) and specialized knowledge. The logic behind a RAG system lies in its ability to leverage pre-trained LLMs to generate informative and contextually relevant responses based on retrieved information. This approach has emerged to tackle the hallucination issue observed in generative models. It achieves this by enriching the context necessary for these models with knowledge sourced directly from relevant documents.

Consequently, RAG systems offer a promising solution to TEI adoption and learning challenges by providing a more adaptive and interactive content. Through advanced natural language generation capabilities, RAG systems can generate tailored explanations, examples, and walkthroughs of TEI encoding practices, catering to the specific needs and skill levels of users. By leveraging retrieval mechanisms, RAG systems can retrieve relevant TEI guidelines and examples from its extended documentation, facilitating self-paced learning and knowledge acquisition. Such interactive systems can empower users by providing assistance in the creation of TEI-compliant markup, streamlining the encoding process, and thereby reducing errors and inconsistencies. Furthermore, RAG-generated summaries and explanations elucidate the rationale behind TEI encoding decisions, enhancing transparency and reproducibility in digital humanities research.

Upon experimentation with state of the art models, we observed the persistence of some technical challenges for this ultimate goal. First, a pre-processing of the documentation is necessary to overcome issues related to the tokenization process. Moreover, a chunking strategy for such a rich documentation has to be carefully defined to enable more precise information retrieval and complete response from the AI assistant. In addition, choosing the right prompt remains crucial to frame the context and the expected outcome in order to generate accurate responses.

In conclusion, RAG systems offer a transformative approach to learning and adopting TEI guidelines, mitigating challenges and maximizing benefits for knowledge acquisition. By leveraging RAG systems to facilitate TEI learning, organizations can empower users to unlock the full potential of TEI for data interoperability, scholarly communication and digital humanities research.

Mejora de la adquisición de conocimientos técnicos con RAG Systems: el caso de uso de TEI

Mohamed Khemakhem¹, Houcem Eddine Rekik², Oussama Bouaziz¹

¹MandaNetwork, ²ENSTA París (Francia)

En una era dominada por una explosión de documentación técnica en diversos ámbitos, se ha vuelto primordial la necesidad de contar con mecanismos eficaces para la adquisición de conocimientos. La asimilación de las directrices TEI, por ejemplo, presenta desafíos para las organizaciones e individuos que buscan adoptar, de manera eficaz, sus principios de codificación. Los sistemas RAG (Generación Aumentada de Recuperación) emergen como un paradigma prometedor para abordar este desafío al integrar perfectamente la recuperación de información con la generación de lenguaje natural a fin de facilitar la adquisición de conocimientos técnicos a partir de material de documentación de gran tamaño. En esta publicación, exploramos cómo los sistemas RAG pueden mitigar estos desafíos y, al mismo tiempo, maximizar los beneficios de la adopción de TEI, particularmente en el contexto del aprendizaje y la implementación.

Los desafíos para aprender y adoptar las directrices TEI giran en torno a la complejidad del lenguaje de marcado y los diversos niveles de habilidad de los usuarios. Dominar TEI requiere familiaridad con su intrincada sintaxis, sus convenciones de codificación y sus aplicaciones de dominio específico, lo que plantea un arduo proceso de aprendizaje para los principiantes. Asimismo, el gran volumen de sus directrices publicadas plantea otro desafío, incluso para los usuarios experimentados, a la hora de recuperar información relevante de forma eficiente.

Los sistemas RAG proporcionan un enfoque novedoso para la adquisición de conocimientos técnicos al integrar perfectamente el poder de los modelos de lenguaje grande (LLMs) y el conocimiento especializado. La lógica detrás de un sistema RAG radica en su capacidad de aprovechar los LLM previamente capacitados para generar respuestas informativas y contextualmente relevantes basadas en la información recuperada. Este enfoque surgió para abordar el problema de las alucinaciones observado en los modelos generativos. Lo logra enriqueciendo el contexto necesario para estos modelos con conocimiento obtenido directamente de documentos relevantes.

En consecuencia, los sistemas RAG ofrecen una solución prometedora para los desafíos de aprendizaje y la adopción de la TEI al proporcionar un contenido más adaptable e interactivo. Por medio de capacidades avanzadas de generación de lenguaje natural, los sistemas RAG pueden generar explicaciones, ejemplos y tutoriales personalizados de las prácticas de codificación TEI, atendiendo a las

necesidades específicas y los niveles de habilidad de los usuarios. Al utilizar los mecanismos de recuperación, los sistemas RAG pueden recuperar directrices y ejemplos relevantes de la TEI desde su documentación ampliada, lo que facilita el aprendizaje y la adquisición de conocimientos de manera personalizada. Estos sistemas interactivos pueden empoderar a los usuarios al brindarles asistencia en la creación de marcado compatible con TEI, simplificando el proceso de codificación y, por lo tanto, reduciendo errores e inconsistencias. Por otra parte, los resúmenes y explicaciones generados por RAG aclaran el fundamento detrás de las decisiones de codificación TEI, lo que resulta en una mejora en cuanto a la transparencia y la reproducibilidad en la investigación de humanidades digitales.

Al experimentar con modelos de última generación, observamos la persistencia de algunos desafíos técnicos para alcanzar este objetivo final. En primer lugar, es necesario un preprocesamiento de la documentación para superar los problemas relacionados con el proceso de tokenización. Por otro lado, se debe definir cuidadosamente una estrategia de fragmentación para una documentación tan rica a fin de permitir una recuperación de información más precisa y una respuesta completa del asistente de IA. Más aún, elegir el mensaje correcto sigue siendo crucial para enmarcar el contexto y el resultado esperado con el fin de generar respuestas precisas.

En conclusión, los sistemas RAG ofrecen un enfoque transformador para aprender y adoptar directrices TEI, mitigando desafíos y maximizando beneficios para la adquisición de conocimientos. Al hacer uso de los sistemas RAG para facilitar el aprendizaje de TEI, las organizaciones pueden capacitar a los usuarios con el objetivo de desbloquear todo el potencial de TEI para la interoperabilidad de datos, la comunicación académica y la investigación en humanidades digitales.

From Catullus to Wikidata: Language Models, Metadata Schemas, and Ontologies in a Digital Edition in TEI

Carlos Nusch^{1,2,3}, Gabriel Calarco², Gimena del Rio Riande^{2,3}, Leticia Cecilia Cagnina^{2,4}, Marcelo Luis Errecalde^{2,4}, Leandro Antonelli¹

¹Universidad Nacional de La Plata, ²Consejo Nacional de Investigaciones Científicas y Técnicas, ³Asociación Argentina de Humanidades Digitales, ⁴Universidad Nacional de San Luis (Argentine Republic)

This paper details various tasks of markup and natural language processing conducted in Aetatis Amoris, a project dedicated to exploring love poetry throughout literary history. The project focuses on the enriched digital edition of works by classical Latin poets such as Gaius Valerius Catullus, Albius Tibullus, and Sextus Propertius, using the

TEI-XML standard for encoding the texts. Initial tasks included the automatic generation of the main document elements, such as header and body, and the counting and automatic tagging of verses and stanzas.

Subsequently, using the advanced spaCy model for Latin, LatinCy, names of people and places were automatically extracted and tagged with the corresponding labels. In addition, searches for standardized metadata were carried out using external APIs, and consulting databases such as Virtual International Authority File (VIAF), Pleiades project, and Wikidata. This allowed for the retrieval of standardized identifiers, rich and curated information, and images of historical places and characters. As a conclusion, we can state that while the automatic tools significantly facilitated the digital editing process, the vast amount of information recovered also posed significant challenges in data curation and quality assessment, redefining the digital scholarly editor role in the process.

De Catulo a Wikidata: modelos de lenguaje, esquemas de metadatos y ontologías en una edición digital en TEI

Carlos Nusch^{1,2,3}, Gabriel Calarco², Gimena del Rio Riande^{2,3}, Leticia Cecilia Cagnina^{2,4}, Marcelo Luis Errecalde^{2,4}, Leandro Antonelli¹

¹Universidad Nacional de La Plata, ²Consejo Nacional de Investigaciones Científicas y Técnicas, ³Asociación Argentina de Humanidades Digitales, ⁴Universidad Nacional de San Luis (República Argentina)

Este artículo detalla varias tareas de marcado y procesamiento del lenguaje natural realizadas en Aetatis Amoris, un proyecto dedicado a explorar la poesía amorosa a lo largo de la historia literaria. El proyecto se centra en la edición digital enriquecida de obras de poetas latinos clásicos como Cayo Valerio Catulo, Albio Tibulo y Sexto Propercio, utilizando el estándar TEI-XML para codificar los textos. Las tareas iniciales incluyeron la generación automática de los elementos principales del documento tales como el encabezado y el cuerpo, y el conteo y etiquetado automático de versos y estrofas.

Posteriormente, utilizando el modelo avanzado de spaCy para latín, LatinCy, se extrajeron, de manera automática, nombres de personas y lugares, y se etiquetaron con las etiquetas correspondientes. Asimismo, se realizaron búsquedas de metadatos estandarizados utilizando APIs externas y consultando bases de datos como Virtual International Authority File (VIAF), proyecto Pleiades y Wikidata. Esto permitió la recuperación de identificadores estandarizados, de información rica y curada, y de imágenes de lugares y personajes históricos. Como conclusión, podemos afirmar que, si bien las herramientas automáticas facilitaron significativamente el proceso de edición

digital, la gran cantidad de información recuperada también planteó desafíos importantes acerca de la curación de datos y la evaluación de la calidad, redefiniendo el papel del editor académico digital en el proceso.

TEI and Right to Left Scripts: How We Got Here and What We Can Do About It

Hugh Cayless

Duke University (United States of America)

A variety of problems face those who wish to use TEI with texts using right-to-left (rtl) scripts, making the baseline “plain text” mode of editing unusable. This paper will survey the technical history of why we ended up in such a situation, discuss possible mitigations and their benefits and drawbacks, and make a call for investment in addressing this serious deficiency in our tooling.

TEI y la escritura de derecha a izquierda: cómo llegamos aquí y qué podemos hacer al respecto

Hugh Cayless

Universidad de Duke (Estados Unidos de América)

Quienes desean utilizar TEI con textos escritos de derecha a izquierda (rtl) enfrentan una variedad de problemas que hacen que el modo de edición básico de “texto sin formato” se vuelva inutilizable. Examinaremos la historia técnica respecto de los motivos por los que nos encontramos en dicha situación, discutiremos posibles mitigaciones y sus beneficios e inconvenientes, y haremos un llamado a la inversión para abordar esta grave deficiencia en nuestras herramientas.

Dubsar: A New Approach to Rendering Non-Western and Encoded Text

Kevin Graaf

Independent Researcher (United States of America)

Rendering text is hard. A typical internationalized text rendering system cumulatively represents hundreds of thousands of lines of code, spread between operating systems,

software libraries, and application programs. And yet, despite millions of person-hours and dollars worth of engineering, much of contemporary software still struggles to handle non-European, non-alphabetic text correctly.

Programmers have heretofore typically assumed that "text", simpliciter, is fundamentally a one-dimensional sequence of symbols that is divided into a series of horizontal lines. In fact, neither of these assumptions are universally true. Unfortunately, modern text rendering stacks are frequently still oriented towards the needs of, and assumptions peculiar to, European alphabetic scripts. Attempting to retrofit non-European scripts onto an essentially European-Alphabetic model has only exacerbated the inherent complexity of the problem.

The highly specialized domain knowledge required to handle non-European writing systems correctly is a serious impediment to internationalizing software. In the case of new or minority scripts, it might well be impossible at present. That such scripts are thus systematically excluded from software can contribute to their marginalization. The predominant font technology, OpenType (<https://learn.microsoft.com/en-us/typography/opentype/spec/>), addresses this problem by providing built-in "shaping models" that attempt to capture the logic of how the world's writing systems work. OpenType is highly effective at handling those classes of scripts within its purview, but it cannot easily account for any behavior beyond what its creators anticipated.

I therefore present Dubsar, a new computer typography system that provides high quality rendering of complex and minority writing systems and TEI-encoded documents. Dubsar collapses nearly all aspects of text rendering into a single mechanism: a simple programming language, DubsarScript, which allows users to programmatically describe how glyphs are drawn and positioned. Dubsar is not limited in what kinds of writing systems it can express precisely because Dubsar fonts just are arbitrary DubsarScript programs. Moreover, Dubsar embeds an expansive notion of text designed to accommodate writing systems such as Maya, which has so far resisted computerization due to its incredible complexity. Technologists who wish to support internationalized text in their software need only implement an interpreter for the DubsarScript language, which can be readily accomplished without any knowledge of how potentially unfamiliar writing systems work. Due to its relative simplicity, Dubsar furthermore empowers users of uncomputerized minority scripts to create fonts which would otherwise be impossible.

The difficulties of text rendering are only magnified in the context of TEI documents. Assumptions that were already tenuous within the plain text regime, such as unidimensionality, here cease to be true in any capacity. Dubsar is especially conducive to directly rendering TEI documents directly by virtue of the same properties that lend it to handling internationalized text.

This paper will first introduce the landscape of contemporary computer typography. It will then describe the Dubsar system: how it works, how it is implemented, and how it can be used to render TEI documents. Finally, the paper will give a brief overview of how Dubsar fonts can be authored.

Dubsar: un nuevo enfoque para renderizar texto codificado y no occidental

Kevin Graaf

Investigador independiente (Estados Unidos de América)

Renderizar texto es difícil. Un sistema típico de renderización de texto internacionalizado representa, en conjunto, cientos de miles de líneas de código repartidas entre sistemas operativos, bibliotecas de software y programas de aplicación. Y, sin embargo, a pesar de millones de horas-persona y dólares destinados a ingeniería, gran parte del software contemporáneo todavía tiene dificultades para manejar correctamente textos no europeos y no alfabéticos.

Hasta ahora, los programadores suponen que el "texto" es simplemente una secuencia unidimensional de símbolos que se divide en una serie de líneas horizontales. De hecho, ninguna de estas suposiciones es universalmente cierta. Desafortunadamente, las pilas de renderización de texto modernas, con frecuencia, aún se encuentran orientadas hacia las necesidades y suposiciones características de las escrituras alfabéticas europeas. Intentar adaptar escrituras no europeas a un modelo alfabético esencialmente europeo solo ha exacerbado la complejidad inherente del problema.

El hecho de que se requiera de un conocimiento del área altamente especializado para el correcto manejo de los sistemas de escritura no europeos, es un gran impedimento para la internacionalización del software. En la actualidad, esto podría resultar imposible en el caso de escrituras nuevas o minoritarias. Que estos documentos queden sistemáticamente excluidos del software puede contribuir a su marginación. La tecnología de fuentes predominante, OpenType (<https://learn.microsoft.com/en-us/typography/opentype/spec/>), aborda este problema proporcionando "modelos de configuración" integrados que intentan captar la lógica respecto del modo en que funcionan los sistemas de escritura. OpenType es muy eficaz en el manejo de esas clases de documentos dentro de su ámbito, pero no puede explicar fácilmente ningún comportamiento más allá de lo que anticiparon sus creadores.

Por consiguiente, presento Dubsar, un nuevo sistema de tipografía por computadora que proporciona una renderización de alta calidad de sistemas de escritura complejos y minoritarios, y documentos codificados con TEI. Dubsar contrae casi todos los aspectos de la representación gráfica del texto en un solo mecanismo: un lenguaje de

programación simple, DubsarScript, que permite a los usuarios describir, mediante programación, cómo se dibujan y posicionan los glifos. Dubsar no está limitado respecto de los tipos de sistemas de escritura que puede expresar precisamente porque las fuentes Dubsar son simplemente programas DubsarScript arbitrarios. Asimismo, Dubsar incorpora una noción amplia de texto diseñada para adaptarse a sistemas de escritura como el maya, que hasta ahora se ha resistido a la informatización debido a su increíble complejidad. Los tecnólogos que deseen admitir texto internacionalizado en su software sólo necesitan implementar un intérprete para el lenguaje DubsarScript, lo que se puede lograr fácilmente sin ningún conocimiento acerca de cómo funcionan los sistemas de escritura potencialmente desconocidos. Debido a su relativa simplicidad, Dubsar además permite que los usuarios de escrituras minoritarias no informatizadas creen fuentes que, de otro modo, serían imposibles.

Las dificultades de la renderización de texto sólo se magnifican en el contexto de los documentos TEI. Suposiciones que ya eran vagas dentro del régimen del texto plano tal como la unidimensionalidad, dejan de ser ciertas en cualquier capacidad. Dubsar es especialmente propicio para renderizar directamente documentos TEI en virtud de las mismas propiedades que lo facilitan para manejar texto internacionalizado.

En primer lugar, este artículo introducirá el panorama de la tipografía informática contemporánea. Luego, describirá el sistema Dubsar: cómo funciona, cómo se implementa y cómo se puede utilizar para representar documentos TEI. Por último, se brindará una breve descripción general sobre la forma de crear las fuentes Dubsar.

Panels / Paneles

TEI for Black DH: A Conversation between *Revue des Colonies* and Keywords for Black Louisiana

María Belieava Solomon¹, Leila K. Blackbird², Alex Gil³, Jessica Marie Johnson⁴, E. G. Palazzolo⁴, Grégory Pierrot⁵, Raffaele Vigiante¹

¹University of Maryland, ²University of Chicago; ³Yale University, ⁴Johns Hopkins University; ⁵University of Connecticut (United States of America)

This seven-person roundtable brings together the team from *The Revue des Colonies* with co-editors of the microedition of *Kinship and Longing: Keywords for Black Louisiana* to discuss experiences working with TEI and make the case for modifications

to better accommodate representation of Black life in slavery's archives and translation of language of race and empire in the colonial archive.

The *Revue des Colonies* project focuses on the eponymous journal, edited by Martinican abolitionist Cyrille Bisette. Published between 1834 and 1842, it was the first French periodical for and by people of color. Its monthly issues provided news about ongoing struggles for civil rights across the French colonial world and beyond, alongside original and reprinted fiction and poetry by global Black writers. Led by an international and interdisciplinary team of scholars including project director Maria Beliaeva Solomon, technical director Raffaele Viglianti and co-editor and translator Grégory Pierrot, the project to digitally annotate and translate this invaluable record of the global history of colonization, enslavement and abolition emerged in response to the absence of any complete and searchable collection, let alone translation, of the journal's complete print run. Supported by the Foundation for the Remembrance of Slavery, the Schomburg Center for Research in Black Culture (New York Public Library), the National Archives, and the Andrew W. Mellon Foundation, the project aims to restore the emancipatory rhetoric of the *Revue des Colonies* within the political, material, and cultural contexts of its publication and make it accessible to new generations.

At the heart of this endeavor lies the Text Encoding Initiative (TEI), which serves as "technology of recovery," as articulated by Kim Gallon in "Making a Case for the Black Digital Humanities" (*Debates in the Digital Humanities*, 2016). TEI not only facilitates the digitization process but also becomes an instrument to guard against the unwitting reproduction of power dynamics embedded within the original texts. Building upon the imperative articulated by Kelly Baker Josephs and Roopika Risam in their introduction to *The Digital Black Atlantic* to resist technology's historical perpetuation of dominant narratives of oppression, our edition's critical apparatus aims to highlight the original contributions of Black authors, editors, journalists and activists to the political and cultural transformations of the nineteenth century. Our TEI customization focuses on the encoding of named entities in order to provide contextual and critical annotation. The tagging features provide a perfect opportunity to create substantial, and cross navigable entries for individuals, events, and organizations that have been overlooked in scholarly discourse. To increase accessibility to the *Revue* sources, we also provide professional English translations.

On the Keywords side, beginning in Fall of 2020, with the leadership and collaboration of founding director Jessica Marie Johnson, guidance of Alex Gil, and editorial assistance of Raffaele Viglianti, Leila Blackbird, Olivia Barnard, Emma Katherine Bilski, and Ellie Palazzolo began working on a microedition for the journal *Scholarly Editing* (<https://scholarlyediting.org/>). The vision was to edit and publish a handful of documents from eighteenth-century colonial Louisiana –transcribed and translated via the Louisiana Historical Center Colonial Documents Digitization Project– to draw attention

to stories of African and African-descended people in that archive. Keywords for Black Louisiana is supported by the National Historical Publications and Records Commission.

The team has edited fifteen stories composed of twenty-one documents spanning from 1740 to 1795, bridging French and Spanish colonialism in Louisiana and reflecting carefully on the possibilities and limitations of TEI. Some of the original documents are in French and others are in Spanish. All have been transcribed and translated into English. The transcriptions and translations are marked up in TEI. The primary tags used are <persName> to identify named and unnamed individuals and <seg> with @ana for Keywords. One cornerstone presentation on Black DH and TEI, Caitlin Pollock and Jessica Lu's 2019 talk "Hacking TEI for Black Digital Humanities," influenced the Keywords project's approach to textual encoding early on in the editorial process. Pollock and Lu invited editors to work with TEI on Black history and archives to push some of the boundaries and conventions that the dictionary so meticulously documents. We encountered a number of questions and roadblocks that validated Pollock and Lu's call, and did a bit of hacking ourselves that we propose to present at the annual meeting while speculating on and encouraging further formal interventions in TEI.

By bringing together these two projects, we aim to foster conversation about how TEI can become more accountable to critical, postcolonial, and Black DH. We will build on the similarities and shared challenges and commitments as well as the differences between these two archives towards a far-reaching set of interventions. We propose additions to the TEI dictionary related to representation and translation of racialized language, to print as well as manuscript documents, and to documents written by and for as well as documents written about Black actors.

TEI para las Humanidades Digitales negras: una conversación entre Revue des Colonies y Keywords for Black Louisiana

M. Belieava Solomon¹, Lila K. Blackbird², Alex Gil³, Jessica Marie Johnson⁴, E. G. Palazzolo⁴, Grégory Pierrot⁵, Raffaele Viglianti¹

¹Universidad de Maryland, ²Universidad de Chicago, ³Universidad de Yale, ⁴Universidad Johns Hopkins, ⁵Universidad de Connecticut (Estados Unidos de América)

Esta mesa redonda de siete personas reúne al equipo de The Revue des Colonies con los coeditores de la microedición de Kinship and Longing: Keywords for Black Louisiana (Parentesco y anhelo: palabras clave para la Luisiana negra) para compartir experiencias al trabajar con TEI y promover modificaciones para que se adapte mejor a

la representación de la vida negra en los archivos de la esclavitud y la traducción del lenguaje de la raza y el imperio en el archivo colonial.

El proyecto *Revue des Colonies* se centra en la revista del mismo nombre, editada por el abolicionista martiniqués Cyrille Bissette. Publicada de 1834 a 1842, fue la primera revista francesa escrita por y para personas de color. Sus números mensuales brindaban noticias sobre las luchas en curso por los derechos civiles en todo el mundo colonial francés y más allá, junto con ficción y poesía originales y reimpresas de escritores negros del mundo entero. El proyecto, dirigido por un equipo internacional e interdisciplinario de académicos que incluye a la directora del proyecto Maria Beliaeva Solomon, al editor técnico Raffaele Vigiante, y al coeditor y traductor Grégory Pierrot, tiene como objetivo anotar y traducir digitalmente este invaluable registro de la historia global de la colonización, la esclavitud y la abolición en respuesta a la ausencia de una colección entera y con capacidad de búsqueda, y mucho menos traducción, de la tirada completa de la revista. Con el apoyo de la Fundación para el Recuerdo de la Esclavitud, el Centro Schomburg para la Investigación de la Cultura Negra (Biblioteca Pública de Nueva York), los Archivos Nacionales y la Fundación Andrew W. Mellon, el proyecto tiene como objetivo restaurar la retórica emancipadora de la *Revue des Colonies* dentro de los contextos políticos, materiales y culturales de su publicación y hacerlo accesible a las nuevas generaciones.

En un lugar central de esta labor se encuentra la TEI, que funciona como "tecnología de recuperación", tal como lo expresa Kim Gallon en "Making a Case for the Black Digital Humanities" (*Debates in the Digital Humanities*, 2016). TEI no sólo facilita el proceso de digitalización sino que también se convierte en un instrumento para protegerse contra la reproducción involuntaria de las dinámicas de poder incorporadas en los textos originales. Partiendo del imperativo enunciado por Kelly Baker Josephs y Roopika Risam en su introducción a *The Digital Black Atlantic* respecto de resistir la perpetuación histórica de las narrativas dominantes de opresión por parte de la tecnología, el aparato crítico de nuestra edición tiene como objetivo resaltar las contribuciones originales de autores, editores, periodistas y activistas negros a las transformaciones políticas y culturales del siglo XIX. Nuestra personalización de TEI se centra en la codificación de entidades nombradas para proporcionar anotaciones contextuales y críticas. Las funciones de etiquetado representan una gran oportunidad para crear entradas sustanciales y de navegación cruzada para individuos, eventos y organizaciones que han sido ignorados en el discurso académico. Asimismo, ofrecemos traducciones profesionales al inglés para aumentar la accesibilidad a las fuentes de *Revue*.

Por el lado de *Keywords for Black Louisiana (K4BL)*, a partir del otoño de 2020, Leila Blackbird, Olivia Barnard, Emma Katherine Bilski y Ellie Palazzolo, bajo el liderazgo y la colaboración de la directora fundadora Jessica Marie Johnson, la orientación de Alex Gil y la asistencia editorial de Raffaele Vigiante, comenzaron a trabajar en una

microedición de la revista *Scholarly Editing* (<https://scholarlyediting.org/>). El objetivo era editar y publicar un puñado de documentos de la Luisiana colonial del siglo XVIII, transcritos y traducidos a través del Proyecto de Digitalización de Documentos Coloniales del Centro Histórico de Luisiana, para poner el acento en las historias de los africanos y afrodescendientes en ese archivo. K4BL cuenta con el respaldo de la Comisión Nacional de Registros y Publicaciones Históricas.

El equipo ha editado quince historias compuestas por veintidós documentos que abarcan desde 1740 hasta 1795, relacionando el colonialismo francés y el español en Luisiana y ha reflexionado cuidadosamente sobre las posibilidades y las limitaciones de la TEI. Algunos de los documentos originales están en francés y otros en español. Todos han sido transcritos y traducidos al inglés. Las transcripciones y traducciones están marcadas en TEI. Las etiquetas principales utilizadas son <persName> para identificar personas con y sin nombre y <seg> con @ana para palabras clave. La conferencia que dieron Caitlin Pollock y Jessica Lu en 2019, "Hacking TEI for Black Digital Humanities", fue la piedra angular de las Humanidades Digitales negras y la TEI e influyó en el enfoque del proyecto Keywords respecto de la codificación textual desde las primeras etapas del proceso editorial. Pollock y Lu invitaron a los editores a utilizar TEI para trabajar con la historia y los archivos negros, y así ampliar ciertos límites y convenciones que el diccionario documenta tan meticulosamente. Nos encontramos con una serie de preguntas y obstáculos que validaron el llamado de Pollock y Lu, e hicimos un pequeño hackeo que nos proponemos presentar en la reunión anual mientras teorizamos y alentamos futuras intervenciones formales en TEI.

Al reunir estos dos proyectos, buscamos fomentar la conversación sobre cómo la TEI puede volverse más responsable ante las Humanidades Digitales críticas, poscoloniales y negras. Nos basaremos en las similitudes y los desafíos y compromisos compartidos, así como también en las diferencias entre estos dos archivos para lograr un conjunto de intervenciones de gran alcance. Proponemos adiciones al diccionario TEI relacionadas con la representación y traducción del lenguaje racializado, tanto para documentos impresos como manuscritos, y para documentos escritos por, para y sobre personas negras.

Short Papers / Presentaciones breves

Lessons learned from developing a customisable tool for TEI processing and handling of various TEI schemas

Balázs Indig^{1,2}, Mihály Nagy^{2,3}, Luca Horváth⁴

¹Eötvös Loránd University, Department of Digital Humanities; ²National Laboratory for Digital Humanities; ³Eötvös Loránd University, Atelier Department of Interdisciplinary History; ⁴Eötvös Loránd University, Doctoral School of Informatics (Hungary)

Throughout maintenance and systematic extension of five (currently medium-sized) corpora from different genres encoded in TEI schemas we have observed limitations regarding handling and enrichment faced by non-technical researchers. This bottleneck in further document processing steps has hindered our efforts to attract a larger userbase among students and researchers.

While pursuing our goal to standardize the common processing steps that connect to the already standardized data storage format (i.e. TEI schema) we developed a lightweight Python library intended for solving conversion, linguistic annotation, and metadata extraction tasks in a unified manner. Intended for users with minimal technical knowledge, our tool provides a high-level API for a range of TEI-XML-related tasks including validation, format conversion/text and metadata extraction for downstream tasks, and TEI-compatible linguistic annotation.

We distinguish TEI schemas (e.g. for poems, dramas, novels, folk song, news articles) as genres, where each genre represents a unique (valid) TEI document structure. Our library (teitools) consists of an API skeleton that provides handling of built-in genres and allows the easy development of custom bundles to be attached as Python modules without further restrictions. This approach creates a standardized framework extensible to numerous genres.

Our library allows using multiple NLP pipelines to accommodate different languages, while supporting conversion to common output formats (JSONL, customizable HTML, sentence per line, vertical XML format for Sketch Engine corpus query framework) for using our corpora outside of TEI. We have also defined different TEI schema levels to fit NLP and genre-specific annotations, while adhering to the original text. This enables users to generate different annotation levels from raw TEI documents then convert them into another format in batch with only a few API calls programmatically.

Furthermore, we present our observations and experiences and developed best practices regarding compatibility between annotations, TEI structures, and fidelity to original source texts.

Lecciones aprendidas a partir del desarrollo de una herramienta personalizable para el procesamiento TEI y el manejo de diferentes esquemas TEI

Balázs Indig^{1,2}, Mihály Nagy^{2,3}, Luca Horváth⁴

¹Universidad Eötvös Loránd, Departamento de Humanidades Digitales; ²Laboratorio Nacional de Humanidades Digitales; ³Universidad Eötvös Loránd, Departamento de Historia Interdisciplinaria; ⁴Universidad Eötvös Loránd, Escuela de Doctorado en Informática (Hungría)

Hemos observado limitaciones respecto del manejo y enriquecimiento que los investigadores no técnicos enfrentan a lo largo del mantenimiento y la extensión sistemática de cinco corpus (actualmente de tamaño mediano) de diferentes géneros codificados en esquemas TEI. Este impedimento en los pasos posteriores del procesamiento de documentos ha dificultado nuestros esfuerzos por atraer una base de usuarios más grande entre estudiantes e investigadores.

Mientras perseguimos nuestro objetivo de estandarizar los pasos de procesamiento comunes que se conectan al formato de almacenamiento de datos ya estandarizado (es decir, el esquema TEI), desarrollamos una biblioteca Python liviana destinada a resolver tareas de conversión, anotación lingüística y extracción de metadatos de manera unificada. Dirigida a usuarios con conocimientos técnicos mínimos, nuestra herramienta proporciona una API de alto nivel para una variedad de tareas relacionadas con TEI-XML que incluyen validación, conversión de formato/texto y extracción de metadatos para tareas posteriores, y anotaciones lingüísticas compatibles con TEI.

Distinguimos los esquemas TEI, por ejemplo, para poemas, dramas, novelas, canciones populares, artículos de noticias, como géneros en donde cada género representa una única (válida) estructura de documento TEI. Nuestra biblioteca (teiutils) consta de un esqueleto de API que proporciona el manejo de géneros integrados y permite la creación de paquetes personalizados que se adjuntarán como módulos de Python sin mayores restricciones. Este enfoque crea un marco estandarizado extensible a numerosos géneros.

Nuestra biblioteca permite el uso de múltiples canales de PNL para adaptarse a diferentes lenguajes, al tiempo que admite la conversión a formatos de salida comunes (JSONL, HTML personalizable, oración por línea, formato XML vertical para el marco de consulta de corpus de Sketch Engine) para usar nuestros corpus fuera de TEI. Asimismo, hemos definido diferentes niveles de esquema TEI para que se adapten a la PNL y a las anotaciones específicas de género, sin dejar de respetar el texto original. Esto permite que los usuarios generen diferentes niveles de anotación a partir de documentos TEI sin procesar y luego puedan convertirlos, en lote, a otro formato con solo realizar unas pocas llamadas API programáticamente.

Por otro lado, presentamos nuestras observaciones y experiencias y desarrollamos las mejores prácticas en cuanto a la compatibilidad entre anotaciones, estructuras TEI y fidelidad a los textos originales.

EasyDrama: a lightweight solution for encoding plays in TEI-XML

Daniil Skorinkin

Digital Humanities Potsdam (Germany)

Although in many cases TEI-XML markup can be automated, a big share of TEI/XML documents are still encoded manually due to limitations of technology, complexity of the annotated phenomena, or simply the desire of the researcher(s) to stay close to the material and be in full control of the annotation. In such cases, the entry threshold for manual TEI encoding becomes a challenge. To turn raw text into TEI, one has to familiarize oneself with XML and learn heavy-weight annotation tools like Oxygen or CATMA. When it comes to markup workshops with students or non-digital scholars, one has to spend considerable time getting the participants familiar with the tools and the format before the annotation can even begin.

In the DraCor project (dracor.org), it is important to enable people without technical background to encode drama in TEI-XML. Therefore we are working on lowering the encoding threshold. One approach is EasyDrama (github.com/dracor-org/ezdrama) — a lightweight markdown-like language to encode the main structural elements of drama. In EasyDrama, speeches (TEI element `<sp>`), speakers (`<speaker>`), stage directions (`<stage>`), as well as acts and scenes (nested `<div>`-s) are encoded with just a handful of metasympols (`#@$%`) and line breaks. This encoding is automatically translated to TEI/XML following a deterministic procedure. A demo is available at ezdrama.eu.pythonanywhere.com or in the Colab notebook in the repository.

EasyDrama quickly became popular with drama corpora creators. It is even sometimes preferred by people with technical skills and knowledge of XML. The balance between the simplicity of the markup and its unambiguous translation to TEI-XML seems to appeal to encoders. Depending on the uniformity of the source, markup can be accelerated with simple search-replaces, regexes, or LLMs. It is fairly easy to few-shot learn GPT to output EasyDrama, and one still has more control than in end-to-end XML generation.

EasyDrama: una solución ligera para codificar obras de teatro en TEI-XML

Daniil Skorinkin

Humanidades Digitales Potsdam (Alemania)

A pesar de que, en muchos casos, el marcado TEI-XML puede automatizarse, una gran parte de los documentos TEI-XML todavía se codifican manualmente debido a las limitaciones de la tecnología, la complejidad de los fenómenos anotados o, simplemente, el deseo de los investigadores de permanecer cerca del material y tener el control total de la anotación. En tales casos, el umbral de entrada para la codificación manual en TEI se convierte en un desafío. Para convertir texto sin formato en TEI, uno debe familiarizarse con XML y aprender acerca de herramientas de anotación pesadas tales como Oxygen o CATMA. Cuando se trata de talleres de marcado con estudiantes o académicos no digitales, se debe dedicar un tiempo considerable a familiarizar a los participantes con las herramientas y el formato antes de que pueda comenzar la anotación.

En DraCor (dracor.org) consideramos importante permitir que personas sin conocimientos técnicos codifiquen teatro en TEI-XML. Por eso, estamos trabajando para reducir el umbral de codificación. Un enfoque es EasyDrama (github.com/dracor-org/ezdrama), un lenguaje ligero similar a markdown para codificar los principales elementos estructurales del teatro. En EasyDrama, se codifican los discursos (elemento TEI `<sp>`), los oradores (`<speaker>`), las acotaciones escénicas (`<stage>`), así como los actos y escenas (nested `<div>`-s) utilizando tan sólo un puñado de metasímbolos (`#@$%`) y saltos de línea. Esta codificación se traduce automáticamente a TEI/XML siguiendo un procedimiento determinista. Hay una demostración disponible en ezdrama.eu.pythonanywhere.com o en el cuaderno de Colab en el repositorio.

Easy Drama se popularizó rápidamente entre los creadores de corpus dramáticos. Incluso, es, a veces, preferido por personas con habilidades técnicas y conocimientos de XML. El equilibrio entre la simplicidad del marcado y su traducción inequívoca a TEI-XML parece atraer a los codificadores. Dependiendo de la uniformidad de la fuente, se puede acelerar el marcado con simples búsquedas y reemplazos, expresiones regulares o LLMs. Es bastante fácil aprender GPT con pocos ejemplos para generar EasyDrama, y se tiene un control aún mayor que en la generación XML de extremo a extremo.

From Annotations in TEI to Natural Language Processing: A Computational Analysis of Characters in Calderón Drama Corpus

Hanno Ehrlicher¹, Antonio Rojas Castro¹, Sebastian Padó², Kerstin Jung², Allison Keith²

¹University Eberhard Karl of Tübingen, ²University of Stuttgart (Germany)

The TEI-encoded Calderón Drama Corpus (<https://dracor.org/cal>) represents an important milestone by enabling the use of the digital methods for investigating Calderón's work, such as the extent to which rules or genre conventions are followed (Ehrlicher et al., 2020; Lehmann & Padó, 2022). An aspect of this corpus that is highly promising for future research concerns the treatment of characters and character types such as the 'gracioso'. However, character-level information such as gender, social role, honorifics, or character types, is scarce: some of it can be recovered from set lists, some from secondary literature, but independently of the source, normalization and representation remain a challenge. On this poster, we report on our studies in which we enhance the TEI encoding of the Calderón Drama Corpus with character information, where available, and outline how this information can be used to recognize other characters within the vast corpus that fit into these archetypes, based on their speech and social relations, with minimal manual intervention and employing machine learning approaches.

De las anotaciones en TEI al Procesamiento del Lenguaje Natural: un análisis computacional de los personajes del corpus dramático de Calderón

Hanno Ehrlicher¹, Antonio Rojas Castro¹, Sebastian Padó², Kerstin Jung², Allison Keith²

¹Universidad Eberhard Karl de Tübingen, ²Universidad de Stuttgart (Alemania)

El Calderón Drama Corpus codificado en TEI (<https://dracor.org/cal>) representa un hito importante ya que permite el uso de métodos digitales para investigar la obra de Calderón, tales como el grado en que se siguen las reglas o convenciones de género (Ehrlicher et al., 2020; Lehmann y Padó, 2022). Un aspecto de este corpus, que lo hace muy prometedor para futuras investigaciones, tiene que ver con el tratamiento de personajes y los tipos de personajes como el "gracioso". Sin embargo, la información a nivel de carácter como, por ejemplo, género, rol social, honoríficos o tipos de personajes, es escasa. Parte de ella puede recuperarse por medio de listas establecidas, parte de literatura secundaria, pero, independientemente de la fuente, la normalización y la representación siguen siendo un desafío. En este poster informamos sobre nuestros estudios donde mejoramos la codificación TEI del Calderón Drama Corpus con información de personajes, cuando se encuentra disponible, y describimos cómo esta información puede usarse para reconocer otros personajes dentro del vasto corpus que encajan en estos arquetipos en función de su habla y relaciones sociales, mediante una mínima intervención manual y empleando enfoques de aprendizaje automático.

Evolving Hands: Investigating Conversion Transformations for HTR to TEI Data Workflows

James Cummings¹, Diane Jakacki²

¹Newcastle University (United Kingdom), ²Bucknell University (United States of America)

This short paper proposal originates from the Evolving Hands project, but the topic is not the project itself, or Handwritten Text Recognition (HTR), but the conversion workflows for transformation, up-conversion, and enrichment of TEI data produced from HTR. There will only be a short introduction to the project since we've discussed it at previous TEI conferences:

- Transformation: Transkribus exports to Page XML (though also has premium formats like ALTO and TEI). The in-built TEI export had a number of issues when tested, so some of the case studies in the project used the more complete Page XML export and then Dario Kampkaspar's Page2TEI XSLT conversion for transforming the HTR transcripts. The paper will investigate some of the options in transformation of files to TEI P5 XML.
- XML Up-Conversion: Up-conversion is the generation of more detailed markup from a less-detailed source, usually through recognising implicit structural patterns. During up-conversion the data may be supplemented by techniques such as:
 - the probabilistic intuiting of existing data structures based on less-detailed markup;
 - the wholesale replacement of existing hierarchies with more detailed and expressive substitutes;
 - or the determination of extra annotation based on retrieving data from external sources using contextual clues.

As a case study in XML up-conversion the paper uses a dataset of structured print volumes where intra-word formatting is important but ignored by OCR. A cipher-based approach enabled us to preserve the intellectual content.

- Data Enrichment: As a conclusion the paper will point to other forms of data enrichment such as editors like LEAF-Writer and related tools to provide the marking of Named Entities as Linked Open Data as a form of data enrichment.

The presentation will highlight the ease of use of such tools for data enrichment, especially where LEAF-Writer has a dedicated import for HTR-generated TEI P5 XML files.

Manos en evolución: investigando los flujos de trabajo en la conversión de las transformaciones de HTR a TEI

James Cummings¹, Diane Jakacki²

¹Universidad de Newcastle (Reino Unido), ²Universidad Bucknell (Estados Unidos de América)

La propuesta de este artículo breve tiene su origen en el proyecto Evolving Hands, pero el tema no es el proyecto en sí mismo ni el reconocimiento de texto escrito a mano (HTR), sino los flujos de trabajo de conversión para la transformación, conversión ascendente y enriquecimiento de los datos TEI producidos a partir de HTR. Dado que ya hemos discutido el proyecto en conferencias TEI anteriores, haremos solamente una breve introducción:

- Transformación: Transkribus exporta a Page XML (aunque también tiene formatos premium como ALTO y TEI). Debido a que la exportación TEI incorporada tuvo varios problemas cuando se probó, algunos de los estudios de caso del proyecto utilizaron la exportación más completa de Page XML y luego la conversión Page2TEI XSLT de Dario Kampkaspar para transformar las transcripciones HTR. El artículo investigará algunas de las opciones acerca de la transformación de archivos a TEI P5 XML.
- Conversión ascendente XML: La conversión ascendente es la generación de un marcado más detallado a partir de una fuente menos detallada, por lo general, mediante el reconocimiento de patrones estructurales implícitos. Durante la conversión ascendente, los datos pueden complementarse con técnicas tales como:
 - la intuición probabilística de estructuras de datos existentes basada en un marcado menos detallado;
 - el reemplazo total de las jerarquías existentes por sustitutos más detallados y expresivos;
 - o la determinación de anotaciones adicionales basadas en la recuperación de datos de fuentes externas utilizando pistas contextuales.

Como estudio de caso sobre conversión ascendente XML, se utiliza un conjunto de datos de volúmenes impresos estructurados donde el formato intrapalabra es

importante pero el OCR lo ignora. Un enfoque basado en cifrado nos permitió preservar el contenido intelectual.

- Enriquecimiento de datos: Como conclusión, se señalarán otras formas de enriquecimiento de datos, por ejemplo, editores como LEAF-Writer, y herramientas relacionadas para marcar entidades nombradas como Datos Abiertos Enlazados como una forma de enriquecimiento de datos. Se destacará la facilidad de uso de dichas herramientas para el enriquecimiento de datos, especialmente donde LEAF-Writer tiene una importación dedicada para los archivos XML TEI P5 generados por el HTR.

C is for Co(nsortium|uncil|laboration|ntributors|mmunity), or What Can GitHub Issues Tell Us About the TEI?

Joey Takeda

Simon Fraser University (Canada)

The TEI's GitHub organization is the central home for development of the TEI Guidelines, Stylesheets, and many other associated tools, projects, and working groups. Among other things, each repository contains every version of every source file, a full log of every change committed, a list of all releases and their source files, and a list of all completed and outstanding issues (or "tickets"). These issues are key to the distributed, asynchronous, and transparent work of TEI Technical Council and, much like the TEI listServ, the repositories provide an important channel for the TEI community to propose and suggest changes, raise issues, and ask questions.

But they also serve as an incredibly useful record of the TEI's development work over the last decade or so (since the migration of the TEI's codebase from SourceForge in 2015). Every bug or feature request is (theoretically) logged in the repository, which also (theoretically) chronicles a history of that particular issue: who raised it, who resolved it, who responded, when was it closed, and under what circumstances.

Drawing on recent research and initiatives into evaluating and measuring the "health" of open source code and communities, this paper investigates what an analysis of the TEI's GitHub data might yield for understanding the relationship between the various groups that make up the TEI-C: council, consortium, contributors, collaborators, and community. Using metrics defined by the Linux Foundation project "Community Health Analytics in Open Source Software" (CHAOSS) (e.g. time to first response, time to close, and "bus factor"), this paper will present a critical analysis of the issues raised on the TEI's two primary GitHub repositories (TEIC/TEI and TEI/Stylesheets) and a

discussion of what, if anything, these metrics can tell us about the past, present, and future of the TEI.

C quiere decir Co(nsortio|nsejo|laboración|laboradores|munidad), o ¿Qué nos pueden decir los GitHub Issues sobre TEI?

Joey Takeda

Universidad Simon Fraser (Canadá)

La plataforma GitHub de la TEI es el lugar central para el desarrollo de las directrices, hojas de estilo y muchas otras herramientas, proyectos y grupos de trabajo asociados con TEI. Entre otras cosas, cada repositorio contiene cada versión de cada archivo fuente, un registro completo de cada modificación, una lista de todas las publicaciones y de sus archivos fuente, y una enumeración de todos los problemas completados y pendientes (o "tickets"). Estas cuestiones son clave para el trabajo distribuido, asincrónico y transparente del Consejo Técnico de TEI y, al igual que la mailing list de la TEI (TEI listServ), los repositorios proporcionan un canal importante para que la comunidad de la TEI proponga y sugiera cambios, plantee problemas y haga preguntas.

Pero también sirven como un registro increíblemente útil del trabajo de desarrollo de la TEI durante la última década (a partir de la migración de la base de código TEI de la plataforma SourceForge en 2015). Cada error o solicitud de función se registra (teóricamente) en el repositorio, que también (teóricamente) registra una historia de ese problema en particular: quién lo planteó, quién lo resolvió, quién respondió, cuándo se cerró y bajo qué circunstancias.

Basándose en investigaciones e iniciativas recientes para evaluar y medir la "salud" del código fuente abierto y las comunidades, este trabajo investiga qué podría arrojar un análisis de los datos TEI en GitHub a fin de comprender la relación entre los diversos grupos que conforman el Consorcio de la TEI (TEI-C): consejo, consorcio, contribuyentes, colaboradores y comunidad. Utilizando métricas definidas por el proyecto "Análisis de salud comunitaria en software de código abierto" (CHAOSS) de la Fundación Linux (p. ej., tiempo de respuesta inicial, momento de cierre y "factor autobús"), este documento presentará un análisis crítico de las cuestiones planteadas en los dos repositorios principales de GitHub de la TEI (TEIC/TEI y TEI/Stylesheets) y una discusión respecto de lo que estas métricas podrían decirnos sobre el pasado, presente y futuro de la TEI.

A minimalist solution to the problem of representing the network of artistic motifs in the Libro de Alexandre

Gabriel Calarco

National Council for Scientific and Technical Research (Argentine Republic)

This short paper aims to address a challenge encountered in the initial stages of developing my digital edition of ekphrasis fragments from the Libro de Alexandre. Specifically, the markup of artistic motifs within these fragments emerged as a pivotal aspect of my research, alongside the creation of a website to enable reader interaction with these marked motifs through digital means. I'll discuss the solution adopted within the framework of minimal computing, which facilitated achieving this goal via a static website. Additionally, I'll explore the creation of a controlled vocabulary, where each motif, submotif, and specific term is assigned a unique identifier within the hierarchical structure of a taxonomy. Finally, I'll reflect on the practical and theoretical implications of navigating this challenge, arguing that what began as a practical solution to technological and financial constraints evolved into a fundamental aspect of my theoretical approach to representing ekphrasis in medieval Spanish poetry through digital means.

Una solución minimalista al problema de representar la red de motivos artísticos en el Libro de Alexandre

Gabriel Calarco

Consejo Nacional de Investigaciones Científicas y Técnicas (República Argentina)

Esta presentación tiene como objetivo abordar un desafío encontrado en las etapas iniciales del desarrollo de mi edición digital de fragmentos de ékfrasis del Libro de Alexandre. Específicamente, el marcado de motivos artísticos dentro de estos fragmentos surgió como un aspecto fundamental de mi investigación, junto con la creación de un sitio web para permitir la interacción del lector con estos motivos marcados a través de medios digitales. Discutiré la solución adoptada en el marco de la computación mínima, que facilitó el logro de este objetivo a través de un sitio web estático. Asimismo, exploraré la creación de un vocabulario controlado en el que se asigna un identificador único dentro de la estructura jerárquica de una taxonomía a cada motivo, submotivo y término específico. Finalmente, reflexionaré sobre las implicaciones prácticas y teóricas de afrontar este desafío, argumentando que, lo que comenzó como una solución práctica a las limitaciones tecnológicas y financieras,

evolució hasta convertirse en un aspecto fundamental de mi enfoque teórico para representar la écfrafrasis en la poesía española medieval a través de medios digitales.

The accidental maximalist: Or, what's so minimal about minimal editions?

Federico Giannetti

Rutgers University (United States of America)

Minimal editions have become popular because they answer the problem of curating texts under a set of constraints. However, encoding and publication workflows, even for minimal editions, run the risk of becoming bloated if the editor loses sight of the project's priorities. Over a project's life cycle, it is normal that these priorities will shift under the pressure of this or that constraint or challenge. Several authors have noted that minimal computing stacks displace complexity away from users and onto the editor or technical partner (Dombrowski, 2022; Giannetti, 2019; Hughes, 2016). Identifying the necessary technical complexity is rarely easy. I focus on the latter two of Risam and Gil's four-question heuristic for minimal computing –“what must we prioritize?” and “what are we willing to give up?” –in order to demonstrate the importance of reengaging with these questions as new challenges come to light (Risam and Gil, 2022).

With the Personal Correspondence from the Rutgers College War Service Bureau, I initially made choices consonant with a minimal approach. The edition navigation replicated the file structure of the archival collection, organized by the name of the alumnus serving in World War I. The project schema exhibited straightforward choices regarding the encoding of people, places, and events. However, the plan to capture biographical information on each Rutgers person mentioned nearly ran aground when a single soldier mentioned fifty classmates in his letters. A desire for a more enriching reader experience led to a choice to provide subject access to the letters, which in turn brought classification and UI difficulties. In this case study, I provide substance to the hard decisions editors face amidst evolving priorities and the need to curtail some plans. Documentation of best practices in this area will serve other editors who must make pragmatic choices in the service of local knowledge production.

El maximalista accidental o ¿qué tienen de mínimo las ediciones mínimas?

Federico Giannetti

Universidad de Rutgers (Estados Unidos de América)

Las ediciones mínimas se han vuelto populares porque responden al problema de curar textos bajo una serie de limitaciones. Sin embargo, los flujos de trabajo de codificación y publicación, incluso para ediciones mínimas, corren el riesgo de ser desmesurados si el editor pierde de vista las prioridades del proyecto. A lo largo del ciclo de vida de un proyecto, es normal que estas prioridades cambien por la presión de tal o cual limitación o desafío. Varios autores han observado que las pilas de computación mínima desplazan la complejidad de los usuarios hacia el editor o el socio técnico (Dombrowski, 2022; Giannetti, 2019; Hughes, 2016). Rara vez es fácil identificar la complejidad técnica necesaria. Me concentro en las dos últimas de las cuatro preguntas de evaluación heurística respecto de la computación mínima formuladas por Risam y Gil: “¿qué debemos priorizar?”; y “¿a qué estamos dispuestos a renunciar?”, para demostrar la importancia de volver a abordar estas preguntas a medida que nuevos desafíos salen a la luz (Risam y Gil, 2022).

Al trabajar con la correspondencia personal de la Oficina de Servicios de Guerra de Rutgers College, inicialmente, tomé decisiones en consonancia con un enfoque mínimo. La navegación de la edición replicaba la estructura de archivos de la colección de archivos, organizada por el nombre del ex alumno que sirvió en la Primera Guerra Mundial. El esquema del proyecto mostraba opciones sencillas con respecto a la codificación de personas, lugares y eventos. Sin embargo, el plan para capturar información biográfica sobre cada persona de Rutgers mencionada casi se estancó cuando un solo soldado mencionó a cincuenta compañeros de clase en sus cartas. El deseo de lograr una experiencia de lectura más enriquecedora llevó a darle acceso a los sujetos a las cartas, que, a su vez, trajo dificultades de clasificación y de interfaz de usuario. En este estudio de caso, confirmé las decisiones difíciles que deben afrontar los editores frente a las prioridades cambiantes y la necesidad de restringir algunos planes. La documentación respecto de las mejores prácticas en esta área será útil para otros editores que deben tomar decisiones pragmáticas al servicio de la producción de conocimiento local.

MiniFierro: la voz del gaucho a través del mercado TEI

María Agustina Ryckeboer

Universidad del Salvador (República Argentina)

Desde el Laboratorio de Humanidades Digitales del IIBICRIT del CONICET (HD LAB) estamos elaborando la edición digital de la primera publicación del poema gaucho Martín Fierro del año 1872. Esta edición busca ser una lectura cercana del mismo. A través del mercado TEI, se rescatan aspectos que no se encuentran presentes en otras

ediciones digitales. En este proceso se nos han presentado varios desafíos: desde localizar una digitalización de buena calidad a realizar elecciones dentro del mercado.

Nuestro objetivo es rescatar todos los aspectos del poema que resultan llamativos en su presentación y presentar un vocabulario que haga más ligera su lectura y comprensión. Lo más llamativo de este poema es que, no solo relata la vida del gaucho y su cultura, sino que refleja su modo de hablar particular, no solo relata la vida del gaucho y su cultura, sino que refleja su modo de hablar particular, lo que aporta su singular riqueza al poema.

Nuestra edición resalta la edición príncipe del poema, siendo lo más fiel posible a la misma para, así, destacar su valor como una de las obras cumbres de la literatura argentina y de la cultura gauchesca.

MiniFierro: the voice of the gaucho through TEI markup

María Agustina Ryckeboer

University of Salvador (Argentine Republic)

At the Laboratorio de Humanidades Digitales del IIBICRIT del CONICET (HD LAB) we are currently working on the digital edition of the first publication of the gaucho poem *El Gaucho Martín Fierro* (1872). This edition aims to be a close reading of it. By means of the TEI markup, aspects that are not present in other digital editions are conserved. Several challenges have arisen in this process –from locating a good quality digitization to making choices within the marking.

Our objective is to preserve all the elements of the poem that are remarkable regarding its presentation, and to provide a controlled vocabulary so as to make its reading and understanding easier. What is most striking about this poem is that it not only describes the life of the gaucho and his culture, but also reflects his particular way of speaking, which gives the poem unique richness.

Our edition highlights the first edition of the poem, remaining as faithful as possible to it so as to stress the value of this masterpiece of Argentine literature and gaucho culture.

Edición digital anotada del Diálogo sobre la historia de la pintura en México

María Teresa Ravelo Sánchez

Instituto de Investigaciones Estéticas (México)

El Diálogo sobre la historia de la pintura en México es considerado un clásico en la historia del arte mexicano por ser el primer texto que ofrece una síntesis sobre la pintura colonial. Me propuse crear un recurso primario de consulta, con materiales no descargables que incluya el texto codificado con el estándar XML-TEI y criterios de Minimal Computing, notas de contexto, el facsímil digital del libro de 1899 y una línea de tiempo con los hechos y pinturas mencionados. Para la edición he usado los módulos de TEI obligatorios (core, header, tei y textstructure) y los opcionales drama, namesdates, transcr y verse; lenguajes de marcado de hipertexto (HTML), de diseño (CSS) y programación (JavaScript y Ruby), y un generador simple para sitios web estáticos (Jekyll). El proyecto es parte del trabajo final del Máster en Humanidades Digitales de la UNED dirigido por la doctora Gimena del Río Riande.

Annotated digital edition of the Diálogo sobre la historia de la pintura en México

María Teresa Ravelo Sánchez

Aesthetic Research Institute (México)

The Diálogo sobre la historia de la pintura en México is considered a classic text about the history of Mexican art, as it is the first one to offer a synthesis of colonial painting. I set out to create a primary reference resource, with non-downloadable materials, which includes the text encoded following the XML-TEI standard and Minimal Computing criteria, context notes, the digital facsimile of the 1899 book and a timeline with the facts and paintings mentioned. For the editing process, I have used the minimal TEI structure modules (core, header, tei and textstructure) as well as drama, namesdates, transcr and verse; hypertext markup (HTML), design (CSS) and programming languages (JavaScript and Ruby), and a static site generator (Jekyll). The project is part of the thesis of the MA in Digital Humanities at UNED directed by Dr. Gimena del Río Riande.

Descripción del póster de la Edición crítica digital de los documentos del archivo de María Mercedes Carranza relacionados con la ANC

Lorenzo A. Almarío

Universidad de los Andes (Colombia)

El póster presentará el diseño preliminar de una edición crítica digital de los documentos vinculados a la participación de María Mercedes Carranza (1945-2003), poeta, periodista y gestora cultural colombiana, en la Asamblea Nacional Constituyente (ANC) de Colombia en 1991. El objetivo principal de esta investigación es contribuir al

reconocimiento de la relevancia de la participación de Carranza en la ANC mediante la creación de una edición crítica digital basada en su archivo personal. Esta iniciativa estará acompañada de una estrategia de divulgación web destinada a ampliar el alcance de la edición.

El contenido del póster inicia por una introducción concisa que presenta el problema de investigación y sus objetivos. A continuación, se encuentra una sección dedicada a la metodología. Esta sección se centra en el uso del estándar TEI (Text Encoding Initiative) para el marcado de todo el material epistolar de Carranza relacionado con el proceso constituyente, así como la cobertura periodística del mismo y las leyes discutidas en la comisión primera, de la cual ella fue parte. Se muestra en esta sección los detalles metodológicos clave que se entrelazan con debates teóricos relevantes, como la idoneidad del uso del estándar TEI, sus desafíos (como las etiquetas en inglés y la escasez de documentación en español, así como la interoperabilidad), y sus beneficios (alta flexibilidad y facilidad de aprendizaje).

También se aborda la metodología para la publicación de la edición, que implica el uso de computación mínima y se plantea una discusión en torno a qué podemos entender por minimalismo en las Humanidades Digitales del Sur Global. Por último, se presentan los resultados esperados y la bibliografía pertinente. Este diseño permite visualizar de manera clara y concisa los objetivos, métodos y potenciales contribuciones de la investigación, así como las implicaciones teóricas y prácticas de esta.

Description of the poster regarding the Digital Critical Edition of the documents in María Mercedes Carranza's archive related to the ANC

Lorenzo A. Almario

University of the Andes (Colombia)

This poster will show the preliminary design of a digital critical edition of the documents connected to the participation of the Colombian poet, journalist and cultural manager María Mercedes Carranza (1945-2003) in Colombia's 1991 National Constituent Assembly (ANC). The main goal of this research is to contribute to the acknowledgement of Carranza's participation in the ANC by creating a digital critical edition based on her personal archive. This initiative will be accompanied by a web dissemination strategy aimed at expanding the range of the edition.

This poster includes a brief introduction stating the research problem and its objectives. Moreover, there is a section dedicated to methodology. This section focuses on the use of the TEI standard for marking all of Carranza's epistolary material related to the constituent process, as well as the journalistic coverage of it and the laws discussed in the first commission she took part in. This section shows key methodological details

that intertwine with relevant theoretical debates like the adequacy of using the TEI standard, its challenges –English tags and scarce documentation in Spanish, as well as interoperability– and its benefits –high flexibility and ease of learning.

The methodology followed in publishing the edition, which involves the use of minimal computing, is also addressed and a discussion regarding what we understand by minimalism in the Digital Humanities of the Global South is raised. Last but not least, the expected results and the relevant bibliography are presented. This design provides a clear and concise visualization of the objectives, methods and potential contributions of this research, as well as its theoretical and practical implications.

Long Papers / Presentaciones extensas

Interfaces de visualización de ediciones digitales académicas en XML-TEI: el caso de la Biblioteca Digital del Progetto Mambrino

Stefano Bazzaco

Universidad de Verona (Italia)

En el contexto de las Humanidades Digitales la creación de ediciones digitales académicas es uno de los campos de estudio que sigue suscitando más interés entre los especialistas del texto. Desde los años 90, cuando las ediciones en formato electrónico aún se distribuían en soportes físicos, se han dado varios pasos, entre los cuales los más importantes han interesado la modelización de los contenidos y la fijación de un lenguaje de marcado como XML-TEI, pormenorizado y flexible, que se ha convertido en un estándar de facto para la remediación de documentos analógicos al entorno digital. Adaptar una edición digital académica al formato XML-TEI, por lo tanto, asegura que los datos se preserven y sobrevivan más allá de las herramientas de publicación y contemporáneamente estimula la interoperabilidad entre proyectos. Sin embargo, la difusión de este lenguaje de marcado en parte ha dificultado la creación y difusión de interfaces de publicación estándar, que, si no pasan por una customización, no son aptas para reproducir la complejidad intrínseca del documento procesado.

La presente comunicación tiene el objetivo de reflexionar sobre el estado actual de las interfaces de visualización de las ediciones digitales académicas y aportar nuevos datos acerca de las posibilidades de publicación de ficheros textuales en el formato XML-TEI. El tema se abordará a partir de la experiencia madurada por el Progetto Mambrino, grupo de investigación de la Universidad de Verona, en el contexto de la

creación de una Biblioteca Digital de las novelas caballerescas italianas del Renacimiento, que son traducciones y continuaciones del famoso género español de los libros de caballerías. En concreto, se relatan las distintas fases que dentro del proyecto han llevado a la personalización de la aplicación TEI Publisher para la visualización de aspectos textuales de naturaleza variada, como la estructura lógica y los aspectos semánticos del documento fuente. Al respecto, se presenta un modelo de visualización que integra dos distintos ficheros TEI y los comulga en un único artefacto digital, permitiendo la consulta del texto desde distintas perspectivas de investigación. Finalmente, se centra la atención en aspectos relacionados con la accesibilidad y la experiencia del usuario, indicando posibles vías para la exportación del modelo de representación empleado y su posible migración a otros contextos de estudio.

Visualization interfaces of scholarly digital editions in XML-TEI: the case of the Progetto Mambrino Digital Library

Stefano Bazzaco

University of Verona (Italy)

Creating scholarly digital editions is one of the fields of study that continues to arouse interest among text specialists in the context of the Digital Humanities. Since the 90s, when editions in electronic format were still distributed on physical media, several steps have been taken, among which the most important have concerned modelling the contents and establishing a markup language such as XML-TEI, both detailed and flexible, which has become a de facto standard for the remediation of analog documents to the digital environment. Adapting a scholarly digital edition to the XML-TEI format therefore ensures that data is preserved and survives beyond publishing tools, and, at the same time, stimulates interoperability between projects. However, the diffusion of this markup language has partly hindered the creation and dissemination of standard publishing interfaces, which are not suitable for reproducing the intrinsic complexity of the processed document unless they undergo customization.

The aim of this work is to reflect on the current state of the visualization interfaces of scholarly digital editions and to provide new data regarding the possibilities of publishing text files in XML-TEI format. Progetto Mambrino, a research group at the University of Verona, will address this topic from the experience gained after creating a Digital Library of Italian chivalric literature in the Italian Renaissance, mainly translations and continuations of the Spanish books of chivalry. Specifically, we describe the different phases within the project leading to the customization of the TEI Publisher application for visualizing diverse textual aspects, such as the logical structure and semantic aspects of the source document. In this regard, we present a visualization

model integrating two different TEI files into a single digital artifact which allows consultation of the text from different research perspectives. Eventually, we focus on aspects related to accessibility and user experience, showing possible ways for exporting the representation model used and its possible migration to other contexts.

TEI and IIIF Technologies in the Native Bound Unbound Project

Nick Laiacona

Performant Software (United States of America)

The Native Bound Unbound project seeks to document the lives of indigenous enslaved people in the Americas. Led by Dr. Estevan Rael-Gálvez, the project aims to tell the stories of the millions of indigenous people who were enslaved since the arrival of Columbus. Our firm, Performant Software Solutions, was selected to develop and implement the data infrastructure and website for Native Bound Unbound. In this presentation, we will discuss our approach, which utilizes technologies based on the TEI Guidelines and the International Image Interoperability Framework (IIIF).

To create the fullest picture we can of the individuals whose lives are being documented, we need to gather information from as many sources as possible. This means we must manage many different types of primary sources, including: baptismal records, census records, court cases, tombstones, and oral histories. Texts are often handwritten in a variety of languages including Spanish, English, and Dutch. From this material, we need to distill information about individuals, events in their lives, and the places they lived and worked.

One possible approach would be to read all this material and simply key it into a database. However, this would not make it possible for other researchers to examine our evidence and draw their own or further conclusions. Furthermore, this project is for the public as much as it is for researchers. So we need to both transcribe and translate the material we are working with so that we can display it on the website alongside the page images. These sources are then linked to the records generated from them and the whole process is open to inspection.

Structuring the texts and marking up the people, places, and events in them allows us to programmatically enter the records into the database. These then need to be scanned for duplication and de-duplicated. Updates to the database or the transcriptions should not cause a loss of data or duplicate data. And, of course, our understanding of what information we want to collect and can collect has been evolving throughout this process, so we need a certain level of flexibility in our data models.

Performant brought to this project an array of open source projects that we have developed for previous digital humanities projects, plus some projects developed by other groups that aligned with our client's needs. These tools include: FairCopy, Core Data, EditionCrafter, FromThePage, Splink, and our IIIF Cloud CMS. These tools, taken together, provide a complete workflow for the data infrastructure of this ambitious project. They also provide an application programming interface on top of which we were able to construct the public facing website. This presentation will provide a detailed look at our technical approach, these tools, and the final project website.

Tecnologías TEI e IIIF en el proyecto Native Bound Unbound

Nick Laiacona

Performant Software (Estados Unidos de América)

El proyecto Native Bound Unbound busca documentar las vidas de los pueblos indígenas esclavizados en las Américas. El proyecto, dirigido por el Dr. Estevan Rael-Gálvez, tiene como objetivo contar las historias de los millones de indígenas que fueron esclavizados desde la llegada de Colón. Nuestra empresa, Performant Software Solutions, fue seleccionada para desarrollar e implementar la infraestructura de datos y el sitio web de Native Bound Unbound. En esta presentación, discutiremos nuestro enfoque, que utiliza tecnologías basadas en las Directrices TEI y el Marco Internacional de Interoperabilidad de Imágenes (IIIF).

Para crear la imagen más completa que podamos de las personas cuyas vidas se están documentando, necesitamos recopilar información de tantas fuentes como sea posible. Esto significa que debemos gestionar diversos tipos de fuentes primarias, incluidos: registros bautismales, registros censales, casos judiciales, lápidas e historias orales. Los textos suelen estar escritos a mano en una variedad de idiomas, incluidos español, inglés y holandés. A partir de este material, necesitamos extraer información sobre las personas, los acontecimientos de sus vidas y los lugares en los que vivieron y trabajaron.

Un posible enfoque sería leer todo este material y simplemente ingresarlo en una base de datos. Sin embargo, esto no permitiría a otros investigadores examinar nuestra evidencia y sacar sus propias conclusiones o conclusiones adicionales. Asimismo, este proyecto está dirigido tanto al público como a los investigadores. Por lo tanto, necesitamos transcribir y traducir el material con el que estamos trabajando para poder mostrarlo en el sitio web junto con las imágenes de la página. Estas fuentes se vinculan luego a los registros generados a partir de ellas y todo el proceso queda abierto a inspección.

Estructurar los textos y marcar las personas, lugares y eventos que contienen nos permite ingresar los registros en la base de datos mediante programación. Es necesario escanearlos luego para detectar duplicaciones y eliminar duplicados. Las actualizaciones de la base de datos o de las transcripciones no deberían causar pérdida ni duplicación de datos. Y, claramente, nuestra comprensión respecto de qué información queremos y podemos recopilar ha ido evolucionando a lo largo de este proceso, por lo que necesitamos un cierto nivel de flexibilidad en nuestros modelos de datos.

Performant aportó a este proyecto una serie de proyectos de código abierto que hemos desarrollado para proyectos de humanidades digitales previos, sumado a algunos proyectos desarrollados por otros grupos que se alinearon con las necesidades de nuestros clientes. Estas herramientas incluyen: FairCopy, Core Data, EditionCrafter, FromThePage, Splink y nuestro IIF Cloud CMS. Este conjunto de herramientas proporciona un flujo de trabajo completo para la infraestructura de datos de este ambicioso proyecto. También proporcionan una interfaz de programación de aplicaciones sobre la cual pudimos construir el sitio web público. Esta presentación brindará una visión detallada de nuestro enfoque técnico, de estas herramientas, y del sitio web final del proyecto.

Good practice built in: case study of e-editions community

Magdalena Turska

e-editions (Poland)

Digital Humanities research usually works with very diverse datasets and tools: from facsimiles, through HTR or OCR, transcription, collation, entity recognition, authority links to further annotation and, eventually, publication online or in print. So much more is within our easy reach than even just a decade ago and these opportunities are widely embraced in our community. Nevertheless, creation of a digital resource is an intrinsically complex subject, as measured by a random assortment of acronyms from TEI to IIF, so the average scholar can't be expected to have an overview of the ever fluctuating technological landscape. And yet she is still ultimately responsible for designing and orchestrating complex, collaborative workflows and has to directly face the short- and long-term consequences of each decision.

I will discuss what happened when a community of practitioners instead of accepting the unfavourable situation, worked together firmly prioritizing the development of a generic, systematic solution above the specific, without compromising on the latter, analyzing the connected cases of TEI Simple, TEI Processing Model, TEI Publisher

framework, and collectives like e-editions and Sources Online. I would like to demonstrate the visible impact these collaborations, grounded in the mottos of "power to the editor" and "standardize where you can, customize where you must" had on the landscape of digital scholarly editions in recent years.

With Publisher now in version 9 and an impressive international community of e-editions, we can make some interesting observations. About 40 projects chose to register on the e-editions website, similar number is currently in preparation and we estimate perhaps another hundred that we never yet heard about. I will talk about various observable trends among editorial projects that we can trace back to resources - software releases, demos and samples - made available by the community across the years. These range from functional features to adopting the text encoding strategies and data organization practices, to standards and open libraries, e.g. XML vocabularies like TEI, JATS or DocBook; MathML and TEX; IIF for the images together with OpenSeaDragon and Tify viewers; Verovio for sheet music; OpenAPI for the specification of the programmatic interface and DTS for machine-consumption of digital text collections and so on.

This has a profound effect on establishing a consensual good practice and base level of interoperability for digital humanities projects. For example, each and every TEI Publisher-based application exposes well documented programmatic interface, e.g. DTS protocol or authority registries. When it's available as default and by design integrated into the framework we treat interoperability as a priority and achieve it out of the box. Similarly, the re-use approach led to emergence of specialized data models, successfully adopted wholesale or customized by different projects.

Concluding, the availability of an open source framework which allows users to mix existing solutions, strengthened by a supportive community of practice has proven extremely successful. For a practical demonstration of the power this approach affords, I will discuss the Sources Online project.

Buenas prácticas incorporadas: estudio de caso de la comunidad e-ediciones

Magdalena Turska

e-ediciones (Polonia)

La investigación en Humanidades Digitales generalmente trabaja con conjuntos de datos y herramientas muy diversos: desde facsímiles, pasando por HTR u OCR, transcripción, cotejo, reconocimiento de entidades, enlaces de autoridad hasta anotaciones adicionales y, eventualmente, publicación en línea o impresa. Hay mucho más a nuestro alcance que hace apenas una década y estas oportunidades son ampliamente aceptadas en nuestra comunidad. Sin embargo, dado que la creación de

un recurso digital es un tema intrínsecamente complejo al ser medido desde TEI hasta IIF, no se puede esperar que el académico promedio tenga una visión general del panorama tecnológico en constante fluctuación. Y, sin embargo, tiene la responsabilidad de diseñar y gestionar flujos de trabajo complejos y colaborativos, y el deber de afrontar directamente las consecuencias a corto y largo plazo de cada decisión tomada.

Discutiré lo que sucedió cuando una comunidad de profesionales, en lugar de aceptar la situación desfavorable, trabajaron juntos priorizando firmemente el desarrollo de una solución genérica y sistemática por encima de la específica, sin comprometer a esta última, analizando los casos relacionados de TEI Simple, el modelo de procesamiento TEI, el marco TEI Publisher y colectivos tales como e-ediciones y Sources Online. Me gustaría demostrar el impacto visible que estas colaboraciones, basadas en los lemas de "poder para el editor" y "estandarizar donde puedas, personalizar donde debes", tuvieron en el ámbito de las ediciones filológicas digitales en los últimos años.

Ahora que se lanzó la versión 9 de Publisher y que hay una comunidad internacional notable de e-ediciones, podemos hacer algunas observaciones interesantes. Alrededor de 40 proyectos optaron por registrarse en el sitio web de e-ediciones, un número similar está actualmente en preparación y estimamos quizás otros cien de los que todavía no hemos oído hablar. Me referiré a varias tendencias observables entre los proyectos editoriales de los que podemos rastrear hasta los recursos (lanzamientos de software, demostraciones y muestras) puestos a disposición por la comunidad a lo largo de los años. Estos van desde características funcionales hasta la adopción de estrategias de codificación de texto y prácticas de organización de datos, pasando por estándares y bibliotecas abiertas, p.ej. vocabularios XML tales como TEI, JATS o DocBook; MathML y TEX; IIF para las imágenes junto con los visores OpenSeaDragon y Tify; Verovio para partituras; OpenAPI para la especificación de la interfaz programática y DTS para el consumo automático de colecciones de texto digitales, etc.

Esto tiene un efecto profundo en el establecimiento de buenas prácticas consensuadas y de un nivel básico de interoperabilidad para proyectos de humanidades digitales. Por ejemplo, todas y cada una de las aplicaciones basadas en TEI Publisher exponen una interfaz programática bien documentada, p. ej. protocolo DTS o registros de autoridad. Cuando la interoperabilidad se encuentra disponible de forma predeterminada y por diseño integrado en el marco, la tratamos como una prioridad y la alcanzamos de manera inmediata. De manera similar, el enfoque de reutilización condujo a la aparición de modelos de datos especializados, adoptados con éxito de forma total o personalizados de acuerdo a los diferentes proyectos.

En conclusión, la disponibilidad de un marco de código abierto que permite a los usuarios combinar soluciones existentes, fortalecido por una comunidad colaborativa

de práctica, ha demostrado ser extremadamente exitosa. Para llevar a cabo una demostración práctica del poder de este enfoque, analizaré el proyecto Sources Online.

¿Cómo editar los censos de esclavos?

Clayton McCarl

University of North Florida (United States of America)

Entre los siglos XVI y XIX, se prepararon en varias partes de las Américas enumeraciones de “propiedad humana”. Algunas son documentos mercantiles relacionados con el transporte y la venta de personas esclavizadas, otras son inventarios que figuran en los testamentos de los esclavizadores, y otras son registros burocráticos a propósito del cobro de impuestos y de los procesos de manumisión. Frecuentemente señalan el nombre, el sexo, y la edad de cada persona. A veces también indican una ocupación y, con una indiferencia espeluznante, una valoración monetaria. Son textos que nos pueden ayudar a rellenar vacíos que existen en la historiografía sobre la esclavitud y las poblaciones afrodescendientes, tanto en Latinoamérica como en los Estados Unidos. Son también documentos que nos permiten entrever aspectos íntimos de las historias de individuos y a veces de familias, además de pensar en la relación entre la escritura y la opresión.

En esta presentación reflexiono sobre los retos a que nos enfrentamos a la hora de editar estos textos. ¿Cómo transmitir la información que contienen sin reducir a las personas a meros datos? ¿Cómo no victimizar de nuevo a los individuos que aparecen en ellos? ¿Cuáles son nuestras responsabilidades como editores ante textos tan deshumanizantes? Repaso primero las ideas de otros estudiosos que han editado documentos relacionados con la esclavitud. Después, explico los experimentos que he hecho en colaboración con mis estudiantes a propósito de dos textos del siglo XIX: un censo de esclavos de la década de 1840 en la región de Antioquia, Colombia, y el testamento de 1857 de Abraham DuPont, el dueño de una plantación en el condado de St. Johns en Florida. Comparto las ideas preliminares que hemos desarrollado en cuanto a un modelo para el marcado que tome en cuenta la humanidad de las personas enumeradas. También considero el diseño de interfaces que nos dejen interactuar con los textos marcados como conjuntos de seres humanos y no como simples aglomeraciones de datos.

How can we edit slave schedules?

Clayton McCarl

Universidad de North Florida (Estados Unidos de América)

Between the 16th and 19th centuries, enumerations of “human property” were compiled in various parts of the Americas. Some are commercial documents related to the transportation and sale of enslaved people, others are inventories that appear in the wills of the enslavers, and others are bureaucratic records regarding tax collection and manumission. They frequently indicate the name, sex and age of each person. Sometimes, they also state an occupation and, with horrifying indifference, a monetary value. These are texts that can help us fill in the existing gaps regarding the historiography of slavery and Afro-descendants both in Latin America and in the United States. These are also documents that offer a glimpse into intimate aspects of the stories of individuals and, sometimes, families. In addition, they enable us to think about the relationship between writing and oppression.

In this presentation I reflect on the challenges we face when editing these texts. How do we transmit the information they contain without reducing people to mere data? How can we avoid revictimising the individuals who appear in them? What are our responsibilities as editors when confronted with such dehumanization? First, I review the ideas of other scholars who have edited documents related to slavery. Next, I explain the experiments I have carried out in collaboration with my students concerning two 19th-century texts: a census of slaves from the 1840s in the region of Antioquia, Colombia, and the 1857 will of Abraham Dupont, the owner of a plantation in St. John's County in Florida. I share our preliminary ideas regarding a markup model that takes the humanity of the people listed into account. In addition, I take the interface design into consideration since it enables us to interact with those marked texts viewing them as a group of human beings and not as mere data agglomeration.

Encoding Travel Literature: Analyzing European Cultural Heritage through the Perspective of Latin American Women Writers

Alba Comino

NOVA University of Lisbon (Portugal)

The representation of cultural heritage in travel literature often reflects a curated selection influenced by various factors, from authorised discourses shaped by prior readings or guided tours to the personal interests of the authors. We unveil a dynamic historical narrative by conducting a comparative analysis of the insights provided by Latin American women authors regarding European cultural heritage at the turn of the

20th century. This exploration serves as a witness to the multitude of sensibilities surrounding European cultural heritage and the discordances therein. Thus, from the perspective of otherness, we study which European artistic manifestations or cultural practices arouse the interest of these Latin American women travellers since they decide to dedicate a few words to them and even express their emotions.

The methodology of the REWIND project integrates different methods and tools from the field of Digital Humanities that consider aspects of cultural heritage related to its spatiotemporal coordinates and its impact on the construction of historical memory. In the framework of the Deep Data concept, we use a structured intelligent dataset to organise the information on cultural heritage in an ontological model that allows detecting patterns or relationships challenging to observe by humans, as well as to apply Distant Reading techniques such as, for example, Sentiment Analysis. From the perspective of literary geography, GIS tools are used to analyse the relationship between space and cultural heritage. This dataset complies with FAIR principles to facilitate the reproducibility of the results and promote the reuse of the data in other research.

We create the dataset by encoding the travel books using XML mark-up language and the standards proposed by the Text Encoding Initiative to extract information on the cultural heritage elements mentioned by female travellers. On the one hand, a modernised diplomatic digital edition of each publication is created so that not only the structure of the books is annotated, but also the original formatting of the text (italics, capital letters, indentation, etc.) and typos are preserved while adapting the Spanish to the latest edition of the *Ortografía de la Lengua Española* (2010). On the other hand, tagging is used to identify people, places, objects, and events, turning each book into a kind of XML database, which can be queried using XQuery expressions. This annotation system makes it possible to differentiate all cultural heritage elements by linking them, whenever possible, both with geographical coordinates, through the GeoNames gazetteer, and with linked open data from Wikidata. In addition, the ROSSIO Thesaurus is used to categorise the different European cultural heritage elements through the @type attribute according to the UNESCO typology that differentiates between intangible cultural heritage and tangible movable and immovable cultural heritage.

Listening to the voice of these non-European women, committed to feminism and socio-cultural diversity, offers a counterbalance to the dominant Eurocentric discourse on cultural heritage. Their storytelling, at times, intertwined with European traditions yet often underscored by a distinct sense of otherness, provides fertile ground for constructing decolonial, inclusive, and plural historical narratives.

Codificación de la literatura de viajes: análisis del patrimonio cultural europeo a través de la perspectiva de escritoras latinoamericanas

Alba Comino

Universidade NOVA de Lisboa (Portugal)

La representación del patrimonio cultural en la literatura de viajes a menudo refleja una selección curada influenciada por diversos factores, desde discursos autorizados moldeados por lecturas previas o visitas guiadas hasta los intereses personales de los autores. Revelamos una narrativa histórica dinámica al realizar un análisis comparativo de los aportes de autoras latinoamericanas acerca del patrimonio cultural europeo a principios del siglo XX. Esta exploración sirve como testimonio de la multitud de apreciaciones que rodean el patrimonio cultural europeo y las discordancias que contiene. Así, desde la perspectiva de la alteridad, estudiamos qué manifestaciones artísticas o prácticas culturales europeas despiertan el interés de estas mujeres viajeras latinoamericanas que deciden dedicarles unas palabras e incluso expresar sus emociones.

La metodología del proyecto REWIND integra diferentes métodos y herramientas del campo de las Humanidades Digitales que consideran aspectos del patrimonio cultural relacionados con sus coordenadas espacio-temporales y su impacto en la construcción de la memoria histórica. En el marco del concepto Deep Data, utilizamos un conjunto de datos inteligente estructurado para organizar la información sobre el patrimonio cultural en un modelo ontológico que permite detectar patrones o relaciones difíciles de observar por el ser humano, así como aplicar técnicas de Lectura a Distancia como, por ejemplo, Análisis de los Sentimientos. Desde la perspectiva de la geografía literaria, se utilizan las herramientas SIG para analizar la relación entre el espacio y el patrimonio cultural. Este conjunto de datos cumple con los principios FAIR para facilitar la reproducibilidad de los resultados y promover la reutilización de los datos en otras investigaciones.

Creamos el conjunto de datos al codificar los libros de viajes para lo que utilizamos el lenguaje de marcado XML y los estándares propuestos por la Text Encoding Initiative a fin de extraer información sobre los elementos del patrimonio cultural mencionados por las viajeras. Por un lado, se crea una edición digital diplomática modernizada de cada publicación donde no sólo se anota la estructura de los libros, sino que también se preserva el formato original del texto (cursiva, mayúsculas, sangría, etc.) y los errores tipográficos adaptando al mismo tiempo el español a la última edición de la Ortografía de la Lengua Española (2010). Por otro lado, se utiliza el etiquetado para identificar personas, lugares, objetos y eventos, que convierte a cada libro en una especie de base de datos XML que se puede consultar mediante expresiones XQuery. Este sistema de anotación permite diferenciar todos los elementos del patrimonio cultural

vinculándolos, siempre que sea posible, tanto con coordenadas geográficas, a través del diccionario geográfico GeoNames, como con datos abiertos vinculados de Wikidata. Asimismo, se utiliza el Tesoro ROSSIO para categorizar los diferentes elementos del patrimonio cultural europeo a través del atributo @type de acuerdo con la tipología de la UNESCO que diferencia entre patrimonio cultural inmaterial y patrimonio cultural tangible, mueble e inmueble.

Escuchar la voz de estas mujeres no europeas, comprometidas con el feminismo y la diversidad sociocultural, ofrece un contrapeso al discurso eurocéntrico dominante sobre el patrimonio cultural. Su narración, a veces entrelazada con tradiciones europeas pero a menudo destacada por un claro sentido de alteridad, proporciona un terreno fértil para construir narrativas históricas decoloniales, inclusivas y plurales.

French Drama in TEI: A Workflow for the Continuous Integration of the "Théâtre classique" Corpus into the DraCor Infrastructure

Frank Fischer¹, Carsten Milling², Jan Jokisch³, Peer Trilcke²

¹Freie Universität Berlin, ²Universität Potsdam, ³Max-Planck-Institut für empirische Ästhetik (Germany)

Paul Fièvre's Théâtre Classique, a growing corpus of French classical theater, has been maintained since 2007 and can be accessed at <http://www.theatre-classique.fr/> (Schöch, 2007). It is a unique source of currently 1.850 French-language plays and available in different formats, including a TEI P4 version. TC is a dynamic corpus: Paul Fièvre keeps adding new plays, corrects errors and also updates the markup. However, the files tend to have several markup problems, including non-valid TEI, there is no version control of changes and it is hard to keep track. Also, there is no simple way to prepare this corpus for research purposes.

To address these challenges and bring the corpus in line with corpora of plays in other languages, we started a project to onboard the corpus to DraCor (Fischer et al., 2019) and keep it up-to-date. The French Drama Corpus (FreDraCor) as we call it, is by far the biggest corpus in the DraCor collection and can be accessed at <https://dracor.org/fre>.

The most significant modifications performed on the original documents include simple things such as adding TEI namespace and XML declaration. The headers were updated with links to authority files like national libraries and Wikidata to provide machine-readable data on authors and works, including premiere and publication dates. We added the particDesc element to collect characters appearing in a play and to give

them an ID and add gender info, which is the basis for many research approaches and visualisations based on structural data.

For building the FreDraCor documents from the "Théâtre Classique" source, a scripted workflow has been set up that processes the original files with an XQuery transformation. To speed up the process, multiple eXist DB instances can be started in parallel using either Podman or Docker.

Some types of corrections cannot be automatised but ultimately rely on manual changes to orthography and markup. For example, about 65% of all stage directions needed to be extracted from headings and speech prefixes via careful examination of suspicious strings. To keep control over the source corpus and our own adaptation and in order not to lose our own enhancements, be they automatic or manual, all changes on FreDraCor are version controlled.

In our presentation, we will detail our workflow and show how it is adaptable for similar projects. We will also demonstrate how DraCor corpora are used in research, i.e. how this collection of thousands of TEI documents in more than a dozen different languages can be used to its full potential.

Teatro francés en TEI: un flujo de trabajo para la integración continua del corpus Théâtre classique en la infraestructura DraCor

Frank Fischer¹, Carsten Milling², Jan Niklas Jokisch³, Peer Trilcke²

¹Universidad Libre de Berlín, ²Universidad de Potsdam, ³Instituto Max Planck para la Estética Empírica (Alemania)

El corpus Théâtre Classique (TC) de Paul Fièvre, un corpus creciente de teatro clásico francés, se mantiene desde 2007 y se puede acceder a él en <http://www.theatre-classique.fr/> (Schöch, 2007). Es una fuente única de 1.850 obras en francés y está disponible en diferentes formatos, incluida una versión TEI P4. TC es un corpus dinámico: Paul Fièvre continúa agregando nuevas obras, corrige errores y también actualiza el marcado. Sin embargo, los archivos tienden a tener varios problemas de marcado, incluido TEI no válido, no hay control de versiones de los cambios y es difícil realizar un seguimiento. Asimismo, no existe una forma sencilla de organizar este corpus con fines de investigación.

Para abordar estos desafíos y alinear el corpus con corpus de obras de teatro en otros idiomas, iniciamos un proyecto para integrar el corpus a DraCor (Fischer et al., 2019) y mantenerlo actualizado. El Corpus de Drama Francés (FreDraCor), como lo llamamos, es el corpus más grande de la colección DraCor y se puede acceder a él en <https://dracor.org/fre>.

Las modificaciones más importantes realizadas en los documentos originales incluyen cuestiones simples como agregar el espacio de nombres de TEI y la declaración XML. Los encabezados se actualizaron con enlaces a archivos de autoridad tales como bibliotecas nacionales y Wikidata, a fin de proporcionar datos legibles por máquina sobre autores y obras, incluidas fechas de estreno y publicación. Agregamos el elemento `particDesc` para recopilar personajes que aparecen en una obra de teatro, y darles una identificación y agregar información de género, que es el principio de muchos enfoques de investigación y visualizaciones basadas en datos estructurales.

Se ha configurado un flujo de trabajo programado que procesa los archivos originales con una transformación XQuery para crear los documentos `FreDraCor` a partir de la fuente "Théâtre Classique". A fin de acelerar el proceso, pueden iniciarse varias instancias de `eXist DB` en paralelo utilizando `Podman` o `Docker`.

Algunos tipos de correcciones no se pueden automatizar sino que, en última instancia, dependen de cambios manuales en ortografía y marcado. Por ejemplo, alrededor del 65% de todas las acotaciones escénicas debían extraerse de títulos y prefijos de discurso mediante un examen minucioso de cadenas sospechosas. Para mantener el control sobre el corpus fuente y nuestra propia adaptación, y para no perder nuestras propias mejoras, ya sean automáticas o manuales, todos los cambios realizados en `FreDraCor` están controlados por versión.

En esta presentación, detallaremos nuestro flujo de trabajo y mostraremos el modo en que puede adaptarse a proyectos similares. También demostraremos cómo se utilizan los corpus `DraCor` en la investigación, es decir, de qué manera puede utilizarse plenamente esta colección de miles de documentos TEI en más de una docena de idiomas diferentes.

La codificación en XML-TEI del teatro español del Siglo de Oro: problemas, soluciones, ortodoxia y heterodoxia

Emanuele Leboffe

Universidad Autónoma de Barcelona (España)

El creciente interés hacia las Humanidades Digitales en la última década ha llevado a la elaboración de muchos proyectos de edición crítica digital, algunos de los cuales se realizaron en el ámbito del teatro español del Siglo de Oro. Dentro de este grupo destacan *La dama boba*, proyecto coordinado por Marco Presotto (2015); las ediciones de la Biblioteca Gondomar Digital (GD; dir. Luigi Giuliani, 2017); las ediciones de la Biblioteca digital PROLOPE (BDP), relacionadas con el grupo de investigación

PROLOPE; La discreta enamorada de Gemma Burgos (2019) y La estrella de Sevilla de Nadia Revenga (2021), dos tesis doctorales surgidas en la Universitat de València. Cada uno ha significado una contribución inestimable para el desarrollo de este campo, permitiendo evidenciar tanto las potencialidades como las dificultades que se presentan al elaborar un producto digital de este tipo y al codificar textos de teatro áureo. A tal respecto, cabe decir que el teatro español del Siglo de Oro implica desafíos importantes en la codificación XML-TEI debido a que combina la necesidad de marcado de versos y estrofas con los distintos parlamentos de los personajes, lo que favorece el fenómeno de *overlapping*, si se siguen rigurosamente las directrices de TEI. Además, analizando el marcado de los proyectos mencionados, se puede constatar que estos abordan la codificación de algunos fenómenos textuales (estrofas, versos partidos, diferentes variantes y errores básicos tipificados en los manuales de crítica textual) de manera distinta. En unos casos, de hecho, se sigue una codificación más ortodoxa respecto a las directrices de TEI. En otros, sin embargo, ellos mismos confiesan acudir a soluciones de compromiso (también condicionados por las visualizaciones que se persiguen), pero en la medida en que esas marcaciones son heterodoxas y diferentes en cada proyecto, se renuncia a un marcado uniforme y realmente interoperable.

A continuación, se ofrecen algunos ejemplos sacados directamente de los proyectos en cuestión:

- Codificación de estrofas:
 - Los donaires de Matico (GD): Se usa <lg> para el marcado de estrofas (correcto, pero multiplicando así las posibilidades de overlapping).
 - Los ramilletes de Madrid (BDP): Se usa <milestone/> (soluciona en parte la cuestión overlapping; aunque no aparezca como ejemplo en las directrices TEI, se ofrecen soluciones similares en casos de conflictos entre elementos <speaker> y <lg>).
- Omisiones:
 - La dama boba: Omisiones de versos en un testimonio específico codificadas con <rdg/> (solución ortodoxa).
 - La discreta enamorada: Se pone una nota en el elemento <rdg> para explicar, por ejemplo, que una acotación se ha omitido (sin marcarlo, como se debe, como elemento vacío).

Mi intención, por tanto, es aportar un catálogo de la casuística, discutiendo las soluciones propuestas en los distintos proyectos y dando mi perspectiva sobre su efectividad. Al mismo tiempo, cabría estudiar si realmente las soluciones heterodoxas son las únicas posibles o si hay alternativas ortodoxas preferibles para una buena

interoperabilidad. Si no fuera así, sería deseable alcanzar soluciones consensuadas que se pudieran perseguir en proyectos distintos o realizar propuestas al Consorcio TEI para la resolución de estos problemas.

THE XML-TEI encoding of Spanish theatre of the Golden Age: problems, solutions, orthodoxy and heterodoxy

Emanuele Leboffe

Autonomous University of Barcelona (Spain)

The growing interest in Digital Humanities in the last decade has led to the development of several projects on digital critical edition, some of which were carried out in the field of Spanish theatre of the Golden Age. Within this group, *La dama boba*, a project coordinated by Marco Presotto (2015); *Gondomar's Digital Library editions (GD)*; dir. Luigi Giuliani, (2017); *PROLOPE's Digital Library editions (BDP)*, related to the research group PROLOPE; *La discreta enamorada* by Gemma Burgos (2019) and *La estrella de Sevilla*, by Nadia Revenga (2021), two doctoral theses that emerged at the University of Valencia, stand out. Their invaluable contributions to the development of this field has made it possible for us to show both the potentialities and difficulties that arise when developing a digital product of this type and when encoding golden theatre texts. In this regard, it should be said that the Spanish theatre of the Golden Age entails important challenges when facing XML-TEI coding since it combines the need to mark verses and stanzas with the different characters speeches, which facilitates the "overlapping" phenomenon if the TEI guidelines are rigorously followed. Furthermore, the different approach adopted to the codification of some textual phenomena (stanzas, split verses, different variants and basic errors typified in textual criticism manuals) can be seen by analysing the marking of the mentioned projects. As a matter of fact, in some cases a more orthodox encoding is followed concerning the TEI guidelines. In others, however, they confess to resorting to compromise solutions (also conditioned by the visualizations that are pursued), but giving up a uniform and truly interoperable marking as these markings are heterodox and different in each project.

Some examples excerpted from the projects involved are listed below:

- Encoding stanzas:
 - *Los donaires de Matico (GD)*: <lg> is used to mark stanzas (correct, but increasing the possibilities of overlapping as a result).

- Los ramilletes de Madrid (BDP): <milestone/> is used (It partially solves the overlapping issue. Although it is not listed as an example in the TEI guidelines, similar solutions are offered in cases of conflicts between <speaker> and <lg> elements).
- Omissions:
 - La dama boba: Verse omissions in specific narrations encoded with <rdg/> (orthodox solution).
 - La discreta enamorada: A note is placed in the <rdg> element to explain, for example, that an annotation has been omitted (not marking it as an empty element as it should be done).

My intention, therefore, is to provide a catalogue of the cases, to discuss the solutions proposed in the different projects and to give my views on their effectiveness. At the same time, consideration should also be given as to whether heterodox solutions are really the only possibility or if there are preferable orthodox alternatives for good interoperability. If this were not the case, it would be useful to reach consensus-based solutions that could be pursued in different projects, or to make proposals to the TEI Consortium in order to solve these problems.

Hacia el nuevo paradigma de edición filológica digital: contratiempos, prejuicios y el valor de la labor digital

Susanna Allés-Torrent

Universidad de Miami (Estados Unidos de América)

Esta presentación aspira a compartir una serie de reflexiones en torno a la creación de ediciones filológicas digitales y de las diferentes etapas que esta conlleva: desde el modelado de datos, basado necesariamente en un sólido conocimiento del contenido y la tradición textual de la fuente primaria, pasando por el establecimiento de unas guías tanto de edición como de codificación textual, a un sistema de colaboración y de flujo de trabajo colaborativo, hasta el diseño, la publicación web y la preservación digital. Concretamente, las experiencias nacen del proyecto Pronapoli, <https://pronapoli.com/>, cuyos objetivos principales son, por un lado, estudiar la estancia italiana del poeta Garcilaso de la Vega, llevada a cabo en Nápoles a intervalos, desde el verano de 1529 hasta la primavera de 1536, y por el otro, publicar una nueva edición crítica digital de su obra poética. Esfuerzos significativos se han ya dedicado a la creación de una versión beta de un primer grupo de textos, disponibles en <https://pronapoli.com/ediciondigital/>. En una próxima ronda de financiación

(2024-2028), se prevé que el mismo modelo se extienda al resto de textos. Se ofrecerá una visión de conjunto de la edición crítica digital de la obra garcilasiana, así como del proceso ecdótico, el flujo digital filológico, y el resultado final en línea. Interesa especialmente compartir nuestra práctica digital que ha intentado en todo momento correr en paralelo a algunas de las buenas prácticas difundidas en el terreno de las humanidades digitales en general y en el de la edición digital en particular: la codificación en XML-TEI y la consiguiente redacción de una documentación técnica para homogeneizar el trabajo en equipo, una transformación a partir del lenguaje XSLT, así como una estructura mínima; todo ello disponible en acceso abierto en un repositorio en línea. Algunos temas clave a tratar, y estrechamente relacionados con el proceso ecdótico, conciernen, en primer lugar, preconcepciones profundamente arraigadas sobre qué es y cómo debe presentarse una edición crítica, fruto sin duda de una larga tradición secular –especialmente en ámbito hispánico peninsular y en el terreno de la edición de textos medievales y del Siglo de Oro–. Preconcepciones y prejuicios que, en algunos casos, son incluso difíciles de percibir pero que entorpecen el camino hacia el establecimiento de un nuevo paradigma digital. En segundo lugar, se problematiza la búsqueda del equilibrio entre el editor filológico (el que establece el texto crítico) y el que lleva a cabo la codificación, dos roles que raras veces se aglutinan en una sola persona. La labor del primero es entendida, mientras que la del segundo permanece todavía en un limbo, las más de las veces relegada a una actividad “técnica”, “mecánica”. Nada más lejos de la verdad: la codificación, y en general cualquier labor digital (p.ej. creación de datasets), es una actividad intelectual y profundamente hermenéutica. En fin, otras percepciones sobre el supuesto carácter efímero, superficial o perecedero de los contenidos en la red –contrapuestos a las publicaciones en papel, por ejemplo– desafían la solidez y la autoridad de las ediciones críticas digitales.

Towards the new paradigm of digital philological editing: setbacks, prejudices and the value of digital work

Susanna Allés-Torrent

University of Miami (United States of America)

This presentation aims to share a number of reflections on the creation of digital philological editions and the different stages entailed: from data modelling, necessarily based on solid knowledge of the content and textual tradition of the primary source, setting guidelines for both editing and text encoding, to a collaboration system and collaborative workflow, design, web publishing and digital preservation. Specifically, the two main objectives of experiences gained in the Pronapoli project (see <https://pronapoli.com/>) are: firstly, to delve into Garcilaso de la Vega’s stay in Naples,

from the summer of 1529 until the spring of 1536, and, secondly, to release a novel digital critical edition of his poetic works. Significant efforts have already been made towards the creation of a beta version of a set of texts available at <https://pronapoli.com/ediciondigital/>. In a new funding stage (2024-2028), the same model is expected to be applied to the rest of the texts. An overview of the digital critical edition of Garcilaso's work as well as the ecdotic process, the philological digital flow and the final online result will be offered. It is particularly interesting to share our digital practice, which has always intended to run parallel with good practices disseminated in the field of digital humanities in general and in that of digital editions in particular: an XML-TEI encoding and drafting of technical documentation to standardise teamwork, a transformation based on the XSLT language, as well as a minimum structure, available in an online open-access repository. First of all, we will discuss some key issues closely related to the ecdotic process concerning deeply-rooted preconceptions about what a critical edition is and how it should be presented, which are the result a long secular tradition, especially in peninsular Spain and in the field of editing medieval and Golden Age texts. Preconceptions and prejudices that, in some cases, are even difficult to perceive but that hinder the path towards establishing a new digital paradigm. Secondly, reaching a balance between the philological editor (the one who establishes the critical text) and the one who encodes is hard as these tasks are rarely performed by a single person. While the work done by the former is understood, that of the latter still remains in limbo and is usually considered a “technical”, “mechanical” activity. Nothing could not be further from the truth: coding, and any digital work in general (e.g. creating datasets), is an intellectual and deeply hermeneutic activity. In short, other perceptions regarding the nature of online content as ephemeral, superficial or perishable, as contrasted with a paper publication, for example, challenge the authority of digital critical editions.

Keynote Conferences / Conferencias plenarias

¿TEI, o no TEI?, esa es la cuestión

José Manuel Fradejas Rueda

Universidad de Valladolid (España)

Desde hace años, desde la popularización de las Humanidades Digitales, el sistema de codificación TEI se ha convertido en “el” estándar de todos aquellos que quieren producir una edición filológica digital (SDE). La TEI no solo ha sido acogida por quienes

hacen ediciones; otros módulos, como el de corpus y diccionarios, tienen un amplísimo uso.

En esta conferencia hablaré de otro sistema de codificación que tuvo una gran difusión en el ámbito del hispanomedievalismo, y de cómo fue reemplazado por la TEI, aunque produjo interesantes resultados que antecedieron a los actuales modelos de edición digital. Creíamos periclitado este sistema, pero en algunos casos hemos tenido que volverlo a retomar puesto que se ha mostrado mucho más flexible y sencillo de utilizar cuando en un proyecto SDE intervienen plataformas HTR y un sistema de etiquetado morfológico y lematización (PoS tagging) automática.

Debido a la simplicidad del sistema, la conversión entre uno y otro sistema es posible con mínimo esfuerzo y la casi nula pérdida de información. Por lo tanto, la respuesta a la duda que se plantea en el título es: ¡sí, TEI sí!

To TEI or not to TEI, that is the question

José Manuel Fradejas Rueda

University of Valladolid (Spain)

For years now, particularly since the popularization of Digital Humanities, TEI has become “the” standard form of encoding for all those willing to create a scholarly digital edition (SDE). Not only those who make editions have embraced the TEI encoding; in fact, other modules, such as corpus and dictionaries, are being extensively used.

In this lecture, I will talk about another encoding system widely disseminated in the field of Hispano-medievalism and how it was replaced by the TEI encoding, despite having produced interesting results which preceded the current models of digital publishing. Although we thought this system was outdated, there were cases in which we needed to take it up again as it proved to be much more flexible and easier to use when HTR platforms, a morphological tagging system and automatic lemmatization (PoS tagging) are involved in an SDE project.

Because of its simplicity, it is possible to convert from one system to another with minimal effort and almost no loss of information.

Consequently, the answer to the question raised in the title is: Say yes to TEI!

Short Papers / Presentaciones breves

Our History is Missing: Digital Sustainability to Preserve the Legacy of Canadian Lesbian Activism

Constance Crompton

University of Ottawa (Canada)

This short paper introduces sustainability plans for the Lesbian and Gay Liberation in Canada project, as we transition from a Neo4j database and node.js web app to an Endings Project compliant site. In our short presentation, we will introduce our 2SLGBTQ+ digital history project as a case study for why infrastructural and community support is so critical regarding transmission of 2SLGBTQ+ Canadian history to future generations. In undertaking this move we have drawn on concerns arising from the accessibility of material related to the movement we study. We ask how to best represent the politics of liberation for members of the 2SLGBTQ+ community in Canada based on the intersectional principles articulated by Tremblay and Podmore (2015). The Lesbian and Gay Liberation in Canada (lglc.ca) project has long focused on men and English speakers, mainly because the material representing them is more readily archivable. How can we ensure that the material we produce is readily archivable for the future, too? A TEI based project, we would be glad to introduce our move away from a 34,000-record database. The publicly accessible web app that sits atop the database is starting to show its age, leading us to ask: Should we continue to migrate the database and the project web app forward indefinitely, or should we move to a static site model, with iterative releases, to help ensure that the site can live on without constant technical upgrades? We will introduce our plans to apply principles for sustainability and longevity to help keep the stories of 2SLGBTQ+ liberation on the web.

Nos falta nuestra historia: sostenibilidad digital para preservar el legado del activismo lésbico canadiense

Constance Crompton

Universidad de Ottawa (Canadá)

Este breve artículo presenta los planes de sostenibilidad para el proyecto de Liberación de Lesbianas y Gays en Canadá, mientras realizamos la transición de una base de datos Neo4j y una aplicación web node.js a un sitio compatible con el Proyecto

Endings. En nuestra breve presentación, expondremos nuestro proyecto de historia digital 2SLGBTQ+ como un estudio de caso del motivo por el cual el apoyo infraestructural y comunitario es tan crítico respecto de la transmisión de la historia canadiense 2SLGBTQ+ a las generaciones futuras.

Para llevar a cabo nuestra tarea, hemos recurrido a las preocupaciones que surgen de la accesibilidad del material relacionado con el movimiento que estudiamos. Nos preguntamos cómo representar, de la mejor manera, las políticas de liberación para los miembros de la comunidad 2SLGBTQ+ en Canadá basados en los principios de interseccionalidad articulados por Tremblay y Podmore (2015). El proyecto Liberación de Lesbianas y Gays en Canadá (lgc.ca) se ha centrado durante mucho tiempo en hombres y angloparlantes, principalmente, porque el material que los representa es más fácil de archivar.

¿Cómo podemos garantizar que el material que producimos también pueda archiversse fácilmente para el futuro? Un proyecto basado en TEI. Nos complace proponer dejar atrás una base de datos de 34.000 registros. La aplicación web de acceso público que se encuentra encima de la base de datos está quedando anticuada y ello nos lleva a preguntarnos: ¿deberíamos continuar migrando la base de datos y la aplicación web del proyecto indefinidamente, o deberíamos pasar a un modelo de sitio estático con lanzamientos iterativos, para garantizar que el sitio pueda funcionar sin actualizaciones técnicas constantes? Presentaremos nuestros planes de aplicación de los principios de sostenibilidad y longevidad para ayudar a mantener las historias de la liberación 2SLGBTQ+ en la web.

Interlibrary Collaborations: Forging Antiracist, Decolonial, and Inclusive Markup Interventions in Partnership

Joey Takeda¹, Sydney Lines², Ekatarina Grgurić², Rebecca Dowson²

¹Simon Fraser University, ²University of British Columbia (Canada)

Collaboration and community have long been central to the TEI-C's mission and, as many people and projects across institutions and libraries have demonstrated, a powerful way for rethinking texts, encoding, and markup practices (e.g. Warwick, 2012; Flanders, 2012; Flanders & Hamlin, 2013; Green, 2014; Lu & Pollock, 2019). Yet while inter-institutional partnerships offer a promising model for building both infrastructures and capacity for collaborative text encoding projects, there remains, as Bonn et al. (2021) note, "much work to do" in developing best practices, frameworks, and working methods in support of such inter-institutional collaborations.

This paper describes some of the challenges and opportunities of inter-institutional partnership that have arisen through the emergent partnership between the Digital Humanities Innovation Lab at Simon Fraser University and the Digital Scholarship in Arts (DiSA) initiative at University of British Columbia, initially developed in support of UBC's Adaptive TEI Network (ATN). Co-led by doctoral students and faculty, the ATN unites several TEI projects to implement antiracist, decolonial, and inclusive encoding practices as well as challenge the stigma of multi-authorship and collaboration that persists within much humanities scholarship. The ATN has also served as the pilot for our cross-institutional collaboration, allowing us to bridge the two institutions to share resources, expertise, and infrastructure.

In this short paper, we will describe how the need for anti-racist markup strategies enabled this partnership as well as discuss the administrative complexity –and necessity– in structuring embedded support for TEI projects across institutions.

Colaboraciones interbibliotecas: forjando intervenciones de marcado antirracistas, decoloniales e inclusivas

Joey Takeda¹, Sydney Lines², Ekatarina Grgurić², Rebecca Dowson²

¹Universidad Simon Fraser, ²Universidad de Columbia Británica (Canadá)

Desde hace mucho tiempo, la colaboración y la comunidad han sido fundamentales en la misión del Consorcio TEI. Tal como lo han demostrado un gran número de personas y proyectos llevados a cabo en instituciones y bibliotecas, representan una forma poderosa de repensar los textos, la codificación y las prácticas de marcado (por ejemplo, Warwick, 2012; Flanders, 2012; Flanders & Hamlin, 2013; Verde, 2014; Lu y Pollock, 2019). Sin embargo, si bien las asociaciones interinstitucionales ofrecen un modelo prometedor a fin de construir infraestructuras y planificar capacidades para proyectos colaborativos de codificación de texto, aún queda, como señalan Bonn et al. (2021), “mucho trabajo por hacer” en el desarrollo de mejores prácticas, marcos y métodos de trabajo en apoyo de dichas colaboraciones interinstitucionales.

Este artículo reseña algunos de los desafíos y las oportunidades de la asociación interinstitucional que han surgido de la asociación emergente entre el Laboratorio de Innovación en Humanidades Digitales de la Universidad Simon Fraser y la iniciativa Digital Scholarship in Arts (DiSA) de la Universidad de Columbia Británica (UBC), desarrollada inicialmente en apoyo de la Adaptive TEI Network (ATN) de la UBC. Codirigida por estudiantes de doctorado y profesores, la ATN reúne varios proyectos TEI para implementar prácticas de codificación antirracistas, decoloniales e inclusivas, así como para desafiar el estigma de la autoría múltiple y la colaboración que persiste en gran parte de los estudios de humanidades. Asimismo, la ATN ha servido como

piloto de nuestra colaboración interinstitucional, permitiéndonos unir a las dos instituciones para compartir recursos, experiencia e infraestructura.

En este breve artículo, describiremos cómo la necesidad de estrategias de mercado antirracistas ha permitido esta asociación y también discutiremos la complejidad administrativa (y la necesidad) de estructurar el apoyo integrado para los proyectos TEI en todas las instituciones.

Using FairCopy Editor to Encode Blackness and Indigeneity in Sor Juana's Villancicos Negros

Alondra Ceballos

Texas A&M University (United States of America)

The ten minute tool demonstration will utilize FairCopy Editor to encode the voices of Afro Mexicans and the Nahuas people of Central Mexico that are depicted in Sor Juana Inés de la Cruz's late seventeenth century villancicos negros. Specifically, I aim to demonstrate how TEI markup language can be used to capture the remixes of Black and indigenous voices in Sor Juana Inés de la Cruz's Villancico 224 (1676): A la aclamación festiva.

Uso del editor FairCopy para codificar la negritud y la indigeneidad en los Villancicos Negros de Sor Juana

Alondra Ceballos

Universidad de Texas A&M (Estados Unidos de América)

La demostración llevará diez minutos y se utilizará FairCopy para codificar las voces de los afromexicanos y los nahuas del centro de México representados en los villancicos negros de Sor Juana Inés de la Cruz de finales del siglo XVII. Específicamente, el objetivo es demostrar de qué modo se puede utilizar el lenguaje de marcado TEI para capturar las remezclas de voces negras e indígenas en el Villancico 224 (1676): "A la aclamación festiva" de Sor Juana Inés de la Cruz.

Edición digital e hipertexto: Plataforma multi editorial y multilingüe para textos académicos y literarios

Julieta Y. Goldsmidt

Universidad de Buenos Aires (República Argentina)

La proliferación de contenidos digitales habilita diversos modos de concebir un texto escrito. El presente trabajo reside en la exposición de una plataforma de lectura hipertextual que permite en tiempo real visualizar distintas ediciones y traducciones provenientes de un mismo texto origen. Las ediciones digitales realizadas contienen metadatos y etiquetas compatibles con los lineamientos de la Text Encoding Initiative (TEI), basado en el metalenguaje XML. Así, esta herramienta permite generar comparaciones y análisis de contenidos multi editoriales o multilingües en una misma pantalla de manera personalizada. Dada la visualización flexible de este sistema, es posible exponer intra e intertextualidades que enriquezcan el análisis de escritos académicos y literarios. No obstante, los recursos mencionados conducen a reflexionar sobre la tensión entre lectura, edición y autoría. A su vez, se problematizan los límites y alcances de la edición digital mínima. Resulta de interés indagar la viabilidad de una edición digital híbrida que mantenga las ventajas de la minimal computing e incorpore recursos característicos de los sitios web dinámicos. Por consiguiente, la propuesta implica un recorrido práctico, técnico y teórico que articula los distintos avatares del hipertexto. Considerando que la escritura se asienta en la cultura posmoderna como una multiplicidad, la plataforma y el debate propuestos pretenden enfatizar un destino posible para el hipertexto en el que su lectoescritura no permanezca infinitizada en contenidos fragmentarios, sino bajo una lógica que habilite la lectura crítica, rigurosa y contextualizada.

Digital publishing and hypertext: Multi-editorial and multilingual platform for academic and literary texts

Julieta Y. Goldsmidt

University of Buenos Aires (Argentine Republic)

The proliferation of digital content enables different ways of conceiving a written text. Our work displays a hypertextual reading platform that makes it possible to view different editions and translations from the same source text in real time. The digital editions contain metadata and tags compatible with the Text Encoding Initiative Guidelines (TEI Guidelines), based on the XML metalanguage. Thus, this tool allows users to make comparisons and perform analysis of multi-editorial or multilingual content on the same screen in a personalized way. Given the flexible visualization of this system, it is possible to expose intra- and intertextualities that enrich the analysis of academic and literary writings. However, the mentioned resources lead to reflection on

the tension between reading, editing and authorship. At the same time, the limits and scope of minimal digital editing are problematized. It is interesting to report the viability of a hybrid digital edition that maintains the advantages of minimal computing and incorporates characteristic resources from dynamic websites. Consequently, the proposal implies a practical, technical and theoretical journey that articulates the different problems of the hypertext. Considering that writing is viewed as multiple in postmodern culture, our aim is to discuss a possible destiny for the hypertext where its reading and writing are not seen as infinite fragments, but where a critical, rigorous and contextualized reading is enabled.

The “Foreign” Element in Yiddish

Jonah Lubin, Frank Fischer

Free University of Berlin (Germany)

Despite its interesting and problematic theoretical nature, little attention has been given to the “foreign” element in TEI. An exception is the work of Barbaric and Halonja, who investigated this issue in terms of Croatian. They claim that the “foreign” element should be applied to instances of code switching (Barbaric, 2012). Distinguishing between code switching and borrowing is a contentious issue among linguists (Mahootian, 2006). The subtleties of these distinctions are not represented by the “foreign” element, and solutions, like Schlitz’s “borrowing” element, have not been incorporated into the standard (“TEI element foreign (foreign)”, Schlitz, 2009).

This investigation proceeds from observations made during the encoding of plays for DraCor’s new corpus of Yiddish drama (YiDraCor), and questions raised over the use of the “foreign” element in the context of Yiddish. Yiddish is a mixed language which has usually been spoken in multilingual contexts (Weinreich, 1955). It is thus difficult to distinguish between the mixedness present within Yiddish, and borrowings from without.

In the interest of coding, we have adopted a model of “foreign” element use that is primarily morphological, secondarily lexical. If a form evinces productive morphological features, then it should be tagged as “foreign.” Failing that, following Stutchkoff and Weinreich, we tag as “foreign” those words and phrases which had entered local Yiddish usage, but had not yet entered Yiddish standards, and which would not likely be understood by those not familiar with their language of origin (Stutchkoff & Weinreich, 1951). Generally, we contend that those terms which should be marked as “foreign,” are those that evince a certain degree of multilingualism (Grimstad, 2017). Ultimately, we find that the “foreign” tag is best modeled pragmatically: since there is no inherently

foreign component to any language, its use should follow the stated goals of the encoders.

El elemento “extranjero” en yiddish

Jonah Lubin, Frank Fischer

Universidad Libre de Berlín (Alemania)

A pesar de su interesante y problemática naturaleza teórica, se ha prestado poca atención al elemento “extranjero” en la TEI. Una excepción a ello es el trabajo realizado por Barbaric y Halonja, quienes investigaron esta cuestión en términos croatas. Afirman que el elemento "extranjero" debería aplicarse a casos de cambio de código (Barbaric, 2012). Distinguir entre cambio de código y préstamo es un tema polémico entre los lingüistas (Mahootian, 2006). Las sutilezas de estas distinciones no están representadas por el elemento "extranjero", y las soluciones, tales como el elemento "préstamo" de Schlitz, no han sido incorporadas a la norma ("elemento TEI extranjero", Schlitz, 2009).

Esta investigación procede de observaciones realizadas durante la codificación de obras de teatro para el nuevo corpus de drama yiddish de DraCor (YiDraCor), y de las preguntas planteadas respecto del uso del elemento "extranjero" en el contexto del yiddish. El yiddish es una lengua mixta que, en general, se ha hablado en contextos multilingües (Weinreich, 1955). Por tanto, es difícil distinguir entre la mezcla presente dentro del yiddish y los préstamos externos.

En aras de la codificación, hemos adoptado un modelo de uso de elemento "extranjero" que es principalmente morfológico y, de manera secundaria, léxico. Si una forma evidencia características morfológicas productivas, entonces debería etiquetarse como "extranjera". En su defecto, siguiendo a Stutchkoff y Weinreich, etiquetamos como "extranjeras" aquellas palabras y frases que han entrado en el uso local del yiddish pero que aún no han entrado en los estándares del yiddish, y que, probablemente, no serían entendidas por aquellos que no están familiarizados con su idioma de origen (Stutchkoff y Weinreich, 1951). Generalmente, sostenemos que aquellos términos que deben marcarse como "extranjeros" son aquellos que evidencian un cierto grado de multilingüismo (Grimstad, 2017). En última instancia, encontramos que la etiqueta "extranjera" se modela mejor de manera pragmática: dado que ningún idioma tiene un componente inherentemente extranjero, su uso debe seguir los objetivos declarados por los codificadores.

Japanese Texts and TEI: A Gap Analysis

Yifan Wang^{1,4}, Kiyonori Nagasaki^{1,2}, Masahiro Shimoda³

¹International Institute for Digital Humanities, ²Keio University, ³Musashino University,

⁴National Institute for Japanese Language and Linguistics (Japan)

Japan is among regions where one of the most complex writing system as well as practice - which include typography, text structure, document formats, and the combination of them - have been utilized on a daily basis throughout history. The Japanese manuscript and printing tradition does not only inherit the traditional Chinese typographical conventions, but also adds its own set of expansions, via ad hoc annotations or modifications, that brought forth subsystems visually and structurally distinct as well as diverse, and often incompatible. Imaginably, there are still a number of those idiomatic concepts do not transfer well into TEI, due to the limitation either in its vocabulary or in the environment surrounding it. We give a brief overview of the current progress and challenges in transcribing premodern Japanese textual elements fully digitally, mostly learned during our TEI encoding project of a large-scale text corpus of a Buddhist canon. As concrete examples, we will address the following topics: portability of characters uncoded and coded in legacy character sets, especially on the limitation of the current gajji module; support of universal structures found in East Asian documents by guidelines, such as "interlinear annotations" (rubi, bōchū), "subline annotations" (warichū), "closing titles", etc; shortage of available vocabulary for metadata in various ISO and other international standards, most conspicuously with regard to the interaction with written variants of Japanese, which often indicates the genre of the document; community building towards standardization and long-term preservability of digital texts, with non-Anglophone user base in mind; and so on.

Textos japoneses y TEI: un análisis de brechas

Yifan Wang^{1,4}, Kiyonori Nagasaki^{1,2}, Masahiro Shimoda³

¹Instituto Internacional de Humanidades Digitales, ²Universidad de Keio, ³Universidad Musashino, ⁴Instituto Nacional de Lengua y Lingüística Japonesas (Japón)

Japón es una de las regiones donde se ha utilizado diariamente, a lo largo de la historia, uno de los sistemas y prácticas de escritura más complejos, que incluye tipografía, estructura de texto, formatos de documentos y la combinación de ellos. La tradición japonesa de manuscritos e impresión no solo hereda las convenciones tipográficas tradicionales chinas, sino que agrega su propio conjunto de expansiones, mediante anotaciones o modificaciones ad hoc, que produjeron subsistemas visual y

estructuralmente distintos, así como también diversos y, a menudo, incompatibles. Es posible que aún varios de esos conceptos idiomáticos no posean una transferencia adecuada en la TEI debido a la limitación de su vocabulario o del entorno que lo rodea. Ofrecemos una breve descripción general del progreso y los desafíos actuales en la transcripción totalmente digital de elementos textuales del japonés premoderno, aprendidos, en su mayoría, durante nuestro proyecto de codificación TEI de un corpus de texto a gran escala de un canon budista. A modo de brindar ejemplos concretos, abordaremos los siguientes temas: portabilidad de caracteres no codificados y codificados en conjuntos de caracteres heredados, especialmente en la limitación del módulo gaiji actual; apoyo de estructuras universales que se encuentran en documentos de Asia oriental mediante directrices tales como "anotaciones interlineales" (rubi, bōchū), "anotaciones de sublíneas" (warichū), "títulos de cierre", etc., escasez de vocabulario disponible para metadatos en diversos estándares ISO y otros estándares internacionales, más notoriamente en lo que respecta a la interacción con variantes escritas del japonés, que, a menudo, indica el género del documento; la construcción de comunidades hacia la estandarización y la preservación a largo plazo de los textos digitales, teniendo en cuenta a la base de usuarios no anglófonos; etcétera.

Neolatina Sarmatica - from Web 1.0 to Web 3.0

Iwona Grabska-Gradzińska, Grażyna Urban-Godziek

Jagiellonian University (Poland)

The original resource and reason for the creation of the *Neolatina Sarmatica* project is Cochanovius Latinus, a collection of completed Latin works by Jan Kochanowski, published electronically (2006-2011) with the first contemporary translation and commentary. This attempt to present Kochanowski's Latin in digital form was based on static HTML page structures and the commentary reproduced the approach of the print editions with its structure and visualisation.

As new tools were developed and became more widespread, a new edition based on the TEI-publisher engine was proposed in 2024, with the aim of enriching it with metadata and consequently increasing the reader's ability to search and interact with the main text and supplementary texts (critical apparatus, commentaries), supporting texts, comparisons with other texts, and comparisons of different translations. This approach makes the texts more accessible to readers with different philological competences: scholars who are well versed in the original language and the historical and cultural context of the works, as well as less qualified readers.

The new opening of the texts to different types of readers has made it possible to update the commentaries and adapt them to the requirements of the modern reader, as well as to use the possibilities of varying both the scope and the way in which the texts are visualised.

Work is currently in progress to utilise the metadata collected for editing in an ontology being developed, enabling a semantic approach to the text as data.

Neolatina Sarmatica: de la Web 1.0 a la Web 3.0

Iwona Grabska-Gradzińska, Grażyna Urban-Godziek

Universidad Jagellónica (Polonia)

Tanto el recurso inicial como el propósito de la creación del proyecto Neolatina Sarmatica es Cochanius Latinus, una colección de obras latinas completas de Jan Kochanowski, publicada electrónicamente (2006-2011) con la primera traducción y comentario contemporáneo. Este intento de presentar el latín de Kochanowski en formato digital se basó en estructuras de páginas HTML estáticas y el comentario reprodujo el enfoque de las ediciones impresas con su estructura y visualización.

A medida que se desarrollaron y difundieron nuevas herramientas, en 2024 se propuso una nueva edición basada en el motor de TEI-publisher, con el objetivo de enriquecerlo con metadatos y, en consecuencia, aumentar la capacidad del lector para buscar e interactuar con el texto principal y los textos complementarios (aparatos críticos, comentarios), textos de apoyo, comparaciones con otros textos y comparaciones de diferentes traducciones. Este enfoque hace que los textos sean más accesibles a lectores con diferentes competencias filológicas: tanto académicos que conocen bien el idioma original y el contexto histórico y cultural de las obras, como lectores menos calificados.

La nueva apertura de los textos a diferentes tipos de lectores ha permitido actualizar los comentarios y adaptarlos a las exigencias del lector moderno, así como aprovechar las posibilidades de modificar tanto el alcance como la forma en que se visualizan los textos.

Actualmente, se está trabajando para utilizar los metadatos recopilados a fin de editarlos en una ontología, hoy en desarrollo, que permitirá un abordaje semántico de los textos en tanto datos.

Edición crítico-genética digital de Frutos de mi tierra (1896) de Tomás Carrasquilla

Juan Esteban Hincapié Atehortúa

Universidad de Antioquia (Colombia)

Se trata de la presentación del desarrollo metodológico para la elaboración de la edición crítico-genética digital de *Frutos de mi tierra* (1896) de Tomás Carrasquilla, basada en la conjugación del paradigma ecdótico digital de Susanna Allés Torrent (2020):

- a. Captura: datos informáticos.
- b. Remediación: el proceso a través del cual los nuevos medios transforman los anteriores.
- c. Estructuración y modelado: armazón interno y externo plasmado en un esquema XML-TEI.
- d. Publicación e interfaces.
- e. Archivo y fluidez textual: elementos paratextuales.

Y las tres fases de la ecdótica colombiana de Edwin Carvajal Córdoba (2017) para realizar ediciones críticas impresas que provienen de la Filología hispánica de Alberto Blecua (1983) y Miguel Ángel Pérez Priego (2011):

Recensio

1. Recuento de la búsqueda de los testimonios.
2. Relación de los testimonios.
3. Descripción bibliográfica.
4. Selección del texto base.

Constitutio textus

5. Cotejo y valoración filológica.

Dispositio textus

6. Formulación del aparato crítico.
7. Criterios editoriales.
8. Fijación.

El propósito es plantear un flujo de trabajo para producir la edición crítico-genética digital de *Frutos de mi tierra* (1896) de Tomás Carrasquilla mediante la siguiente estructura:

Recensio

- a. Síntesis de las historias de transmisión.
- b. Descripción material de los testimonios.
- c. Selección del texto base.
- d. Captura.
- e. Remediación.
- f. Formulación del juicio crítico o *iudicium* del editor.

Constitutio textus

- a. Cotejo.
- b. Valoración filológica de los resultados.
- c. TEI: estructuración y modelado.

Dispositio textus

- a. Composición del aparato crítico.
- b. Criterios editoriales.
- c. Fijación.

Digital critical-genetic edition of *Frutos de mi tierra* (1896) by Tomás Carrasquilla

Juan Esteban Hincapié Atehortúa

University of Antioquia (Colombia)

This presentation describes the methodology followed in the preparation of the digital critical and genetic edition of *Frutos de mi tierra* (1896) by Tomás Carrasquilla, based on the digital ecdotic paradigm of Susanna Allés Torrent (2020):

- a. Capture: computer data.
- b. Remediation: the process through which new media transforms old media.
- C. Structuring and modelling: internal and external framework captured in an XML-TEI schema.

- d. Publishing and interfaces
- e. Archive and textual fluidity: paratextual elements.

And the three phases of Colombian ecdotics by Edwin Carvajal Córdoba (2017) to make printed critical editions that come from the Hispanic Philology of Alberto Blecua (1983) and Miguel Ángel Pérez Priego (2011):

Recensio

- 1. Account of the search for testimonies.
- 2. List of testimonies.
- 3. Bibliographic description.
- 4. Selection of the base text.

Constitutio textus

- 5. Comparison and philological assessment.

Dispositio textus

- 6. Critical apparatus.
- 7. Editorial criteria.
- 8. Fixation.

The aim is to propose a workflow to produce the digital critical and genetic edition of *Frutos de mi tierra* (1896) by Tomás Carrasquilla using the following structure:

Recensio

- a. Synthesis of story transmission.
- b. Material description of the testimonies.
- c. Base text selection.
- d. Capture.
- e. Remediation.
- f. Critical judgment (*iudicium*).

Constitutio textus

- a. Comparison.

- b. Philological assessment of results.
- c. TEI: structuring and modelling.

Dispositio textus

- a. Critical apparatus.
- b. Editorial criteria.
- c. Fixation.

Developing An Integrated TEI Viewer for East Asian Classics

Kiyonori Nagasaki¹, Jun Homma², Masahiro Shimoda³

¹International Institute for Digital Humanities and Keio University, ²FLX STYLE CO., LTD., ³Musashino University and The University of Tokyo (Japan)

In East Asia in general, including Japan, the diffusion of TEI has been extremely limited. Therefore, the authors have made various efforts to overcome this barrier. In the process, we have realized the need for a viewer that meets the needs of East Asian Classics, which are mostly written vertically, and that can be easily used while focusing on vertical display. In order to familiarize researchers of the East Asian Classics with TEI, it was necessary to provide a viewer that could be used easily, while focusing on the vertical display of the text. To this end, Nagasaki developed a simple vertical viewer based on CETELcean, and customized it to meet each researcher's needs as Nagasaki presented in the TEI conferences before.

Moreover, Mr. Homma joined the team and integrated the customized viewers into a single viewer, while allowing users to easily customize some functions. This is the TEIviewer4EAJ presented here. It supports several major styles of literature in East Asian classics (especially Japanese), especially those related to critical apparatus, waka poetry, and IIF images, and is currently very popular. If a fragment in IIF image is described by <zone>, the part can be focused in the viewer. It also provides a function to display the relationship between xml:id and @ref or @corresp in a graph. In addition, in response to requests for special customization, we have developed a viewer [2] for trilingual texts based on the TEIviewer4EAJ. This viewer is also equipped with functions such as the ability to display multiple translations in parallel and to magnify images on the corresponding picture scrolls.

In the future, we plan to convert what is possible from this customized version back to the integrated version, while further developing an integrated viewer.

Desarrollo de un visor TEI integrado para los clásicos de Asia oriental

Kiyonori Nagasaki¹, Jun Homma², Masahiro Shimoda³

¹Instituto Internacional de Humanidades Digitales y Universidad de Keio, ²FLX ESTILO CO., LTD., ³Universidad Musashino y Universidad de Tokio (Japón)

La difusión de la TEI ha sido extremadamente limitada en Asia oriental en general, incluido Japón. Por ello, los autores han realizado diversos esfuerzos para superar esta barrera. En el proceso, nos hemos dado cuenta de la necesidad de un visor que satisfaga las necesidades de los clásicos de Asia oriental, que en su mayoría están escritos verticalmente, y que pueda usarse fácilmente mientras se enfoca en la visualización vertical. Para familiarizar a los investigadores de los clásicos de Asia oriental con la TEI, era necesario proporcionar un visor que pudiera usarse fácilmente y, al mismo tiempo, centrarse en la visualización vertical del texto. A tal efecto, Nagasaki desarrolló un visor vertical simple basado en CETEIcean, y lo personalizó para satisfacer las necesidades de cada investigador como Nagasaki presentó anteriormente en las conferencias de la TEI.

Por otra parte, Homma se unió al equipo e integró los visores personalizados en un solo visor, lo que permitió a los usuarios personalizar fácilmente algunas funciones. El mismo es el TEIviewer4EAJ que se presenta aquí. Soporta varios estilos importantes de literatura en los clásicos de Asia oriental (en especial los japoneses), particularmente aquellos relacionados con el aparato crítico, la poesía waka y las imágenes IIIF, y es muy popular en la actualidad. Si un fragmento en una imagen IIIF se describe mediante <zone>, se puede enfocar la parte en el visor. Además, proporciona una función para mostrar la relación entre xml:id y @ref o @corresp en un gráfico. Asimismo, hemos desarrollado un visor para textos trilingües basado en el TEIviewer4EAJ en respuesta a solicitudes de personalización especial. Este visor también está equipado con funciones tales como la capacidad de mostrar múltiples traducciones en paralelo y la de ampliar imágenes en los correspondientes rollos de imágenes.

En el futuro, planeamos convertir lo que sea posible de esta versión personalizada a la versión integrada, mientras seguimos desarrollando un visor integrado.

Aplicación de tecnologías de digitalización y codificación para la preservación del patrimonio musical de Guatemala

Martha Eladia María Thomae Elias

Universidade NOVA de Lisboa (Portugal)

En algunos países como Guatemala hay poco trabajo de digitalización y codificación de documentos musicales con fines de preservación y acceso. Este artículo presenta la labor de digitalización y codificación de los libros de polifonía de la Catedral de Guatemala que fue llevado a cabo con el propósito de preservar dicho patrimonio musical y darlo a conocer fuera de las paredes del archivo de la Catedral. Esta colección de libros de música es un ejemplo de la transmisión cultural de la música europea en Latinoamérica y es parte importante del patrimonio e historia del país. Con el objetivo de colaborar en su preservación y diseminación, obtuve permiso de parte del Canciller de la Curia Eclesiástica para la elaboración de un proyecto piloto para la digitalización y codificación del primero de estos seis libros, el GuatC 1. Este artículo presenta el proceso seguido para estos fines, el cuál involucra diferentes tecnologías de digitalización y la integración de distintas herramientas de codificación musical. Finalmente, el artículo da acceso a las imágenes a color y alta resolución de los folios del libro GuatC 1 y a su corpus de música codificada en MEI, la cuál puede ser renderizada y escuchada en línea.

Applying digitization and coding technologies for the preservation of the musical heritage of Guatemala

Martha Eladia María Thomae Elias

NOVA University of Lisbon (Lisboa)

There is little work regarding digitizing and encoding musical documents for preservation and access purposes in countries like Guatemala. This article presents the digitization and encoding of the polyphonic choir books held at Guatemala City's Metropolitan Cathedral, with the goal of preserving and making its musical heritage known outside of the walls of the archive of the Cathedral. This collection of music books is an example of the cultural transmission of European music in Latin America and is an important part of the country's heritage and history. With the aim of collaborating in its preservation and dissemination, I obtained permission from the Chancellor of the Curia to develop a pilot project on the digitization and encoding of the first of these six books: GuatC 1. This article shows the process followed to that end, which involves different digitization technologies and the integration of different music encoding tools. Eventually, the article gives access to the colour images and high-resolution pages of the book GuatC 1 and to its corpus of music encoded in MEI, which can be rendered and listened to online.

From Cinema to Publishing, there and back: Encoding a Corpus of Letters Between Filmmakers and Publishers in XML-TEI

Marco De Cristofaro

Université de Mons (Belgium)

The connection between literature and cinema has been widely examined, yet the nuanced interplay between the film industry and the publishing market has been underexplored. As part of a larger research project, the present paper aims to fill this gap by examining the dynamics between the film industry and the Italian publishing market from the 1960s to the 1980s. During the selected period, film directors influenced the selection processes of publishing houses and consequently impacted reading patterns and the book market. To analyze such a network, the study reconstructs a corpus of published and unpublished letters between filmmakers and publishers using the TEI P5 guidelines. We will focus on a specific case like the correspondence between Federico Fellini and different publishing houses. Following established models like the Darwin Correspondence Project (<https://www.darwinproject.ac.uk>), the Van Gogh Letters Project (<https://tei-c.org/activities/projects/vincent-van-gogh-the-letters/>), and Bellini Digital Correspondence (Del Grosso & Spampinato, 2023, <http://bellinicorrespondence.cnr.it>), the paper employs TEI elements <CorrespContext> and <CorrespDesc> to provide a comprehensive framework for examining the correspondence. We will then describe the individuals mentioned in the letters and their roles, following module 13 of the TEI Guidelines. Additionally, the study uses the <ptr> and <target> elements to trace recommendations of specific books, linking letters related to the same work and tracking the evolution of proposed works from text to published volume. Drawing on methodologies from the DraCor Project (Fischer et al., 2019), visualization software facilitates network analysis and allows for querying the corpus.

This research introduces an innovative model for exploring interactions between the film industry and publishing houses, offering new insights into the influence of creative industries on the book market. The encoding of letters between publishers and filmmakers in XML-TEI is a foundation for expanding research in this domain and enhances our understanding of the interactions between the two fields.

Del cine a la edición, de ida y vuelta: codificando un corpus de cartas entre cineastas y editores en XML-TEI

Marco De Cristofaro

Universidad de Mons (Bélgica)

La conexión entre la literatura y el cine ha sido ampliamente examinada, pero no así la interacción matizada entre la industria cinematográfica y el mercado editorial. Como parte de un proyecto de investigación más amplio, el presente artículo intenta llenar este vacío estudiando la dinámica entre la industria cinematográfica y el mercado editorial italiano desde los años sesenta hasta los ochenta. Durante el período seleccionado, los directores de cine influyeron en los procesos de selección de las editoriales y, en consecuencia, impactaron en los patrones de lectura y el mercado del libro. Para analizar dicha red, el estudio reconstruye un corpus de cartas publicadas y no publicadas entre cineastas y editores utilizando las directrices TEI P5. Nos centraremos en un caso concreto como es la correspondencia entre Federico Fellini y distintas editoriales. Siguiendo modelos establecidos como el Proyecto de Correspondencia Darwin (<https://www.darwinproject.ac.uk>), el Proyecto Cartas de Van Gogh (<https://tei-c.org/activities/projects/vincent-van-gogh-the-cards/>), y Bellini Digital Correspondence (Del Grosso y Spampinato, 2023, <http://bellinidigitalcorrespondence.cnr.it>), el artículo emplea elementos TEI <CorrespContext> y <CorrespDesc> a fin de proporcionar un marco integral para examinar la correspondencia. Luego, describiremos a las personas mencionadas en las cartas y sus funciones, siguiendo el módulo 13 de las Directrices TEI. Además, el estudio utiliza los elementos <ptr> y <target> para rastrear recomendaciones de libros específicos, vincular cartas relacionadas con el mismo trabajo y rastrear la evolución de los trabajos propuestos desde el texto hasta el volumen publicado. Basándose en metodologías del Proyecto DraCor (Fischer et al., 2019), el software de visualización facilita el análisis de la red y permite consultar el corpus.

Esta investigación presenta un modelo innovador para explorar las interacciones entre la industria cinematográfica y las editoriales, ofreciendo nuevos conocimientos sobre la influencia de las industrias creativas en el mercado del libro. La codificación de cartas entre editores y cineastas en XML-TEI es una base para ampliar la investigación en este ámbito y mejorar nuestra comprensión sobre las interacciones entre los dos campos.

Cartographer App

Hizkiel Mitiku Alemayehu, Peter Stadler, Joachim Veit

University of Paderborn (Germany)

The Cartographer app is an open-source image markup tool that automatically creates corresponding bars (measures) in Music Encoding Initiative (MEI). It integrates with

public libraries ' resources with features like image file uploading, IIF manifest loading, and MEI file generation. Additionally, it uses an AI library, i.e., Measure Detector, to automate bar position detection. While currently focused on MEI, plans include expanding to TEI and setting it apart with its GitHub and IIF integration for annotating musical and non-musical elements.

La aplicación Cartographer

Hizkiel Mitiku Alemayehu, Peter Stadler, Joachim Veit

Universidad de Paderborn (Alemania)

La aplicación Cartographer es una herramienta de marcado de imágenes de código abierto que crea, de manera automática, las barras de compás correspondientes en MEI (Music Encoding Initiative). Se integra a los recursos de las bibliotecas públicas con funciones tales como carga de archivos de imágenes, carga de manifiestos IIF y generación de archivos MEI. Asimismo, utiliza una biblioteca de IA, es decir, Measure Detector, para automatizar la detección de la posición de la barra. Si bien actualmente se centra en MEI, los planes incluyen expandirse a TEI y diferenciarlo con su integración en GitHub e IIF para anotar elementos musicales y no musicales.

Long Papers / Presentaciones extensas

Integrating TEI XML with Existing Semantic Web Practices for Enhanced Accessibility and Interoperability in Scholarly Editions

Zsófia Fellegi

HUN-REN Research Centre for the Humanities (Hungary)

In recent years, the integration of the semantic web and linked open data has emerged as a pivotal area in digital philology. This discussion has observed numerous applications of semantic web technology and graph data models in modelling data for scholarly editions. Notably, there has been a shift away from the traditional TEI-XML format, initially tailored for digital scholarly editing needs.

My presentation will propose a refined architecture that aims to overcome the limitations observed in recent digital philological experiments. The prevalent approach –developing bespoke ontologies, data structures, and corresponding software– tends to isolate

digital philology from broader scholarly engagement, inadvertently perpetuating the exclusivity of each edition.

Our framework, hosted at DigiPhil (digiphil.hu), leverages TEI-XML to ensure the accessibility of texts to a diverse academic audience, including historians and linguists. This accessibility is not limited to traditional close reading but extends to computer-assisted distant reading. We are actively developing tools that facilitate the conversion of XML into various data formats, such as plain text, CSV, and LaTeX.

While TEI-XML inherently possesses semantic properties, it does not alone bridge the gap between a document network and a data network as envisaged by Berners-Lee. Previous efforts to integrate TEI XML with the semantic web have fallen short, failing to elevate the use of philological data beyond its original academic confines to a broader cultural and scientific arena.

Our proposed architecture emphasizes the integration of services and software such as WikiData, GitHub, Zenodo.org, Wikibase, and Invenio RDM –cornerstones of the open data philosophy. However, linking these platforms is not straightforward. This presentation will outline the entire workflow, from the initial editing of scholarly texts to the publishing of semantically rich linked data, describing the metadata relational network of larger text units and the practices of semantic annotation and linking of smaller text segments.

Ultimately, I will showcase the DigiPhil infrastructure and the comprehensive workflow we employ, from inception to both print and digital publication, through a case study of a multilingual (Latin and English) digital scholarly edition.

Integración de TEI XML con prácticas de la web semántica existentes a fin de mejorar la accesibilidad y la interoperabilidad en las ediciones filológicas

Zsófia Fellegi

Centro de Investigación en Humanidades HUN-REN (Hungría)

En los últimos años, la integración de la web semántica y los datos abiertos enlazados se ha convertido en un área fundamental en la filología digital. Se han observado numerosas aplicaciones de la tecnología de web semántica y modelos de datos gráficos en el modelado de datos para ediciones filológicas. En particular, se ha producido un alejamiento del formato tradicional TEI XML, inicialmente diseñado para satisfacer las necesidades de la edición digital filológica.

Mi presentación propondrá una arquitectura refinada cuyo objetivo es superar las limitaciones observadas en los recientes experimentos filológicos digitales. El enfoque predominante (desarrollar ontologías, estructuras de datos y el software

correspondiente a medida) tiende a aislar la filología digital de un compromiso académico más amplio, perpetuando, de forma involuntaria, la exclusividad de cada edición.

Nuestro marco, alojado en DigiPhil (digiphil.hu), utiliza XML-TEI para garantizar la accesibilidad de los textos a una audiencia académica diversa, incluidos historiadores y lingüistas. Esta accesibilidad no se limita a la lectura cercana tradicional, sino que se extiende a la lectura a distancia asistida por computadora. Estamos desarrollando activamente herramientas que facilitan la conversión de XML a varios formatos de datos, como texto sin formato, CSV y LaTeX.

Si bien XML-TEI posee propiedades semánticas de modo inherente, no cierra la brecha entre una red de documentos y una red de datos prevista por Berners-Lee. Los esfuerzos realizados anteriormente, a fin de integrar XML-TEI con la web semántica, han fracasado y no han logrado promover el uso de datos filológicos más allá de sus límites académicos originales a un ámbito cultural y científico más amplio.

Nuestra propuesta enfatiza la integración de servicios y software, tales como WikiData, GitHub, Zenodo.org, Wikibase e Invenio RDM, piedras angulares de la filosofía de datos abiertos. Sin embargo, vincular estas plataformas no es sencillo. Esta presentación describirá todo el flujo de trabajo, desde la edición inicial de textos hasta la publicación de datos enlazados enriquecidos semánticamente, al detallar la red relacional de metadatos de unidades mayores de texto, las prácticas de anotación semántica y la vinculación de segmentos de texto más pequeños.

En última instancia, mostraré la infraestructura de DigiPhil y el flujo de trabajo integral que empleamos, desde el inicio hasta la publicación impresa y digital, mediante un estudio de caso de una edición digital filológica multilingüe (latín e inglés).

Can we apply the new CMC chapter to the TEI Listserv Archives? An experiment with TEI for Correspondence and Computer-Mediated Communication

Elisa Beshero-Bondar¹, Syd Bauman²

¹Penn State Erie, ²Northeastern University (United States of America)

The TEI Technical Council is working with the Computer-Mediated Communication (CMC) special interest group (SIG) on introducing a new chapter on CMC for the TEI Guidelines; they expect the new chapter to be released either a few months before or a month after the conference. We propose to test the applicability of the new module to e-mail by encoding a subset of the TEI Listserv archive, and reporting on our successes, failures, and problems.

The authors have been involved in a project to transfer those TEI mailing lists currently on the Brown University Listserv server, many of them dating from the 1990s, to a Listserv server at Penn State University. Simultaneously we have been reviewing and working on the introduction of the draft CMC chapter in its late stages of development in 2024. It is not clear to us how well the TEI with the new CMC Guidelines would apply to the archiving of e-mail in general, and in particular to e-mail from a mailing list. At the time of this writing, the draft CMC chapter primarily addresses the kinds of dialogic, conversational messages we encounter in chat forums, social media platforms, and discussion boards as amenable to TEI encoding. These are media that the CMC draft authors Michael Beißwenger and Harald Lungen described as requiring packaging into “products,” or “posts” prior to transmission over a network.

While the draft chapter does not specifically address e-mail, we want to determine how much of the encoding proposed for CMC could apply to e-mail messages posted in the conversational space of an email listserv. We expect that e-mail messages can be addressed in the encoding provided for correspondence introduced to the Guidelines, but we also wonder to what extent the CMC encoding can be blended with TEI encoding for correspondence in encoding in representing metadata about the transmission and distribution of messages. We expect that the CMC encoding is better suited to documenting the distribution of a Listserv post, and that Listserv messages shared to a community of recipients are better served by CMC encoding, while an e-mail archive representing an individual's personal communications over time may be better represented by the TEI correspondence encoding.

The authors propose to explore the new encoding of CMC introduced to the TEI Guidelines by applying them at scale to a subset of TEI Listserv archives. We will work with the Listserv archive format currently supplied by Brown and Penn State Universities (versions 16.5 and 17.0 respectively). We will document our steps in transforming the Listserv archive format to TEI using text manipulation tools like Perl, Python, and XSLT. The authors expect that they will each have distinct ideas about how to apply the CMC and correspondence encoding to the data we are working with. We will document and present where our perspectives diverge, and we will seek the thoughts of TEI conference attendees on elaborating best practices for the encoding of e-mail list archives.

¿Podemos aplicar el nuevo capítulo de CMC a los archivos *Listserv* de la TEI? Experimentando con la TEI para la correspondencia y la comunicación mediada por computadora

Elisa Beshero-Bondar¹, Syd Bauman²

¹Penn State Erie, ²Universidad del Noreste (Estados Unidos de América)

El Consejo Técnico de TEI está trabajando con el grupo de interés especial (SIG) de Comunicación Mediada por Computadora (CMC) para introducir un nuevo capítulo sobre CMC para las Directrices de TEI. Los integrantes esperan que el nuevo capítulo se publique unos meses antes o un mes después de la conferencia. Proponemos probar la aplicabilidad del nuevo módulo al correo electrónico codificando un subconjunto del archivo Listserv (servidor de listas de correo electrónico) de la TEI e informar sobre nuestros éxitos, fracasos y problemas.

Los autores han participado en un proyecto para transferir las listas de correo de TEI, que se encuentran actualmente en el servidor Listserv de la Universidad de Brown (muchas datan de la década de 1990) a un servidor Listserv en la Universidad de Penn State. Al mismo tiempo, hemos estado revisando y trabajando en la introducción del borrador del capítulo CMC, en sus últimas etapas de desarrollo en 2024. No nos queda claro qué tan bien se aplicaría la TEI con las nuevas Directrices CMC al archivo de correo electrónico en general, y en particular al correo electrónico desde una lista de correo. Al momento de escribir este artículo, el borrador del capítulo de CMC aborda principalmente los tipos de mensajes dialógicos y conversacionales que encontramos en foros de chat, plataformas de redes sociales y foros de discusión que son susceptibles de codificación TEI. Los autores del borrador de CMC, Michael Beißwenger y Harald Lungen, indicaron que estos medios requieren empaquetado en "productos" o "publicaciones" antes de su transmisión a través de una red.

Si bien el borrador del capítulo no aborda específicamente el correo electrónico, queremos determinar qué parte de la codificación propuesta para CMC podría aplicarse a los mensajes de correo electrónico publicados en el espacio conversacional de un servidor de listas de correo electrónico. Esperamos que los mensajes de correo electrónico puedan abordarse en la codificación proporcionada para la correspondencia introducida en las Directrices, pero también nos preguntamos hasta qué punto se puede combinar la codificación CMC con la codificación TEI para la correspondencia en la codificación al representar metadatos sobre la transmisión y distribución de mensajes. Queremos que la codificación CMC sea más adecuada para documentar la distribución de una publicación de Listserv, y que los mensajes de Listserv compartidos con una comunidad de destinatarios sean mejor atendidos por la codificación CMC, mientras que un archivo de correo electrónico que represente las comunicaciones personales de un individuo a lo largo del tiempo pueda ser representado de mejor manera por la codificación de correspondencia TEI.

Los autores proponen explorar la nueva codificación de CMC introducida en las Directrices TEI aplicándolas a escala a un subconjunto de archivos Listserv de la TEI. Trabajaremos con el formato de archivo Listserv proporcionado actualmente por las universidades Brown y Penn State (versiones 16.5 y 17.0 respectivamente).

Explicaremos los pasos seguidos para transformar el formato de archivo Listserv a TEI utilizando herramientas de manipulación de texto tales como Perl, Python y XSLT. Los autores esperan que cada uno tenga ideas distintas acerca del modo de aplicar la CMC y la codificación de correspondencia a los datos con los que estamos trabajando. Documentaremos y presentaremos dónde divergen nuestras perspectivas, y buscaremos las opiniones de los asistentes a la conferencia TEI respecto de la elaboración de mejores prácticas para la codificación de archivos de listas de correo electrónico.

Towards an LLM-powered encoding workflow for plays

L. Giovannini^{1,2}, D. Skorinkin¹

¹University of Potsdam (Germany), ²University of Padova (Italy)

Encoding new texts in TEI-XML format plays a central role in established research projects such as DraCor (Fischer et al. 2019), a major computational infrastructure hosting ‘programmable corpora’ comprising thousands of dramatic texts. As outlined by Giovannini et al. 2023, current DraCor production workflows change according to the initial markup of the texts to be onboarded: while computational transformations (mostly Python or XSLT scripts) are applied to sources with basic (HTML) or advanced (XML) markup, texts with no markup are usually encoded through the application of the lightweight markdown language easydrama (Skorinkin 2024) and a successive scripted conversion.

The rise of large language models (LLMs), however, promises to further automate such encoding tasks, and scholars have been already exploring the potential of generative AI by developing advanced prompt-engineering techniques to enhance outputs (Czmiel et al. 2024, Pollin 2023). Most efforts, however, seem to have been devoted to shorter textual forms, like letters (e.g. Pollin, Steiner, and Zach 2023), which present comparatively fewer challenges in encoding than longer texts like plays.

In this contribution, we present a proof-of-concept demonstrating how LLMs can be efficiently integrated into the corpus building pipeline of a standard DraCor corpus. To this aim, we conduct a series of experiments with several state-of-art models, including both commercial (OpenAI’s GPT, Google’s Gemini) and open-source (Meta’s Llama, Mistral AI’s Mixtral) products, to assess to which extent a largely automated TEI encoding of plays is possible. We therefore discuss strengths and limitations of this approach and eventually propose a set of tailored LLMs prompts which can be used to generate partially DraCor-ready files from raw text input. In future developments, we also envision the possibility of fine-tuning existing LLMs with specific DraCor data in order to improve its performance.

Hacia un flujo de trabajo de codificación para obras de teatro impulsado por LLM

L. Giovannini^{1,2}, D. Skorinkin¹

¹Universidad de Potsdam (Alemania), ²Universidad de Padua (Italia)

La codificación de nuevos textos en formato TEI-XML desempeña un papel central en proyectos de investigación tales como DraCor (Fischer et al. 2019), una importante infraestructura computacional que alberga "corpus programables" que comprenden miles de textos dramáticos. Como lo describen Giovannini et al. 2023, los flujos de trabajo de producción actuales de DraCor cambian según el marcado inicial de los textos que se van a incorporar: mientras que las transformaciones computacionales (principalmente scripts Python o XLST) se aplican a fuentes con marcado básico (HTML) o avanzado (XML), los textos sin marcado generalmente son codificados mediante la aplicación del lenguaje de marcado ligero EasyDrama (Skorinkin 2024) y una conversión secuencial guionizada.

Sin embargo, el auge de los grandes modelos de lenguaje (LLMs) promete automatizar aún más dichas tareas de codificación, y los académicos ya han estado explorando el potencial de la IA generativa mediante el desarrollo de técnicas avanzadas de ingeniería de instrucciones para mejorar los resultados (Czmiel et al. 2024, Pollin 2023). No obstante, la mayoría de los esfuerzos parecen haberse dedicado a formas textuales más breves, como las cartas (p. ej., Pollin, Steiner y Zach, 2023), que presentan comparativamente menos desafíos en cuanto a la codificación que textos más largos como las obras de teatro.

En esta contribución presentamos una prueba de concepto que demuestra cómo los LLMs se pueden integrar, de manera eficiente, al canal de creación de corpus de un corpus DraCor estándar. Para ello, llevamos a cabo una serie de experimentos con varios modelos de última generación, que incluyen tanto productos comerciales (GPT de OpenAI, Gemini de Google) como de código abierto (Llama de Meta, Mixtral de Mistral AI), para evaluar, en qué medida, es posible una codificación TEI automatizada de obras de teatro. Por consiguiente, analizamos las fortalezas y limitaciones de este enfoque y, finalmente, proponemos un conjunto de indicaciones adaptadas a los LLMs que se pueden utilizar para generar archivos parcialmente listos para DraCor a partir de la entrada de texto sin formato. Asimismo, visualizamos la posibilidad de ajustar los LLMs existentes con datos específicos de DraCor para mejorar su rendimiento en desarrollos futuros.

Editing Early-Born Digital Text in TEI

Torsten Roeder

University of Würzburg (Germany)

With the increasing prevalence of digital inheritances and digital literary publication forms, digital cultural heritage is gradually coming into the purview of editorial philology. Due to electronic storage, "born digital" resources do not possess a fixed material form but only temporary visualizations. Insofar, their materiality is transmitted to devices and storage media, to whose zeitgeist-specific technological design the storage is bound. A closer examination of digital remnants from the 1980s makes clear the fundamental technological change that has taken place in just a few decades: Neither software nor files are accessible using today's common systems, and handling original storage media and devices requires care and expertise. Emulation is only possible if corresponding digital images of the original storage media are available. For the scholarly editing of early digital material, it follows that digital is not equal to digital, but that digitality arises in various manifestations depending on the respective digital environment. On a ZX Spectrum, this was entirely different from on an Apple II or a Commodore Amiga 4000. This, in turn, creates different conditions for textual criticism, depending on the specific historic or system-specific concept of digitality.

To approach these conditions, it seems reasonable to examine the electronic representation of text more closely. In the simplest case, there is a direct, symbolic connection between the binary code on a storage medium and the display on a display device, which is defined by a standard such as ASCII (later included in Unicode). However, depending on the historical environment, a multi-tiered chain can be expected: Companies like Commodore maintained proprietary standards in which the outward appearance of the character inventory could also be individually customized, while standard formats for character sets or their documentation did not exist. For a correct mapping of the byte sequence to the semantics of individual characters, sufficient knowledge of the adaptations is therefore necessary, because otherwise neither can the intended character be inferred from the encoding nor can the encoding be inferred from the outward form. However, this also means that the transmission of text cannot occur without the transmission of the system environment. The same applies to historical compression algorithms, which were often applied due to limited storage space but now need to be determined due to the lack of specifications when one wants to access the digital original text.

But if the relationship between the displayed character and the underlying encoding only works best in the context of the original environment: What then is the text to be edited in the digital realm? And how can this relationship be meaningfully incorporated into textual criticism? Do emulations possibly serve a facsimile function? And finally, the

crucial question: What does the TEI need to capture firstly the necessary information in the metadata and secondly to encode the different digital text layers? The contribution discusses these questions theoretically and with reference to several examples from a current editing project.

Edición de textos digitales tempranos en TEI

Torsten Roeder

Universidad de Würzburg (Alemania)

El patrimonio cultural digital está entrando, de manera gradual, en el ámbito de la filología editorial a causa de la creciente preponderancia de herencias digitales y formas de publicación literaria digital. Debido al almacenamiento electrónico, los recursos "de origen digital" no poseen una forma material fija sino sólo visualizaciones temporales. En la medida de lo posible, su materialidad se transmite a dispositivos y medios de almacenamiento vinculado al diseño tecnológico del espíritu de la época. Un examen más detenido de los rastros digitales de la década de 1980 deja claro el profundo cambio tecnológico que ha tenido lugar en tan sólo unas pocas décadas: ni el software ni los archivos son accesibles utilizando los sistemas actuales, y el manejo de dispositivos y medios de almacenamiento originales requiere cuidado y pericia. Sólo es posible emular si las imágenes digitales correspondientes del medio de almacenamiento original están disponibles. En cuanto a la edición filológica de material digital temprano, se deduce que lo digital no es igual a lo digital, sino que la digitalidad surge en diversas manifestaciones dependiendo del entorno digital respectivo. Esto era completamente diferente en un ZX Spectrum que en un Apple II o un Commodore Amiga 4000. A su vez, esto crea diferentes condiciones para la crítica textual, dependiendo del concepto de digitalidad histórico específico o específico del sistema.

Para abordar estas condiciones, parece razonable examinar detenidamente la representación electrónica del texto. En el más sencillo de los casos, existe una conexión directa y simbólica entre el código binario de un medio de almacenamiento y la visualización en un dispositivo de visualización, que se define mediante un estándar tal como ASCII (posteriormente incluido en Unicode). Sin embargo, dependiendo del entorno histórico, se puede esperar una cadena de varios niveles: empresas como Commodore mantenían estándares propietarios en los que el aspecto exterior del inventario de caracteres también podía personalizarse individualmente, mientras que no existían formatos estándar para los conjuntos de caracteres o su documentación. Por lo tanto, se necesita tener un conocimiento suficiente respecto de las adaptaciones para realizar un mapeo correcto de la secuencia de bytes con la semántica de los caracteres individuales, ya que, de lo contrario, ni el carácter deseado puede inferirse

de la codificación ni la decodificación puede deducirse de la forma exterior. Sin embargo, esto también significa que la transmisión de texto no puede realizarse sin la transmisión del entorno del sistema. Lo mismo se aplica a los algoritmos de comprensión históricos, que a menudo se aplicaban debido al espacio de almacenamiento limitado pero que ahora deben determinarse por la falta de especificaciones cuando se quiere acceder al texto digital original.

Pero, si la relación entre el carácter mostrado y la codificación subyacente sólo funciona mejor en el contexto del entorno original: ¿cuál es entonces el texto que se va a editar en el ámbito digital? ¿Y cómo puede incorporarse significativamente esta relación a la crítica textual? ¿Es posible que las emulaciones cumplan una función de facsímil? Y, por último, la pregunta crucial: ¿Qué necesita la TEI para capturar, en primer lugar, la información necesaria en los metadatos y, en segundo lugar, para codificar las diferentes capas de texto digital? En esta ocasión analizaremos dichas cuestiones teóricamente y con referencia a varios ejemplos de un proyecto de edición actual.

Integrating Deep Learning and Philology: Challenges and Opportunities in the Digital Processing of János Arany's Legacy

Gábor Palkó

HUN-REN Research Centre for the Humanities (Hungary)

In the first decades of the 21st century, we can witness two parallel and closely related trends in the fields of culture and science. On the one hand, Artificial Intelligence is transforming and replacing various established cultural practices at an unforeseeable pace. On the other hand, partly due to the digitalization of cultural heritage and partly due to the huge volume of 'born digital' materials being produced, we are literally seeing the creation of data sets and networks of unimaginable scale.

However, within the discourse of digital heritage, alongside easily processable and publishable printed or 'born digital' materials, the "real" –that is, handwritten– manuscripts tend to be overshadowed, as they cannot be made searchable with general models that do not consider the specific characteristics of the document group in question. A particular problem is that AI tools work better in major world languages spoken by populations of hundreds of millions, thus for instance, Hungarian handwritten documents are exceptionally underrepresented in the entirety of the digital cultural heritage.

Addressing this issue is one of the main goals of the National Laboratory for Digital Heritage project. Led and professionally guided by digital humanities experts from ELTE Faculty of Humanities, this project, which is a collaboration of public collections and research institutions, considers its primary task to be the application of AI tools optimized for the Hungarian language in public collections. One of the most significant achievements of this work is the development of a handwriting recognition model that has made it possible to make János Arany's official documents searchable, thereby making an extremely valuable document corpus accessible to researchers and the general public.

The present lecture, in addition to briefly outlining the HTR processing of a significant portion –approximately 30,000 pages– of János Arany's legacy, fundamentally focuses on two issues. The first issue examines the potential and risks of various deep-learning technologies such as synthetic handwriting generation and large language models (LLMs) in “improving” the text quality of a corpus that is too large to be checked by human involvement, and in making it researchable. The second issue, which I discuss in detail in my presentation, addresses how a corpus converted into text by artificial intelligence can be integrated within the framework of critical text editions created by philologists. Specifically, it queries whether the TEI XML markup language is optimal for the publication of uncorrected HTR-ed documents and how texts of high philological demand in digital scholarly editions can coexist on a single platform with documents “read” by machines. I will present these issues in the context of specific IT developments.

Integrando el aprendizaje profundo y la filología: desafíos y oportunidades en el procesamiento digital del legado de János Arany

Gábor Palko

Centro de Investigación en Humanidades HUN-REN (Hungría)

En las primeras décadas del siglo XXI podemos observar dos tendencias paralelas y estrechamente relacionadas en los campos de la cultura y la ciencia. Por un lado, la Inteligencia Artificial está transformando y reemplazando diversas prácticas culturales establecidas a un ritmo imprevisible. Por otro lado, en parte debido a la digitalización del patrimonio cultural y en parte al enorme volumen de producción de materiales “nacidos digitales”, estamos literalmente asistiendo a la creación de conjuntos de datos y redes a una escala inimaginable.

Sin embargo, dentro del discurso del patrimonio digital, junto con los materiales impresos o “nacidos digitales” fácilmente procesables y publicables, los manuscritos “reales” (es decir, escritos a mano) tienden a quedar eclipsados ya que no es posible

buscarlos con modelos generales que no consideran las características específicas del grupo de documentos en cuestión. Un problema particular es que las herramientas de inteligencia artificial funcionan mejor en los principales idiomas del mundo hablados por poblaciones de cientos de millones; así, por ejemplo, los documentos escritos a mano en húngaro están excepcionalmente infrarrepresentados en todo el patrimonio cultural digital.

Abordar esta cuestión es uno de los principales objetivos del proyecto del Laboratorio Nacional de Patrimonio Digital. Este proyecto, dirigido y guiado profesionalmente por expertos en humanidades digitales de la Facultad de Humanidades de ELTE, es una colaboración de colecciones públicas e instituciones de investigación, y considera que su tarea principal es la aplicación de herramientas de inteligencia artificial optimizadas para el idioma húngaro en colecciones públicas. Uno de los logros más significativos de este trabajo es el desarrollo de un modelo de reconocimiento de escritura que ha permitido hacer que los documentos oficiales de János Arany sean consultables, y, de este modo, un corpus documental de gran valor es accesible a los investigadores y al público en general.

La presente conferencia, además de esbozar brevemente el procesamiento HTR de una parte importante (aproximadamente 30.000 páginas) del legado de János Arany, se centra fundamentalmente en dos cuestiones. En primer lugar, examina el potencial y los riesgos de diversas tecnologías de aprendizaje profundo, tales como la generación de escritura sintética y los grandes modelos de lenguaje (LLMs), a fin de “mejorar” la calidad del texto de un corpus que es demasiado grande para ser verificado por la participación humana, y hacerlo investigable. La segunda cuestión, que analizo en detalle en mi presentación, aborda la manera en que un corpus convertido en texto por inteligencia artificial puede integrarse en el marco de las ediciones de textos críticos creados por filólogos. En concreto, se cuestiona si el lenguaje de marcado TEI XML es óptimo para la publicación de documentos editados en HTR sin corregir y de qué modo pueden convivir en una única plataforma textos de alta exigencia filológica en ediciones académicas digitales con documentos “leídos” por máquinas. Presentaré estos temas en el contexto específico de los desarrollos de las Tecnologías de la Información.

Chasing ‘Carmen Nova’: Encoding and Analysis of a TEI Version of the Crime Novella Allegedly Written by Umberto Eco

Frank Fischer¹, Dilan Canan Çakir¹, Viktor J. Illmer¹, Niels Penke², Mark Schwindt¹, Lily Welz¹

¹Free University of Berlin, ²University of Siegen (Germany)

On the first pages of Umberto Eco's world-famous 1980 novel *Il nome della rosa*, we read about a book that the narrator finds in an antiquarian bookshop on Avenida Corrientes in Buenos Aires, which sets the plot in motion. With this contribution to the TEI 2024 Conference, we want to bring Umberto Eco back to Buenos Aires under circumstances that are just as intricate as in Eco's novel. Using a TEI document, our project aims to thoroughly examine an unresolved case concerning the authorship of a crime novella:

At the end of 2022, literary scholar Niels Penke discovered a book titled 'Carmen Nova' on eBay, which had Umberto Eco as the author. The afterword was allegedly written by Roland Barthes. Several copies of the book, which was published by a fictitious Swiss publishing house, are now known to exist. But as it turned out, the 64-page volume is a literary forgery, as nothing of such a work by Umberto Eco is known. The novella disguises itself as a German translation of the alleged Italian original. The plot revolves around the detective search for a certain Carmen, who, however, does not appear to be a concrete person, but rather a world literary concept that has its origins in Mérimée's novella *Carmen*, published in 1845, which in turn was the basis for Bizet's opera of the same name, which premiered in 1875. The *Carmen Nova* of the novella is one of these 'fluctuating individuals' (Eco, 2009, pp. 86-89).

Since Niels Penke's discovery, a community of scholars, interested readers and journalists has formed to find out more about the author (or authors) of this literary forgery, which must have been written in the early 1980s. It is still not known who wrote, printed and circulated this book.

The State and University Library Bremen (SuUB) has made a digital scan of its copy in 2023. In order to be able to carry out digital analyses with the full text of the work, we used OCR to convert the PDF to plain text and, after a round of corrections, encoded it in TEI (Araneda Lavín et al., 2023). We published the encoded file by help of the JavaScript library CETEIcean (Cayless & Viglianti, 2018). Among other things, we have annotated mentioned persons and obvious spelling mistakes, clues to help uncovering the nature of the text. We will present our annotation strategy and present initial analysis results. Using a Jupyter notebook that utilizes the LXML library to extract information from the TEI-encoded version, we generated quantitative results that provide new insights into the making of this mysterious text.

We have released both the TEI and Python code as open source (see bibliography) and hope that the presentation will arouse lively interest among the TEI community in this mystery, which has so far been confined to the German-speaking community. Last not least, it is a nice touch that, through the TEI 2024 Conference, a pseudo-version of Umberto Eco finds his way to Buenos Aires via winding paths.

Persiguiendo a 'Carmen Nova': codificación y análisis de una versión TEI de la novela policíaca presuntamente escrita por Umberto Eco

Frank Fischer¹, Dilan Canan Çakir¹, Viktor J. Illmer¹, Niels Penke², Mark Schwindt¹, Lily Welz¹

¹Universidad Libre de Berlín, ²Universidad de Siegen (Alemania)

En las primeras páginas de la mundialmente famosa novela *Il nome della rosa* de Umberto Eco (1980), leemos acerca de un libro que el narrador encuentra en una librería antigua de la Avenida Corrientes de Buenos Aires, lo que pone en marcha la trama. Con este aporte a la Conferencia TEI 2024, queremos traer a Umberto Eco de regreso a Buenos Aires en circunstancias tan intrincadas como en la novela de Eco. Utilizando un documento TEI, nuestro proyecto pretende examinar en profundidad un caso no resuelto relativo a la autoría de una novela policíaca:

A finales de 2022, el investigador Niels Penke descubrió en eBay un libro titulado "Carmen Nova" cuya autoría se le atribuía a Umberto Eco. El epílogo fue, presuntamente, escrito por Roland Barthes. Ahora se sabe que existen varios ejemplares del libro que fue publicado por una editorial suiza ficticia. Pero se descubrió que el volumen de 64 páginas es una falsificación literaria ya que no se sabe nada de tal obra de Umberto Eco. La novela se disfraza de traducción alemana del supuesto original italiano. La trama gira en torno a la búsqueda detectivesca de una tal Carmen, que, sin embargo, no parece ser una persona concreta sino más bien un concepto literario mundial que tiene su origen en la novela corta *Carmen* de Mérimée, publicada en 1845, que a su vez fue la base para la ópera homónima de Bizet que se estrenó en 1875. La *Carmen Nova* de la novela es uno de estos "individuos fluctuantes" (Eco, 2009, pp. 86-89).

A partir del descubrimiento de Niels Penke, se ha formado una comunidad de académicos, lectores interesados y periodistas para saber más sobre el autor (o autores) de esta falsificación literaria, que debe haber sido escrita a principios de los años 1980. Aún no se sabe quién escribió, imprimió y distribuyó este libro.

La Biblioteca Estatal y Universitaria de Bremen (SuUB) realizó un escaneo digital de su copia en 2023. Con el objetivo de poder realizar análisis digitales con el texto completo de la obra, utilizamos OCR para convertir el PDF a texto sin formato y, luego de una serie de correcciones, lo codificamos en TEI (Araneda Lavín et al., 2023). Publicamos el archivo codificado con la ayuda de CETEIcean, una biblioteca JavaScript (Cayless y Viglianti, 2018). Entre otras cosas, hemos anotado personas mencionadas y errores ortográficos evidentes, pistas que ayudan a descubrir la naturaleza del texto. Presentaremos nuestra estrategia de anotación y los resultados del análisis inicial. Utilizando un cuaderno Jupyter que usa la biblioteca LXML para extraer información de

la versión codificada en TEI, generamos resultados cuantitativos que brindan nuevos conocimientos acerca de la creación de este misterioso texto.

Hemos publicado tanto el código TEI como código Python en abierto y esperamos que la presentación despierte un vivo interés entre la comunidad TEI respecto de este misterio que, hasta ahora, se ha limitado a la comunidad de habla alemana. Por último, si bien no menos importante, es un detalle interesante que, a través de la Conferencia TEI 2024, una pseudo versión de Umberto Eco llegue a Buenos Aires por caminos sinuosos.

Panels / Paneles

¿Qué está pasando con la TEI en español?

Gimena del Rio Riande¹, Susanna Allés-Torrent², Clayton Mc Carl³, Cristian Suárez Giraldo⁴

¹Consejo Nacional de Investigaciones Científicas y Técnicas (República Argentina),

²Universidad de Miami (Estados Unidos de América), ³Universidad de North Florida (Estados Unidos de América), ⁴EAFIT (Colombia)

En los países hispanohablantes el interés creciente por la fijación de textos y su transformación en formato electrónico vino en un primer momento motivado por el análisis de textos a través de concordancias, diccionarios, o de la creación de léxicos específicos. La adopción del sistema Madison para la transcripción de manuscritos, desarrollado en el Hispanic Seminary of Medieval Studies (HSMS) de la Universidad de Wisconsin, ampliamente adoptado para textos medievales hispánicos desde fines de la década del 80 del siglo XX, colaboraría con la necesidad de echar mano de un estándar de codificación que fuera más allá de los formatos y que incluyera tanto información lingüística como semántica. En ese contexto, si bien la TEI surge en estos años en ámbito anglosajón, no tarda en interesar a algunos investigadores activos en los estudios hispánicos durante la siguiente década.

En 1992, en el marco del Congreso de la Lengua Española celebrado en Sevilla, se organizó el panel “La lengua española y las nuevas tecnologías”, donde se empezó a discutir –quizás por primera vez y gracias a Charles Faulhaber– acerca de la TEI para textos en español. No es un dato menor el hecho de que en este evento se presentó la traducción al español de las primeras Guías Directrices de la TEI, elaboradas entre

1988 y 1990. Las “Guidelines for the Encoding and Interchange of Machine-Readable Texts” (TEI P1), traducidas al español por Francisco Marcos Marín y las bibliotecarias argentinas Verónica Zumárraga y Marcela Tabanera, pueden entenderse hoy como claro ejemplo del temprano interés de la comunidad hispanohablante en la TEI.

La TEI es hoy un estándar extendido en la edición de textos hispánicos, aunque, como bien hemos afirmado en otro lugar, muchas veces deslucido por el contexto de trabajo aislado y desfinanciado en muchas academias. Este panel busca ofrecer una serie de reflexiones sobre los derroteros que esta práctica digital ha tenido en los estudios hispánicos y describir los métodos, prácticas, las necesidades y las colaboraciones que en este momento están definiendo a la comunidad de práctica de la TEI en español. También busca ser un espacio de intercambio y debate con los asistentes al panel, con lo que después de las breves presentaciones, se buscará abrir un espacio de conversación acerca de los posibles futuros de la comunidad.

What is happening with the TEI in Spanish?

Gimena del Rio Riande¹, Susanna Allés-Torrent², Clayton Mc Carl³, Cristian Suárez Giraldo⁴

¹National Council for Scientific and Technical Research (Argentine Republic) ²University of Miami (United States of America) ³University of Florida (United States of America), ⁴EAFIT (Colombia)

Spanish-speaking countries' growing interest in text markup and its transformation into electronic format initially stemmed from text analysis resources such as concordances, dictionaries, or the creation of specific lexicons. The Madison system for manuscript transcription, which was developed at the Hispanic Seminary of Medieval Studies (HSMS) at the University of Wisconsin, has been widely adopted for Hispanic medieval texts since the late 1980s and contributed to the need for an encoding standard beyond formats including both linguistic and semantic information. In this context, although the TEI emerged in the Anglophone world in the late 1980s, it did not take long for some active researchers in Hispanic studies to become interested in it during the following decade.

In 1992, within the framework of the Congreso de la Lengua Española held in Seville, a group of participants organized the panel “The Spanish language and new technologies” and discussions regarding the Text Encoding Initiative for texts in Spanish began to take place perhaps for the first time and thanks to Charles Faulhaber. It is important to note that the translation in Spanish of the TEI Guidelines, drafted between 1988 and 1990, was presented at this event. “The Guidelines for the Encoding and Interchange of Machine-Readable Texts” (TEI P1), translated into Spanish by Francisco

Marcos Marín and the Argentinean librarians Verónica Zumárraga and Marcela Tabanera, show the early interest of the Spanish-speaking community regarding this standard.

The TEI is now a widely used standard in the editing process of Hispanic texts, although, as we have already stated, it is often overshadowed by the context of isolated and underfunded work in many academies. This panel seeks to offer a set of reflections on the use of the TEI in Hispanic studies and to describe the methods, practices, needs and collaborations that are currently defining the TEI community of practice in Spanish. Furthermore, it aims to be a space for exchange and debate with the attendees. Therefore, once the brief presentations are over, we will open up the conversation and discuss the possible futures of the TEI Spanish-speaking community.

Posters / Pósters

Establishing Best Practices for TEI Encoding of Newspapers: A Case Study of the Darmstädter Tagblatt

Kevin Kuck, Silke Kalmer

Centre for Digital Editions in Darmstadt, Technische Universität Darmstadt (Germany)

This poster addresses the reuse of periodicals and newspapers and proposes a best practices for TEI encoding in digitization projects, focusing on the Darmstädter Tagblatt. As part of the newspaper working group of DHd (Association for Digital Humanities in the German Speaking Areas), we recognize the growing need for standardized TEI encoding to facilitate data reuse across various newspaper projects. Leveraging insights from a recently initiated series of workshops on the topic of reuse of newspaper data, of which the first is taking place in Darmstadt in May 2024, the objective is to explore the potential of TEI in enhancing access to historical newspaper data while fostering collaboration among researchers. Currently, concepts are being developed and will come to fruition after the initial workshop. One idea is the creation of a universal TEI header that is suitable for editions of newspapers as well as periodicals. An ongoing digitization project at ULB Darmstadt, which is funded by DFG (German Research Foundation), will serve as the scientific basis. The poster will provide an overview of the project *A Darmstadt Newspaper in Three Centuries –Digitization of the Darmstädter Tagblatt, 1740-1986*, by Thomas Stäcker, Marcus Müller, Dario Kampkaspar, et. al., highlighting its significance. As one of the longest running periodicals within Germany, it represents a heterogeneous data set and provides an

excellent case study for establishing a best practice. It will also address the challenges and opportunities in TEI encoding of newspapers, proposing best practices tailored to meet the needs of diverse projects. Engaging with the international TEI community, we seek to foster discussions on standardization, collaborative markup, and the revitalization of the SIG Newspapers and Periodicals. By establishing standardized TEI encoding practices for newspapers, we aim to facilitate collaboration, enhance access to historical resources, and advance research in digital humanities.

Establecimiento de mejores prácticas para la codificación TEI de periódicos: un estudio de caso del Darmstädter Tagblatt

Kevin Kuck, Silke Kalmer

Centro de Ediciones Digitales en Darmstadt (CEiD), Universidad Técnica de Darmstadt (Alemania)

Este póster aborda la reutilización de publicaciones periódicas y diarios, y propone mejores prácticas para la codificación TEI en proyectos de digitalización, centrándose en el Darmstädter Tagblatt. Como integrantes del grupo de trabajo de periódicos de DHd (Asociación para las Humanidades Digitales en las Áreas de Habla Alemana) reconocemos la creciente necesidad de una codificación TEI estandarizada para facilitar la reutilización de datos en varios proyectos de periódicos. El objetivo es explorar el potencial de TEI para mejorar el acceso a los datos históricos de los periódicos y, al mismo tiempo, fomentar la colaboración entre los investigadores aprovechando los conocimientos de una serie de talleres acerca de la reutilización de datos periodísticos que comenzaron recientemente –el primero se está llevando a cabo en Darmstadt en mayo de 2024. Actualmente, se están desarrollando los conceptos y se materializarán después del taller inicial. Se propone la idea de crear un encabezado TEI universal que sea adecuado tanto para ediciones de periódicos como para publicaciones periódicas. La base científica será un proyecto de digitalización en curso en la ULB Darmstadt, financiado por la DFG (Fundación Alemana de Investigación). El póster ofrecerá una visión general del proyecto Un periódico de Darmstadt en tres siglos –Digitalización del Darmstädter Tagblatt, 1740-1986 de Thomas Stäcker, Marcus Müller, Dario Kampkaspar, et. al., destacando su importancia. Debido a que es una de las publicaciones periódicas más antiguas de Alemania, representa un conjunto de datos heterogéneos y proporciona un excelente estudio de caso para establecer una mejor práctica. Asimismo, abordará los desafíos y oportunidades en la codificación TEI de periódicos, proponiendo mejores prácticas diseñadas para satisfacer las necesidades de diversos proyectos. Procuramos fomentar debates sobre la estandarización, el marcado colaborativo y la revitalización de los "Periódicos y publicaciones periódicas" del SIG para, de esa manera, colaborar con la comunidad

internacional de TEI, Buscamos facilitar la colaboración, mejorar el acceso a recursos históricos y promover la investigación en humanidades digitales al establecer prácticas estandarizadas de codificación TEI para periódicos.

Developing a Base Format for Heterogeneous Texts According to the TEI P5 Guidelines

Silke Kalmer

University and State Library Darmstadt (Germany)

The University and State Library Darmstadt (Germany) is collecting and digitizing texts, i.e. scientific journal articles and monographs, digital scholarly editions etc., to make them available for open access. Since not only the text types, but also the file formats differ greatly, for example, JATS, TEI or non-XML formats, the goal is to transform these heterogeneous formats into one standardized format. For this, a so-called base format is being developed using the TEI P5 guidelines. The base format is basically a subset of the TEI, selected and put together with consideration of what kind of information is needed, what information is present and how it has to be depicted.

The basic structure consists of a <teiHeader>, a <standOff> for the entities and a <text> containing at least a <body>, also when applicable <front> and <back>. The main focus lies on the metadata in the <teiHeader>. Problems arise on details, for example how to depict a <persName> (do we have a separation into surname and forename in the source material?) or the title (is the title separated into main title and subtitle or one string?). Moreover, the fact that different text types need different TEI modules has to be taken into consideration, for example an edition additionally requires a <msDesc>.

This base format guarantees not only a structured format that is the same for all texts, but also that the handling of the metadata meets the criteria of the library, since the metadata of all texts will ultimately be put into the library catalogue. The texts will be made available and searchable on the library's own publishing platform for digital texts 'TUeditions' (<https://tueditions.ulb.tu-darmstadt.de/>) at the Centre for Digital Editions. The base format itself will get a RelaxNG schema for validation, as well as a documentation in DITA format.

Desarrollo de un formato base para textos heterogéneos según las directrices TEI P5

Silke Kalmer

Universidad y Biblioteca Estatal de Darmstadt (Alemania)

La Universidad y Biblioteca Estatal de Darmstadt (Alemania) está recopilando y digitalizando textos, i.e. artículos de revistas científicas y monografías, ediciones filológicas digitales, etc., para que estén disponibles en acceso abierto. Dado que no sólo los tipos de texto sino también los formatos de archivos difieren mucho (por ejemplo, JATS, TEI o formatos no XML), el objetivo es transformar estos formatos heterogéneos en un formato estandarizado. Para ello, se está desarrollando el llamado formato base siguiendo las directrices TEI P5. El formato base es básicamente un subconjunto del TEI, seleccionado y elaborado teniendo en cuenta qué tipo de información se necesita, qué información está presente y cómo debe representarse.

La estructura básica consta de un <teiHeader>, un <standOff> para las entidades y un <text> que contiene al menos un <body>, y, cuando corresponda, <front> y <back>. La atención se centra principalmente en los metadatos del <teiHeader>. Surgen problemas con los detalles, por ejemplo, al considerar cómo representar un <persName> (¿tenemos una separación entre apellido y nombre en el material fuente?) o el título (¿el título está separado en título principal y subtítulo o es una cadena?). Asimismo, se debe tener en cuenta el hecho de que diferentes tipos de texto necesitan diferentes módulos TEI; por ejemplo, una edición requiere además un <msDesc>.

Este formato base garantiza no sólo un formato estructurado e igual para todos los textos, sino también que el manejo de los metadatos cumpla con los criterios de la biblioteca, ya que los metadatos de todos los textos finalmente se incluirán en el catálogo de la biblioteca. Los textos estarán disponibles y se podrán buscar en la plataforma de publicación de textos digitales de la biblioteca 'TUeditions' (<https://tueditions.ulb.tu-darmstadt.de/>) del Centro de Ediciones Digitales. El propio formato base obtendrá un esquema RelaxNG para su validación, así como una documentación en formato DITA.

Codificación TEI y análisis de redes: a propósito de Calderón Drama Corpus (CalDraCor) v.2.0

Hanno Ehrlicher¹, Antonio Rojas Castro¹, S. Padó², K. Jung², A. Keith²

¹Universidad Eberhard Karl de Tübingen, ²Universidad de Stuttgart (Alemania)

Tras la publicación en su totalidad en acceso abierto en 2022 en DraCor (Fisher et al., 2019) bajo el nombre CalDraCor, la codificación en formato TEI de un corpus de 205 obras teatrales de Calderón de la Barca abrió nuevos caminos para la investigación del

Siglo de Oro con métodos digitales, incluyendo el análisis cuantitativo (por ejemplo, Ehrlicher et al., 2020). Sin embargo, el desarrollo de nuevas preguntas de investigación ha generado la necesidad de un enriquecimiento del corpus más allá de la codificación original. A modo de ejemplo, este póster se centrará en la codificación TEI para el análisis de las redes sociales. Presentaremos la metodología que desarrollamos para realizar la segmentación de actos en distintas unidades para realizar el análisis y discutiremos la plusvalía de la codificación TEI.

TEI coding and network analysis: about Calderón Drama Corpus (CalDraCor) v.2.0

Hanno Ehrlicher¹, Antonio Rojas Castro¹, S. Padó², K. Jung², A. Keith²

¹University Eberhard Karl of Tübingen, ²University of Stuttgart (Alemania)

After being entirely published in open access in 2022 in DraCor (Fisher et al., 2019) under the name CalDraCor, the encoding of a corpus of 205 plays by Calderón de la Barca in TEI format has opened new paths for research on the Spanish Golden Age with digital methods, including quantitative analysis (e.g., Ehrlicher et al., 2020). However, the development of new research questions has led to a need for corpus enrichment beyond the original coding. By way of example, this poster will focus on TEI encoding for social network analysis. We will present the methodology we have followed to segment acts into different units in order to carry out the analysis, and we will discuss the added value of TEI encoding.

Enter the <stage>. A Review of the Encoding of Stage Directions within the TEI

Jan N. Jokisch

Max Planck Institute for Empirical Aesthetics (Germany)

Stage directions and other non-dialogical aspects of the dramatic texts have been overlooked historically and have received only minor scholarly attention. Thus, it comes as no surprise that the TEI's treatment of them is equally superficial. This is the source of diverse problems. On the one hand, this leads to certain not uncommon structures of drama to be impossible to encode within the TEI. For example, the blend of speech prefix and genuine stage direction in "Pendant ce temps, ADRIEN." has no obvious nor trivial markup within the TEI. On the other hand, the current guidelines tend to mislead practitioners, resulting in poor markup choices. Maybe the most aggravating example occurs whenever a dialog is introduced not with a speech prefix but with a stage

direction. In these cases, the dialog is habitually ignored as a speech or the stage direction is changed to conform to our idea of a classic speech prefix. These problems are augmented by the fact that the term “stage direction” lacks a stable equivalent in many languages. The poster explores "impossible" and unconventional stage directions from French, Spanish, and German plays from the 17th and 18th century as well as their encoding within the TEI(-adjacent) projects Théâtre Classique, EMOTHE, Deutsches Textarchiv, and TextGrid Repository. It aims at a productive criticism of the current TEI schema and guidelines and suggests ways to overcome existing shortcomings.

Entrar en <stage>. Una revisión de la codificación de las de la dirección escénica dentro del TEI

Jan N. Jokisch

Instituto Max Planck de Estética Empírica (Alemania)

La dirección escénica y otros aspectos no dialógicos de los textos dramáticos se han pasado por alto históricamente y sólo han recibido una atención académica menor. Por lo tanto, no sorprende que el tratamiento que les da la TEI sea igualmente superficial. Esta es la fuente de diversos problemas. Por un lado, conduce a que ciertas estructuras dramáticas no inusuales sean imposibles de codificar dentro de la TEI. Por ejemplo, la combinación de prefijo de voz y dirección escénica genuina en "Pendant ce temps, ADRIEN" no tiene un marcado obvio ni trivial dentro de la TEI. Por otro lado, las directrices actuales tienden a engañar a los profesionales, lo que da lugar a malas decisiones en cuanto al marcado. Quizás el ejemplo más inconveniente ocurre cuando no se introduce un diálogo con un prefijo de voz sino con una dirección escénica. En estos casos, habitualmente se ignora el diálogo como un discurso o se cambia la dirección escénica para ajustarse a nuestra idea de un prefijo de voz clásico. Estos problemas se ven agravados por el hecho de que el término “dirección escénica” carece de un equivalente en muchos idiomas. El poster explora direcciones escénicas "imposibles" y poco convencionales de obras de teatro francesas, españolas y alemanas de los siglos XVII y XVIII, así como su codificación dentro de los proyectos TEI (adyacentes): Théâtre Classique, EMOTHE, Deutsches Textarchiv y TextGrid Repository. Su objetivo es realizar una crítica productiva del esquema y de las directrices actuales de la TEI, y sugiere formas de superar las deficiencias existentes.

The Adaptive TEI Network: Antiracist, Decolonial, and Inclusive Markup Interventions

Sarah Revilla-Sanchez, D. Orizaga Doguim

University of British Columbia (Canada)

This poster presentation introduces the “PhD CoLab” project (University of British Columbia, 2024-26, with the collaboration of the SFU Digital Humanities Innovation Lab, DHIL) which brings together graduate students, faculty, and staff from various fields in humanities, languages, and literatures. While based in Vancouver, an English-speaking North American education system, this multifaceted project is concerned with the continuities and limitations of text encoding across languages (English, Spanish, German and Russian), geographic regions, and literary genres. This poster will provide concrete examples to illustrate the larger objectives of the PhD CoLab.

One of the encompassed projects is NovElla, which focuses on making visible and accessible short prose fiction written by early modern Spanish writers. It includes a catalog of annotated bibliographic resources to help promote future research by both students and scholars. Another example, related to Latin America, is Unión Cívica Project that focuses on the newspaper Unión Cívica published by the eponymous political movement founded in 1961 in the aftermath of the Rafael L. Trujillo dictatorship (1930-1961) in Dominican Republic. We will offer high resolution digital reproductions of 140 issues, with annotations, to provide political and historical context.

Furthermore, the very structure of the Adaptive TEI Network, rooted in a team-oriented ethos, disrupts the traditional mode of solitary, humanistic research. PhD students collaborate in a transdisciplinary team-based, project-oriented environment where we learn from and with one another while we propose a new TEI schema for text-encoding projects that consider antiracist, decolonial, inclusive and feminist markup practices.

In short, the TEI schema aims to address some of the projects’ research questions like: Can we adapt current TEI modules or does an antiracist/decolonial and feminist engagement with the literary text necessitate new TEI markup standards or new modules? Is the TEI also robust enough to address/function for multilingual texts?

La Red TEI Adaptativa: Intervenciones de marcado antirracistas, decoloniales e inclusivas

Sarah Revilla-Sánchez, D. Orizaga Doguim

Universidad de Columbia Británica (Canadá)

Este póster presenta el proyecto “PhD CoLab” (Universidad de Columbia Británica, 2024-26, con la colaboración del Laboratorio de Innovación en Humanidades Digitales

de SFU, DHIL) que reúne a estudiantes de posgrado, profesores y personal de diversos campos en humanidades e idiomas y literaturas. Si bien tiene su sede en Vancouver, un sistema educativo norteamericano de habla inglesa, este proyecto multifacético se ocupa de las continuidades y limitaciones de la codificación de textos en todos los idiomas (inglés, español, alemán y ruso), regiones geográficas y géneros literarios. Este póster proporcionará ejemplos concretos para ilustrar los objetivos más amplios del PhD CoLab.

Uno de los proyectos incluidos es NovElla, que se centra en hacer visible y accesible la ficción breve en prosa escrita por escritores españoles de la Edad Moderna. Incluye un catálogo de recursos bibliográficos comentados para ayudar a promover investigaciones futuras tanto por parte de estudiantes como de académicos. Otro ejemplo, relacionado con América Latina, es el Proyecto Unión Cívica que se centra en el periódico Unión Cívica publicado por el movimiento político homónimo fundado en 1961 tras la dictadura de Rafael L. Trujillo (1930-1961) en República Dominicana. Ofreceremos reproducciones digitales de 140 números en alta resolución con anotaciones para proporcionar un contexto político e histórico.

Asimismo, la estructura misma de Adaptive TEI Network, arraigada en un espíritu orientado al equipo, altera el modo tradicional de investigación humanista y solitaria. Los estudiantes de doctorado trabajan en un entorno transdisciplinario basado en equipos y orientado a proyectos donde aprendemos unos de otros mientras proponemos un nuevo esquema TEI para proyectos de codificación de texto que consideran prácticas de marcado antirracistas, decoloniales, inclusivas y feministas.

En resumen, el esquema TEI apunta a abordar algunas de las preguntas de investigación de los proyectos; a saber: ¿Podemos adaptar los módulos TEI actuales o se necesitan nuevos estándares de marcado TEI o nuevos módulos para un compromiso antirracista/decolonial y feminista con el texto literario? ¿Es la TEI también lo suficientemente sólida como para abordar/funcionar con textos multilingües?

Diccionario DCH: plataforma digital, fuentes y debates historiográficos en el derecho canónico hispanoamericano (s. XVI-XVIII)

Nadia dos Santos Aguiar

Instituto Max Planck de Historia del Derecho y Teoría del Derecho (Alemania)

El Proyecto del Diccionario Histórico de Derecho Canónico en Hispanoamérica y Filipinas (s.XVI -XVIII) (DCH), liderado desde el Instituto Max Planck de Historia y Teoría Derecho en Frankfurt (Alemania), tiene como propósito ofrecer a investigadores

de diversas disciplinas una herramienta fiable sobre las normatividades religiosas a partir de las fuentes primarias y actualizadas de acuerdo a la historiografía reciente. Así, a quienes trabajen con expedientes judiciales, obras de doctrina jurídica o canónica, crónicas, legislaciones reales y pontificias, concilios locales, manuales prácticos, costumbres locales, o documentos de archivos les ofrecemos un instrumento de consulta que contribuye a una mejor comprensión de los términos, transformaciones y producción local del derecho canónico en Hispanoamérica durante la edad moderna temprana.

El diccionario histórico, compuesto por 120 entradas (aproximadamente 4.000 páginas) y amplias referencias cruzadas, explora, a través de cada una de sus entradas, el surgimiento de regulaciones especiales e instituciones que se perfilaron de acuerdo a los diversos territorios y actores del Nuevo Mundo. Las primeras 80 entradas publicadas como Research Papers en OpenAccess, las fuentes primarias utilizadas, la estructura y algunos videos que hemos hecho, pueden ser visitadas en nuestro blog: <https://dch.hypotheses.org/>. Además de la versión impresa del Diccionario, que se espera lanzar en 2025, estamos desarrollando también una plataforma DCH, alojada en una aplicación de TEI Publisher, que se encuentra en su fase inicial.

La presentación del Diccionario DCH en TEI Publisher 2024 tiene como objetivo compartir y enriquecerse de la experiencia de los investigadores e investigadoras para que la herramienta sea lo más orgánica e intuitiva posible. Los comentarios técnicos sobre los procesos de modelización y el funcionamiento de TEI Publisher serán muy bienvenidos.

DCH Dictionary: digital platform, sources and historiographical debates in Hispanic American canon law (16th-18th centuries)

Nadia dos Santos Aguiar

Max Planck Institute for Legal History and Legal Theory (Germany)

The Historical Dictionary of Canon Law in Hispanic America and the Philippines (16th-18th centuries) Project (DCH), led by the Max Planck Institute for Legal History and Legal Theory in Frankfurt (Germany), aims to offer researchers from various disciplines a reliable tool on religious regulations based on primary sources and updated according to recent historiography. Thus, we offer those who work with court files, works of legal or canonical doctrine, chronicles, royal and pontifical legislation, local councils, practical handbooks, local customs, or archival documents, an instrument for consultation that contributes to a better understanding of the terms, transformations and local production of canon law in Hispanic America during the early modern age.

The historical dictionary, composed of 120 entries (approx. 4,000 pages) and extensive cross-references, explores, through each of its entries, the emergence of special regulations and institutions that were outlined according to the various territories and actors of the New World. The first 80 entries published as Research Papers in Open Access, the primary sources used, the structure and some videos made, can be visited in our blog: <https://dch.hypotheses.org/>. In addition to the printed version of the Dictionary, which is expected to be launched in 2025, we are also developing a DCH platform, hosted in a TEI Publisher application, which is in its initial phase.

The presentation of the DCH Dictionary in TEI Publisher 2024 aims to share and enrich from researchers' experiences so that the tool is as organic and intuitive as possible. Technical comments on the modelling processes and the operation of TEI Publisher are particularly welcome.

A Python Library for TEI Conversion into Edition Formats

Anastasiya Kostyanitsyna¹, Daniil Skorinkin²

¹Independent researcher, ²Digital Humanities Potsdam (Germany)

TEI-XML was designed to encode the structure and purpose of document parts, and not to represent them visually. Unlike HTML or DOCX, TEI is generally not tied to any particular software that would offer a standard way of rendering it. At the same time, TEI is not as easily interoperable as more popular data exchange formats like CSV or JSON, for which there are tools ranging from Excel to OpenRefine to programming libraries like Pandas. Developers used to dealing with JSON might get confused when confronted with the task of using TEI data.

All this creates the need for easy-to-use transition tools from TEI to these more common formats. Traditionally, this has been accomplished with XSLT stylesheets (Rahtz, 2006). However, this approach has limitations. As of 2024, XSLT is not the most widespread technology in the developer world and, what is more important, it is not part of most programming courses. For continuous support of TEI in the future, it seems crucial to produce tools that wrap TEI transformation into more widely-known data processing ecosystems than bare XSLT.

With that in mind, we have developed and published a TEI transformation library for Python (pypi.org/project/TEItransformer/). Currently the library performs conversion of TEI-XML into three formats: JSON, DOCX, and HTML. Each format is created by a separate class with its own settings. For the DOCX and HTML transformation, the user gets to choose a transformation scenario. The client interface for the user is

implemented by the TEITransformer class. Under the hood, the conversion is divided into three main steps: validation, transformation, and stylization. The library enables specification of an XML schema and the input of ODD or CSS files for the output customization. Since ODD is XML-based, the user does not need to have additional knowledge like XSLT to adjust the transformation.

Una biblioteca de Python para la conversión de TEI a formatos de edición

Anastasiya Kostyanitsyna¹, Daniil Skorinkin²

¹Investigadora independiente, ²Humanidades Digitales Potsdam (Alemania)

TEI-XML fue diseñado para codificar la estructura y el propósito de las partes del documento, y no para representarlas visualmente. A diferencia de HTML o DOCX, TEI generalmente no está vinculado a ningún software en particular que ofrezca una forma estándar de renderizarlo. Al mismo tiempo, TEI no es tan fácilmente interoperable como lo son los formatos de intercambio de datos más populares CSV o JSON, para los cuales existen herramientas que van desde Excel hasta OpenRefine y bibliotecas de programación tales como Pandas. Los desarrolladores acostumbrados a trabajar con JSON pueden confundirse cuando se enfrentan a la tarea de utilizar datos TEI.

Todo esto crea la necesidad de herramientas de transición fáciles de usar de TEI a estos formatos más comunes. Tradicionalmente, esto se ha logrado con hojas de estilo XSLT (Rahtz, 2006). Sin embargo, este enfoque tiene limitaciones. Desde 2024, XSLT ya no es la tecnología más extendida en el mundo de los desarrolladores y, lo que es más importante, no forma parte de la mayoría de cursos de programación. Para un soporte continuo de TEI en el futuro, parece fundamental producir herramientas que contengan la transformación de TEI dentro de ecosistemas de procesamiento de datos más conocidos que el XSLT básico.

En ese sentido, hemos desarrollado y publicado una biblioteca de transformación TEI para Python (pypi.org/project/TEItransformer/). Actualmente, la biblioteca realiza la conversión de TEI/XML a tres formatos: JSON, DOCX y HTML. Cada formato es creado por una clase separada con su propia configuración. Para la transformación DOCX y HTML, el usuario puede elegir un escenario de transformación. TEITransformer implementa la interfaz de cliente para el usuario. En esencia, la conversión se divide en tres pasos principales: validación, transformación y estilización. La biblioteca permite la especificación de un esquema XML y la entrada de archivos ODD o CSS para la personalización de la salida. Dado que ODD está basado en XML, el usuario no necesita tener conocimientos adicionales como XSLT para ajustar la transformación

Visualizing the Frankenstein Variorum

Elisa Beshero-Bondar¹, Raffaele Viglianti², Yuying Jin³

¹Penn State Erie, ²University of Maryland, ³Northeastern University (United States of America)

The Frankenstein Variorum team has completed work on a digital scholarly edition that compares five distinct versions of the novel Frankenstein. While we have presented this project at several conferences in recent years, we propose to share a new view of the project at TEI 2024: a visual summary of our publication method, and a visual survey of our collation data. Now that we have fully published the Frankenstein Variorum's TEI edition files, we are analyzing what we have learned about how much the novel changed over five distinct instantiations. These include:

- The 1816 manuscript notebook,
- The 1818 first anonymous publication,
- The Thomas copy's marginal handwritten revisions that were later lost,
- The 1823 edition produced by Mary Shelley's father,
- The 1831 edition, substantially revised.

For the TEI 2024 conference, we propose a poster to display two things: 1) our publishing process applying JavaScript-based static site generation to publish the edition, and 2) a "big picture" view of Frankenstein's changes over time drawn in Scalable Vector Graphics (SVG) from our TEI data. The poster will show how our project's TEI grounds the edition's interactive visualizations of the novel's transformation over time.

Our TEI standoff spine and edition files store the collation data of the novel's five versions, and the pipeline processing and algorithmic refinement of those files has been the subject of several past conference papers and presentations. Now that the edition is complete, the TEI data invite us to analyze the edition's moments of alignment, divergence, and gaps where material was missing or removed. For this poster and for our edition, we illuminate the Variorum interface design and visualize Frankenstein's transformations, each built directly from the TEI. If successful, our poster will welcome discussion of our publishing architecture and invite a holistic, nonlinear exploration of the digital variorum edition.

Visualizando el Frankenstein Variorum

Elisa Beshero-Bondar¹, Raffaele Viglianti², Yuying Jin³

¹Penn State Erie, ²Universidad de Maryland, ³Universidad del Noreste (Estados Unidos de América)

El equipo de Frankenstein Variorum ha completado el trabajo sobre una edición filológica digital que compara cinco versiones distintas de la novela Frankenstein. Si bien hemos presentado este proyecto en varias conferencias durante los últimos años, proponemos compartir una nueva visión del proyecto en TEI 2024: un resumen visual de nuestro método de publicación y un estudio visual de nuestra recopilación de datos. Ahora que hemos publicado completamente los archivos de la edición TEI del Frankenstein Variorum, estamos analizando lo aprendido respecto de cuánto ha cambiado la novela a lo largo de cinco instancias distintas. Estas incluyen:

- El cuaderno manuscrito de 1816,
- La primera publicación anónima de 1818,
- Las revisiones marginales manuscritas de la copia de Thomas que luego se perdieron,
- La edición de 1823 producida por el padre de Mary Shelley,
- La edición de 1831, sustancialmente revisada.

Para la conferencia TEI 2024, proponemos presentar un póster para mostrar dos cosas: 1) nuestro proceso de publicación al aplicar la generación de sitios estáticos basados en JavaScript para publicar la edición, y 2) un buen panorama de los cambios que Frankenstein ha sufrido a lo largo del tiempo trazados en formato SVG (Gráficos Vectoriales Escalables) de nuestros datos TEI. El póster mostrará cómo la TEI de nuestro proyecto se basa en las visualizaciones interactivas de la edición de la transformación de la novela a lo largo del tiempo.

Nuestros archivos de edición y columna vertebral de TEI almacenan los datos de recopilación de las cinco versiones de la novela, y el procesamiento de proyectos en desarrollo y el refinamiento algorítmico de esos archivos han sido el tema de varios artículos y presentaciones de conferencias anteriores. Ahora que la edición está completa, los datos en TEI nos invitan a analizar los momentos de alineación, divergencia y brechas de la edición en los que faltaba o se eliminaba material. Tanto para este póster como para nuestra edición, iluminamos el diseño de la interfaz Variorum y visualizamos las transformaciones de Frankenstein, cada una construida directamente desde la TEI. Si tiene éxito, nuestro póster dará la bienvenida a la discusión sobre nuestra arquitectura editorial e invitará a una exploración holística y no lineal de la edición variorum digital.

TEI Publisher 9: más allá de TEI y de la publicación / TEI Publisher 9: going beyond TEI and publication

Magdalena Turska¹, Helena Bermúdez Sabel²

¹e-editiones (Suiza), ²Jinntec (Alemania)

La filosofía de TEI Publisher gira en torno a la modularidad, reutilización y sostenibilidad gracias al uso de estándares. TEI Publisher nace con el objetivo de facilitar la producción de ediciones digitales, para que humanistas puedan crear productos académicos que responden a sus objetivos de investigación, con poca o ninguna programación. Esto se consigue gracias a un diseño modular que permite que las funcionalidades se puedan organizar y recombinar libremente. Esta concepción permite que personas con perfiles técnicos puedan hacer ajustes con facilidad, y que usuarios/as sin conocimientos de programación puedan aprovechar los valores predeterminados adaptados a diferentes modelos de edición.

TEI Publisher admite diferentes formatos, tanto de entrada como de salida. Además de TEI, TEI Publisher puede utilizarse para la publicación de documentos en otros estándares como DocBook, MS Word (DOCX) o JATS. Los documentos fuente no tienen que responder a un esquema específico, y serán fácilmente transformados a una variedad de formatos de salida para su publicación: desde una interfaz web, hasta un libro electrónico, un archivo PDF o su fuente LaTeX.

En las versiones más recientes de TEI Publisher se ha respondido a las demandas de la comunidad de usuarios/as que pedían ayuda para convertir e incorporar datos en diferentes formatos, así como para enriquecer la anotación de las fuentes sin tener que editar directamente el XML. TEI Publisher es pues más que una caja de herramientas para la publicación y explotación de documentos XML, convirtiéndose en un instrumento que permite la generación automática de documentos TEI y la anotación automatizada. Este póster presenta las características más importantes de TEI Publisher haciendo hincapié en las novedades introducidas en la versión 9.

TEI Publisher 9: going beyond TEI and publication

Magdalena Turska¹, Helena Bermúdez Sabel²

¹e-editiones (Switzerland), ²Jinntec (Germany)

TEI Publisher's philosophy is based on modularity, reusability and sustainability thanks to the use of standards. TEI Publisher was born with the aim of facilitating the production of digital editions so that humanists can create academic products that respond to their research objectives with little or no programming skills at all. This is achieved on account of a modular design that helps the functionalities to be organized and recombined freely. This conception enables people with technical profiles to make adjustments easily, and users lacking in programming knowledge to take advantage of the default values adapted to different editing models.

TEI Publisher supports both input and output formats. In addition to source documents in TEI, TEI Publisher can be used for publishing documents in other form of XML: DocBook, MS Word (DOCX) or JATS. Source documents do not have to respond to a specific schema, and will be easily transformed into a variety of output formats for publication: from a web interface to an e-book, a PDF file or LaTeX source.

Recent versions of the TEI Publisher have responded to the demands of the user community who were seeking help to incorporate and convert data into different formats, as well as to enrich the annotation of sources without having to edit the XML file directly. Therefore, TEI Publisher is more than a toolbox for the publication and exploitation of XML documents, it is an instrument that makes the automatic generation of TEI documents and automated annotation possible. This poster presentation outlines the most important features of TEI Publisher, highlighting the new features introduced in version 9.

The Revue des Colonies Scholarly Edition and Translation: a Distributed and Bilingual TEI Project

Maria Beliaeva Solomon¹, Grégory Pierrot², Raffaele Viglianti¹

¹University of Maryland, ²University of Connecticut (United States of America)

The Revue des Colonies Scholarly Edition and Translation project marks the first effort to digitally annotate and translate a landmark abolitionist periodical published in Paris between 1834 and 1842. Led by an international team of scholars, the project aims to make the Revue's invaluable store of journalistic and literary contents accessible to the public. This poster will outline the project's operations and workflows, as well as describing the TEI encoding strategies that we have adopted and plan to adopt.

The project's operation depends on a dynamic interplay between the encoding team and the editorial team, both geographically distributed and each with distinct roles and expertise contributing to the project's overarching goal.

The encoding team includes graduate students currently completing training modules in transcription, TEI encoding, and research in online databases.

The editorial team is composed of scholars covering a range of relevant disciplines who bring a deep understanding of the historical and cultural contexts in which the *Revue des Colonies* operated. These scholars compose meticulously researched annotations for the named entities identified in the text. Their annotations thus shed light on otherwise overlooked individuals, events, organizations, and historical documents, ensuring that the journal's original contributions to global discourse are recognized and contextualized within relevant scholarly fields.

Collaboration between these teams is facilitated by an online content management system that exports TEI data and adjusts the project customization ODD as content is added. The project's TEI customization focuses on the encoding of named entities to enable the creation of bilingual, substantial, and cross-navigable entries. Translation of the *Revue* itself is undertaken by members of the editorial team with significant experience in professional French–English translation, ensuring both the accessibility of the text to the widest audience and its fidelity to the rhetorical and stylistic features of the original.

Edición académica y traducción de la *Revue des Colonies*: un proyecto TEI distribuido y bilingüe

Maria Beliaeva Solomon¹, Grégory Pierrot², Raffaele Viglianti¹

¹Universidad de Maryland ²Universidad de Connecticut (Estados Unidos de América)

El proyecto de edición y traducción académica de la *Revue des Colonies* marca el primer esfuerzo por anotar y traducir digitalmente una publicación periódica abolicionista emblemática publicada en París de 1834 a 1842. Dirigido por un equipo internacional de académicos, el proyecto tiene como objetivo hacer que el invaluable acervo de material periodístico y contenidos literarios sea accesible al público. Este póster describirá las operaciones y los flujos de trabajo del proyecto, así como también las estrategias de codificación TEI que hemos adoptado y planeamos adoptar.

El funcionamiento del proyecto depende de una interacción dinámica entre el equipo de codificación y el equipo editorial, ambos distribuidos geográficamente y cada uno con funciones y experiencia distintas, que contribuyen al objetivo general del proyecto.

El equipo de codificación incluye estudiantes graduados que actualmente completan módulos de capacitación en transcripción, codificación TEI e investigación con bases de datos en línea.

El equipo editorial está compuesto por académicos que abarcan una variedad de disciplinas pertinentes y que aportan un profundo conocimiento de los contextos históricos y culturales en los que operaba la Revue des Colonies. Estos académicos redactan anotaciones meticulosamente investigadas para las entidades nombradas identificadas en el texto. Por lo tanto, sus anotaciones arrojan luz sobre personas, eventos, organizaciones y documentos históricos que, de otro modo, se pasarían por alto, y aseguran que las contribuciones originales de la revista al discurso global sean reconocidas y contextualizadas dentro de campos académicos relevantes.

La colaboración entre estos equipos se ve facilitada por un sistema de gestión de contenido en línea que exporta datos de TEI y ajusta el ODD de personalización del proyecto a medida que se agrega contenido. La personalización TEI del proyecto se centra en la codificación de entidades nombradas para permitir la creación de entradas bilingües, sustanciales y de navegación cruzada. La traducción de la Revue está a cargo de miembros del equipo editorial con importante experiencia en traducción profesional francés-inglés, lo que garantiza tanto la accesibilidad del texto para el público más amplio como su fidelidad a las características retóricas y estilísticas del original.

Keynote Conferences / Conferencias plenarias

TEI para las comunidades de Humanidades Digitales: edición digital, análisis computacional, escritura científica

Ulrike Henny-Krahmer

Universidad de Rostock (Alemania)

Las guías directrices de la TEI proporcionan a la comunidad científica interesada en las humanidades digitales un estándar de datos que puede utilizarse para representar una amplia variedad de recursos digitales que, con ayuda del marcado XML, se centra en los recursos textuales que tradicionalmente se procesan y utilizan en las humanidades, como manuscritos, impresos históricos, textos literarios, corpus lingüísticos y diccionarios. Los distintos capítulos de las guías directrices de la TEI reflejan las diferentes áreas de aplicación y perspectivas sobre el texto que pueden adoptarse. Además de esto, también hay diferentes comunidades dentro de las humanidades digitales que investigan los textos digitales de múltiples maneras.

En mi propio trabajo como humanista digital, he encontrado o utilizado TEI junto a varias de estas subcomunidades: la edición académica digital, el análisis de textos literarios asistido por ordenador y las reseñas académicas de recursos digitales. En la ponencia, me gustaría presentar proyectos basados en la TEI de parte de estas tres subcomunidades: la edición digital Fernando Pessoa, mis proyectos y publicaciones, el análisis de contenido y estilo de novelas hispanoamericanas del siglo XIX y la revista de reseñas RIDE.

Me gustaría animar a la reflexión sobre cómo podemos utilizar la TEI para nuestro trabajo específico en las comunidades de texto individuales y, al mismo tiempo, mantener un ojo en el panorama general: ¿cómo pueden los resultados de nuestro trabajo ser beneficiosos para varias comunidades dentro de las Humanidades Digitales o para las Humanidades Digitales en general? La TEI, como estándar, ofrece muy buenas condiciones para ello.

TEI for the Digital Humanities communities: digital edition, computational analysis, and scientific writing

Ulrike Henny-Krahmer

University of Rostock (Germany)

The TEI guidelines provide the Digital Humanities community with data standards that can be used to represent a wide variety of digital resources by means of XML markup: they focus on textual resources traditionally processed and used in the humanities, such as manuscripts, historical documents, literary texts, linguistic corpora and dictionaries. The different chapters of the TEI guidelines show diverse areas of application and perspectives concerning the text which can be adopted. In addition to this, there are different communities within the digital humanities that investigate certain digital texts in particular ways.

In my own work as a digital humanist, I have found or used the TEI in several of these subcommunities: digital academic publishing, computer-aided analysis of literary texts, and academic reviews of digital resources. In my presentation, I would like to talk about projects based on the use of the TEI from these three sub-communities: the digital scholarly edition Fernando Pessoa. Projects and Publications, the content and style analysis of 19th century Latin American novels, and the RIDE journal.

I would like to promote reflection on the way we can use TEI for our specific work in individual text communities and, at the same time, keep an eye on the big picture: how

can the results of our work be beneficial to various communities within the Digital Humanities or to Digital Humanities in general? TEI, as a standard, offers very good conditions for this.

Comisión Local Organizadora / Local Organizing Committee

Gimena del Rio Riande (IIBICRIT, CONICET-Universidad del Salvador)

Alicia Martin (Universidad del Salvador)

Claudia Pelossi (Universidad del Salvador)

Verónica Parselis (Universidad del Salvador)

Mariana Bricchi (Universidad del Salvador)

Mercedes Rodríguez Temperley (IIBICRIT, CONICET-Universidad Nacional de Lomas de Zamora)

Laura Volkind (IIBICRIT, CONICET)

Gabriel Calarco (Universidad de Buenos Aires)

Romina De León (IIBICRIT, CONICET)

Agustina Ryckeboer (Universidad del Salvador)

Melisa Marti (Universidad de Buenos Aires)

Pamela Gionco (Universidad de Buenos Aires-Biblioteca Nacional Mariano Moreno)

Rocío Méndez (Universidad de Buenos Aires)

Verónica Zumárraga (Fundación Ortega y Gasset)

Nicolás Lázaro (Universidad Católica de Santa Fe)

Carlos Nusch (Universidad Nacional de La Plata-CONICET)

Comité de Programa / Program Committee

Gustavo Fernández Riva (Universidad de Heidelberg, Alemania)

Marta López Izquierdo (Paris 3, France)

Martina Scholger (University of Graz, Austria)

Nidia Hernández (IMICIHU-CONICET, Argentina)

Susanna Allés (University of Miami, USA)

Helena Bermúdez Sabel (exist Solutions, Switzerland)

Magdalena Turska (exist Solutions, Switzerland)

Miriam Peña Pimentel (Universidad Nacional Autónoma de México, Mexico)

Clara Martínez Cantón (Universidad Nacional de Educación a Distancia, Spain)

Sandra Cristina de Jesús Boto (Universidade Nova de Lisboa, Portugal)

Martha Thomae (McGill University, Canada)

Anna E. Kijas (Tufts University, USA)

Frederike Neuber (Berlin-Brandenburgische Akademie der Wissenschaften, Germany)

Diane Jakacki (Bucknell University, USA)

Kathryn Tomasek (Wheaton College, USA)
Elisa Beshero Bondar (Behrend College, USA)
Connie Crompton (University of Ottawa, Canada)
Ana Salgado (Universidade Nova de Lisboa, Portugal)
Janelle Jenstad (University of Victoria, Canada)
Katarzyna Anna Kapitan (University of Oxford, UK)
Sandra Cristina de Jesús Boto (Universidade Nova de Lisboa, Portugal)
Pietro Liuzzo (Biblioteca Hertziana, Italy)
Jan Jokisch (Max Planck Institute, Germany)
Nicolás Quiroga (Universidad Nacional de Mar del Plata, Argentina)
Antonio Rojas Castro (Berlin-Brandenburgische Akademie der Wissenschaften/UNIR, Germany-Spain)
Ernesto Piani (Universidad Nacional Autónoma de México, Mexico)
José Calvo Tello (Universidad de Göttingen, Germany)
James Cummings (Newcastle University, UK)
James O'Sullivan (University of Cork, Ireland)
Raff Vigilanti (University of Maryland, USA)
Clayton McCarl (University of Florida, USA)
Cristian Suárez (Universidad EAFIT, Colombia)
Kiyonori Nagasaki (University of Tokyo, Japan)
Franz Fischer (Universidad Venezia C'a Foscari, Italy)
Frank Fischer (University of Berlin, Germany)
Alejandro Bia Plata (Universidad Miguel Hernández, Spain)
Hugh Cayless (University of Duke, USA)
Andreas Wagner (Max Planck Institute, Germany)



Text
Encoding
Initiative



USAL
UNIVERSIDAD
DEL SALVADOR

Ciencia a la mente y virtud al corazón