Dispro učebnica

Abstract

Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world. Whereas disregard and contempt for human rights have resulted in barbarous acts which have outraged the conscience of mankind, and the advent of a world in which human beings shall enjoy freedom of speech and belief and freedom from fear and want has been proclaimed as the highest aspiration of the common people. Whereas it is essential, if man is not to be compelled to have recourse, as a last resort, to rebellion against tyranny and oppression, that human rights should be protected by the rule of law. Whereas it is essential to promote the development of friendly relations between nations. Whereas the peoples of the United Nations have in the Charter reaffirmed their faith in fundamental human rights, in the dignity and worth of the human person and in the equal rights of men and women and have determined to promote social progress and better standards of life in larger freedom. Whereas Member States have pledged themselves to achieve, in co-operation with the United Nations, the promotion of universal respect for and observance of human rights and fundamental freedoms. Whereas a common understanding of these rights and freedoms is of the greatest importance for the full realization of this pledge. Now, therefore, The General Assembly, proclaims this Universal Declaration of Human Rights as a common standard of achievement for all peoples and all nations, to the end that every individual and every organ of society, keeping this Declaration constantly in mind, shall strive by teaching and education to promote respect for these rights and freedoms and by progressive measures, national and international, to secure their universal and effective recognition and observance, both among the peoples of Member States themselves and among the peoples of territories under their jurisdiction.

Obsah

Úvod do digitálnych humanitných vied

Definícia DH

Aliquam erat volutpat. Nunc eleifend leo vitae magna. In id erat non orci commodo lobortis. Proin neque massa, cursus ut, gravida ut, lobortis eget, lacus. Sed diam. Praesent fermentum tempor tellus. Nullam tempus. Mauris ac felis vel velit tristique imperdiet. Donec at pede. Etiam vel neque nec dui dignissim bibendum. Vivamus id enim. Phasellus neque orci, porta a, aliquet quis, semper a, massa. Phasellus purus. Pellentesque tristique imperdiet tortor. Nam euismod tellus id erat.

Aliquam erat volutpat. Nunc eleifend leo vitae magna. In id erat non orci commodo lobortis. Proin neque massa, cursus ut, gravida ut, lobortis eget, lacus. Sed diam. Praesent fermentum tempor tellus. Nullam tempus. Mauris ac felis vel velit tristique imperdiet. Donec at pede. Etiam vel neque nec dui dignissim bibendum. Vivamus id enim. Phasellus neque orci, porta a, aliquet quis, semper a, massa. Phasellus purus. Pellentesque tristique imperdiet tortor. Nam euismod tellus id erat.

História DH

Aliquam erat volutpat. Nunc eleifend leo vitae magna. In id erat non orci commodo lobortis. Proin neque massa, cursus ut, gravida ut, lobortis eget, lacus. Sed diam. Praesent fermentum tempor tellus. Nullam tempus. Mauris ac felis vel velit tristique imperdiet. Donec at pede. Etiam vel neque nec dui dignissim bibendum. Vivamus id enim. Phasellus neque orci, porta a, aliquet quis, semper a, massa. Phasellus purus. Pellentesque tristique imperdiet tortor. Nam euismod tellus id erat.

Aliquam erat volutpat. Nunc eleifend leo vitae magna. In id erat non orci commodo lobortis. Proin neque massa, cursus ut, gravida ut, lobortis eget, lacus. Sed diam. Praesent fermentum tempor tellus. Nullam tempus. Mauris ac felis vel velit tristique imperdiet. Donec at pede. Etiam vel neque nec dui dignissim bibendum. Vivamus id enim. Phasellus neque orci, porta a, aliquet quis, semper a, massa. Phasellus purus. Pellentesque tristique imperdiet tortor. Nam euismod tellus id erat.

Vzťah DH a tradičných humanitných vied

Ústredné teórie a pojmy

Nullam eu ante vel est convallis dignissim. Fusce suscipit, wisi nec facilisis facilisis, est dui fermentum leo, quis tempor ligula erat quis odio. Nunc porta vulputate tellus. Nunc rutrum turpis sed pede. Sed bibendum. Aliquam posuere. Nunc aliquet, augue nec adipiscing interdum, lacus tellus malesuada massa, quis varius mi purus non odio. Pellentesque condimentum, magna ut suscipit hendrerit, ipsum augue ornare nulla, non luctus diam neque sit amet urna. Curabitur vulputate vestibulum lorem. Fusce sagittis, libero non molestie mollis, magna orci ultrices dolor, at vulputate neque nulla lacinia eros. Sed id ligula quis est convallis tempor. Curabitur lacinia pulvinar nibh. Nam a sapien.

Interdisciplinarita v DH

Aliquam erat volutpat. Nunc eleifend leo vitae magna. In id erat non orci commodo lobortis. Proin neque massa, cursus ut, gravida ut, lobortis eget, lacus. Sed diam. Praesent fermentum tempor tellus. Nullam tempus. Mauris ac felis vel velit tristique imperdiet. Donec at pede. Etiam vel neque nec dui dignissim bibendum. Vivamus id enim. Phasellus neque orci, porta a, aliquet quis, semper a, massa. Phasellus purus. Pellentesque tristique imperdiet tortor. Nam euismod tellus id erat.

Pellentesque dapibus suscipit ligula. Donec posuere augue in quam. Etiam vel tortor sodales tellus ultricies commodo. Suspendisse potenti. Aenean in sem ac leo mollis blandit. Donec neque quam, dignissim in, mollis nec, sagittis eu, wisi. Phasellus lacus. Etiam laoreet quam sed arcu. Phasellus at dui in ligula mollis ultricies. Integer placerat tristique nisl. Praesent augue. Fusce commodo. Vestibulum convallis, lorem a tempus semper, dui dui euismod elit, vitae placerat urna tortor vitae lacus. Nullam libero mauris, consequat quis, varius et, dictum id, arcu. Mauris mollis tincidunt felis. Aliquam feugiat tellus ut neque. Nulla facilisis, risus a rhoncus fermentum, tellus tellus lacinia purus, et dictum nunc justo sit amet elit.

Humanitné disciplíny (literatúra, história, filozofia, umenie, lingvistika)

Aliquam erat volutpat. Nunc eleifend leo vitae magna. In id erat non orci commodo lobortis. Proin neque massa, cursus ut, gravida ut, lobortis eget, lacus. Sed diam. Praesent fermentum tempor tellus. Nullam tempus. Mauris ac felis vel velit tristique imperdiet. Donec at pede. Etiam vel neque nec dui dignissim bibendum. Vivamus id enim. Phasellus neque orci, porta a, aliquet quis, semper a, massa. Phasellus purus. Pellentesque tristique imperdiet tortor. Nam euismod tellus id erat.

Výpočtové metódy (programovanie, dátové modelovanie, strojové učenie)

Pellentesque dapibus suscipit ligula. Donec posuere augue in quam. Etiam vel tortor sodales tellus ultricies commodo. Suspendisse potenti. Aenean in sem ac leo mollis blandit. Donec neque quam, dignissim in, mollis nec, sagittis eu, wisi. Phasellus lacus. Etiam laoreet quam sed arcu. Phasellus at dui in ligula mollis ultricies. Integer placerat tristique nisl. Praesent augue. Fusce commodo. Vestibulum convallis, lorem a tempus semper, dui dui euismod elit, vitae placerat urna tortor vitae lacus. Nullam libero mauris, consequat quis, varius et, dictum id, arcu. Mauris mollis tincidunt felis. Aliquam feugiat tellus ut neque. Nulla facilisis, risus a rhoncus fermentum, tellus tellus lacinia purus, et dictum nunc justo sit amet elit.

Pellentesque dapibus suscipit ligula. Donec posuere augue in quam. Etiam vel tortor sodales tellus ultricies commodo. Suspendisse potenti. Aenean in sem ac leo mollis blandit. Donec neque quam, dignissim in, mollis nec, sagittis eu, wisi. Phasellus lacus. Etiam laoreet quam sed arcu. Phasellus at dui in ligula mollis ultricies. Integer placerat tristique nisl. Praesent augue. Fusce commodo. Vestibulum convallis, lorem a tempus semper, dui dui euismod elit, vitae placerat urna tortor vitae lacus. Nullam libero mauris, consequat quis, varius et, dictum id, arcu. Mauris mollis tincidunt felis. Aliquam feugiat tellus ut neque. Nulla facilisis, risus a rhoncus fermentum, tellus tellus lacinia purus, et dictum nunc justo sit amet elit.

Nullam eu ante vel est convallis dignissim. Fusce suscipit, wisi nec facilisis facilisis, est dui fermentum leo, quis tempor ligula erat quis odio. Nunc porta vulputate tellus. Nunc rutrum turpis sed pede. Sed bibendum. Aliquam posuere. Nunc aliquet, augue nec adipiscing interdum, lacus tellus malesuada massa, quis varius mi purus non odio. Pellentesque condimentum, magna ut suscipit hendrerit, ipsum augue ornare nulla, non luctus diam neque sit amet urna. Curabitur vulputate vestibulum lorem. Fusce sagittis, libero non molestie mollis, magna orci ultrices dolor, at vulputate neque nulla lacinia eros. Sed id ligula quis est convallis tempor. Curabitur lacinia pulvinar nibh. Nam a sapien.

Knižničné a informačné vedy (metadáta, digitálne archivovanie)

Aliquam erat volutpat. Nunc eleifend leo vitae magna. In id erat non orci commodo lobortis. Proin neque massa, cursus ut, gravida ut, lobortis eget, lacus. Sed diam. Praesent fermentum tempor tellus. Nullam tempus. Mauris ac felis vel velit tristique imperdiet. Donec at pede. Etiam vel neque nec dui dignissim bibendum. Vivamus id enim. Phasellus neque orci, porta a, aliquet quis, semper a, massa. Phasellus purus. Pellentesque tristique imperdiet tortor. Nam euismod tellus id erat.

Aliquam erat volutpat. Nunc eleifend leo vitae magna. In id erat non orci commodo lobortis. Proin neque massa, cursus ut, gravida ut, lobortis eget, lacus. Sed diam. Praesent fermentum tempor tellus. Nullam tempus. Mauris ac felis vel velit tristique imperdiet. Donec at pede. Etiam vel neque nec dui dignissim bibendum. Vivamus id enim. Phasellus neque orci, porta a, aliquet quis, semper a, massa. Phasellus purus. Pellentesque tristique imperdiet tortor. Nam euismod tellus id erat.

Sociálne vedy

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Donec hendrerit tempor tellus. Donec pretium posuere tellus. Proin quam nisl, tincidunt et, mattis eget, convallis nec, purus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nulla posuere. Donec vitae dolor. Nullam tristique diam non turpis. Cras placerat accumsan nulla. Nullam rutrum. Nam vestibulum accumsan nisl.

Digitálna zbierka slovenskej prózy (prípadová štúdia)

V tejto kapitole opisujeme vývoj digitálneho korpusu slovenskej prózy vydanej pred rokom 1950, pričom osobitnú pozornosť v nej venujeme prieniku literárnej vedy a prostriedkov výpočtových technológií. Cieľom projektu, bolo zostaviť reprezentatívnu a na výskum pripravenú zbierku beletristickej prózy, ktorá by odrážala formálny, tematický a jazykový vývoj slovenskej literatúry od polovice 19. storočia do začiatku povojnového obdobia. Korpus, ktorý čerpal z rôznych archívnych zdrojov - vrátane Slovenskej národnej knižnice, univerzitných repozitárov, jazykovedného inštitútu SAV a historických vydavateľských záznamov - pozostáva z desiatok XML súborov zodpovedajúcich schéme TEI [1] podporujúcej prístupy blízkeho aj vzdialeného čítania. Tento štruktúrovaný súbor údajov umožňuje celý rad vedeckých výskumov, od štylistickej analýzy a klasifikácie žánrov až po sieťové modelovanie autorských a publikačných kotextov. Namiesto vytvárania uzavretého kánonu projekt poskytuje otvorenú platformu na skúmanie literárnej histórie prostredníctvom počítačových nástrojov, pričom zostáva zakotvený v interpretačných tradíciách humanitných vied.

Opis výskumného projektu

Použité technológie

Linux

Dôležitou, ale často opomínanou zložkou pracovného postupu tvorby korpusu, bolo použitie operačného systému Linux ako technologického základu projektu. Linux poskytoval

[1] Text Encoding Initiative

stabilné prostredie s otvoreným zdrojovým kódom, ktoré sa ideálne hodilo na požiadavky rozsiahleho spracovania textu, kontroly verzií a automatizácie. Jeho kompatibilita so základnými nástrojmi - ako sú knižnice na spracovanie TEI, XML a skriptovacími jazykmi, ako sú Python a Bash - nám umožnil výskumnému tímu vytvoriť vlastné postupy na čistenie údajov, kódovanie a správu korpusu. Okrem toho modulárna konštrukcia systému Linux umožnila jemnú kontrolu nad systémom správania, od oprávnení súborov až po plánovanie úloh, čo sa ukázalo ako nevyhnutné, keď pri práci so súbormi údajov archívneho rozsahu. V tejto časti sa uvádza, ako systém Linux podporoval technickú infraštruktúru projektu, pričom sa zdôrazňuje jeho úloha pri zabezpečovaní transparentnosti, reprodukovateľnosti a dlhodobej udržiavateľnosti - hodnôt, ktoré zdieľa digitálnych humanitných vied a komunitách open-source.

Programovacie jazyky

Medzi základné technológie použité v projekte budovania korpusu patria programovacie jazyky Python a Lua, ktoré zohrávali odlišné, ale vzájomne sa dopĺňajúce úlohy. Python slúžil ako primárny jazyk na manipuláciu s údajmi, spracovanie textu, a integráciu s knižnicami na spracovanie TEI-XML, parsovanie regulárnych výrazov, a transformáciu metadát. Jeho čitateľnosť, všestrannosť a rozsiahly ekosystém sa hodil na vytváranie robustných skriptov na automatizáciu úloh, ako sú čistenie OCR, overovanie štrukturálnych značiek a štatistické analýzy. Jazyk Lua sme používali predovšetkým na vývoj vlastných filtrov a zapisovačov pre Pandoc, čo umožnilo jemnú kontrolu nad konverziou dokumentov najmä na generovanie konzistentných výstupov z textov zakódovaných v TEI do formátov ako HTML, Markdown alebo LaTeX. Táto skriptovacia vrstva umožnila tímu prispôsobiť transformáciu zložitých štruktúr XML do použiteľných formátov na vedeckú prezentáciu aj výpočtovú analýzu. Táto časť skúma, ako Python a Lua prispeli k modulárnemu, reprodukovateľnému projektu. pracovných postupov, čím sa posilňuje hodnota ľahkého, účelovo vytvoreného skriptovania v digitálnej humanitnej infraštruktúry.

Among the core technologies employed in the corpus-building project, the programming languages Python and Lua played distinct yet complementary roles. Python served as the primary language for data manipulation, text processing, and integration with libraries for TEI-XML handling, regular expression parsing, and metadata transformation. Its readability, versatility, and extensive ecosystem made it well-suited for building robust scripts to automate tasks such as OCR cleanup, structural markup verification, and corpus-wide statistical analysis. Lua, on the other hand, was primarily used to develop custom filters and writers for Pandoc, enabling fine-grained control over document conversion pipelines—particularly for generating consistent outputs from TEI-encoded texts to formats such as HTML, Markdown, or LaTeX. This scripting layer allowed the team to tailor the transformation of complex XML structures into usable formats for both scholarly presentation and computational analysis. This section explores how Python and Lua contributed to the project's modular, reproducible workflow, reinforcing the value of lightweight, purpose-built scripting in digital humanities infrastructure.

In digital humanities projects, programming is less about building complex software and more about designing flexible tools to help explore, transform, and interpret data. Two languages especially useful in this context are Python and Lua—each with its own strengths and roles within a DH workflow.

Python is one of the most widely used languages in the digital humanities due to its readability, extensive libraries, and active community. It's especially well-suited for tasks such as cleaning text data, analyzing word frequencies, converting file formats, or querying metadata. For example, using libraries like lxml or BeautifulSoup, students can extract information from XML or HTML documents, while tools like pandas allow for powerful data manipulation and statistical summaries with just a few lines of code. Python is ideal for building repeatable, modular scripts that can be shared and reused across projects.

Lua, by contrast, is a lightweight scripting language often embedded within other tools. In DH contexts, it shines when used for customizing workflows inside software like Pandoc—a universal document converter that plays a key role in many text transformation pipelines. With Lua, students can write compact filters that modify how documents are converted, such as reformatting chapter titles, removing footnotes, or extracting specific TEI elements before exporting to HTML or PDF. Lua's simplicity makes it easy to learn for specific, focused tasks, especially when working within structured publishing systems.

Together, Python and Lua offer a powerful toolkit: Python for data processing and analysis, Lua for document transformation and customization. Mastery of even basic scripts in these languages can significantly extend what's possible in digital humanities research, bridging the gap between traditional scholarship and computational methods.

XML

XML (eXtensible Markup Language) je jazyk navrhnutý na reprezentáciu informácií v štruktúrovanom, pre človeka a stroj čitateľnom formáte. Pri jeho návrhu sa kládol dôraz najmä na jednoduchosť, všeobecnosť a použiteľnosť v prostredí internetu[2] a vyznačuje sa silnou podporou takmer všektých ľudských jazykov vďaka kompatibilite s Unicode štandardom.[3] Hoci mal jazyk XML pôvodne slúžiť najmä na reprezentáciu dokumentov, v súčasnosti sa extenzívne používa na reprezentáciu ľubovoľných dátových štruktúr,[4] napríklad tých, ktoré sa vyskytujú vo webových službách.

Pre digitálnych humanistov je XML viac ako len technický nástroj - je to metóda na vyjadrenie významu, štruktúry a vzťahov v texte konzistentným a transparentným spôsobom.

At its core, XML (eXtensible Markup Language) is a way to represent information in a structured, human- and machine-readable format. For digital humanists, XML is more than just a technical tool—it's a method for expressing the meaning, structure, and relationships within a text in a consistent and transparent way.

Unlike word processors that focus on how text looks, XML focuses on how text is organized and interpreted. It allows scholars to mark up a text using custom, descriptive tags that reflect the text's internal structure. These tags are enclosed in angle brackets and come in pairs—for example:

```
<title>The Cross and the Sword</title>
```

This line tells both the human reader and the computer that "The Cross and the Sword" is the title of a work. Tags can represent a wide range of elements, such as:

```
<author>Ľudmila Podjavorinská</author>
<date when="1910">1910</date>
<placeName>Martin</placeName>
<quote>"Freedom must live in the heart before it lives on paper."</quote>
```

[2] "Extensible Markup Language (XML) 1.0 (Fifth Edition)".

[3] Ide o univerzálne kódovanie znakov určené na podporu celosvetovej výmeny, spracovania a zobrazovania písaných textov rôznych jazykov a technických disciplín moderného sveta. ("Unicode Standard")

[4] Fennell, "Extremes of XML".

Each of these elements conveys semantic meaning—not just formatting. You can also nest elements to reflect more complex relationships, like a paragraph that contains a name and a date:

```
<p>In <date when="1923">1923</date>,
<name>Jozef Cíger Hronský</name>
published his second novel.</p>
```

Every well-formed XML document consists of the following basic parts:

- **Elements**: Named sections of content wrapped in opening and closing tags (
  …
).
- **Attributes**: Extra information about an element, written inside the opening tag (1923).
- **Hierarchy**: XML is structured like a tree, with elements nested inside others to show relationships and structure.

A minimal XML document might look like this:

```
<?xml version="1.0" encoding="UTF-8"?>
<novel>
<title>Jarné vody</title>
<author>Božena Slančíková-Timrava</author>
<date when="1914">1914</date>
<text>
<p>Keď sa vrátil z vojny, všetko sa zdalo byť rovnaké, a predsa iné.</p>
</text>
</novel>
```

In digital humanities, XML—especially when guided by frameworks like the Text Encoding Initiative (TEI)—enables researchers to encode literary texts with a level of detail and care that reflects the richness of the material itself. This structured markup makes texts not only easier to preserve, but also searchable, analyzable, and convertible into other formats such as HTML, PDF, or plain text for visualization or presentation.

By learning XML, students of the humanities gain the ability to bridge traditional textual scholarship with digital tools—making their research more sustainable, collaborative, and computationally powerful.

Introducing TEI: A Shared Language for Encoding Texts

Once students understand the basics of XML, the next step in many digital humanities projects—especially those involving historical or literary texts—is learning how to apply those principles consistently and meaningfully. This is where the Text Encoding Initiative (TEI) comes in.

TEI is an international standard for encoding texts in XML, developed by and for scholars in the humanities. Its guidelines provide a shared vocabulary and structure for representing everything from prose and poetry to letters, plays, critical editions, and historical documents. Rather than inventing their own XML tags for each project, researchers can rely on TEI's rich and well-documented set of elements, which ensures interoperability, clarity, and long-term preservation.

For example, a simple TEI-encoded excerpt from a novel might look like this:

Dom v stráni

```
    <author>Martin Kukučín</author>
  </titleStmt>
  <publicationStmt>
    <publisher>Slovenská akadémia vied</publisher>
    <date when="1904">1904</date>
  </publicationStmt>
  <sourceDesc>
    <bibl>Original print edition from 1904</bibl>
  </sourceDesc>
</fileDesc>
```

Bol to dom, akých bolo v dedine málo – biely, s oblôčikmi plnými kvetov.

In this snippet, we see several key features of TEI in action:

```
The <teiHeader> provides essential metadata about the document: its title, author, source

The <text> element holds the actual content, typically structured into <body>, <div> (for

TEI supports a wide range of additional elements, such as <name>, <placeName>, <persName>
```

TEI is designed to be flexible and extensible. Projects can select only the elements they need, or even define custom rules using ODD (One Document Does it all) specifications, which describe how a given TEI customization should behave. This adaptability makes TEI useful for everything from minimalist digital editions to deeply annotated scholarly corpora.

For students and researchers in the digital humanities, TEI is more than just a markup standard—it's a framework for thinking critically about the structure and meaning of texts. Encoding with TEI encourages close reading, editorial reflection, and a heightened awareness of textual variation, paratexts, and publication history. At the same time, it prepares those texts for computational analysis, digital presentation, and long-term preservation.

Getting Started with TEI Encoding: Tools and Tips Starting a TEI-based project doesn't require an advanced technical background—just curiosity, patience, and some basic tools. Here are a few ways students can begin:

- Start small. Pick a short text (e.g., a chapter, a short story, or a single letter) and try encoding just the title, author, paragraphs, and chapter divisions.
- Use a TEI-aware XML editor. Tools like oXygen XML Editor (paid, but with academic licenses), Sublime Text with plugins, or the free TEI Publisher's web-based editor offer validation, autocomplete, and syntax highlighting.
- Follow the Guidelines. The TEI Guidelines are searchable and rich with examples. They're your best companion for understanding what each element means and how to use it.
- Validate your files. Always check your TEI files for well-formedness and against your project's schema—either within an editor like oXygen or using online tools like Roma (TEI's schema builder).

- Join the community. The TEI community is friendly and active. Mailing lists, GitHub repositories, and workshops provide opportunities to ask questions, share schemas, and learn from others.

Whether you're preparing a digital edition, building a corpus, or experimenting with literary analysis, customizing TEI to your needs is a valuable scholarly exercise. It asks you to consider what matters in a text—not just to you, but to future readers and machines—and how to model that meaning clearly and sustainably.

Webové technológie

Transforming a structured digital corpus into an accessible, user-friendly web resource requires more than just solid encoding—it demands thoughtful design and the right technical stack. In this project, a combination of standard web technologies—HTML, JavaScript, and CSS—was used to build a responsive, readable interface for interacting with the corpus. HTML served as the backbone for content display, CSS ensured typographic and structural clarity, and JavaScript enabled interactive features such as search, filtering, and basic visualizations. To bridge the gap between TEI-encoded data and the web presentation layer, the project made extensive use of TEI Publisher, an open-source framework designed specifically for publishing TEI documents online. TEI Publisher provided out-of-the-box support for rendering TEI-XML, managing search indexes, and offering faceted browsing, while also allowing for deep customization through XSLT, CSS, and optional integration with JavaScript frameworks like Vue.js. This section explores how TEI Publisher, combined with modern web technologies, enabled the creation of a platform that is both technically robust and accessible to a broad range of users, from researchers to general readers.

Digitization: From Printed Page to Machine-Readable Text

The foundation of the corpus-building process began with the digitization of physical novels, many of which existed only in aging print editions or archival microfilm. This phase involved careful selection of source materials based on availability, historical significance, and condition, followed by high-resolution scanning and optical character recognition (OCR).

While OCR technologies offer substantial time savings, the process also revealed the limitations of automated text capture when applied to older Slovak orthographies, non-standard typography, or damaged pages. As a result, post-OCR correction—both automated and manual—became a key component of the digitization workflow. This section outlines the practical and methodological considerations that shaped the transition from analog texts to machine-readable data, including the tools, standards, and quality control measures employed to ensure that the digital texts would be suitable for subsequent encoding and analysis.

Encoding: Structuring the Text with TEI

Once the novels were digitized and cleaned, the next step involved enriching the plain text with semantic and structural markup using the Text Encoding Initiative (TEI) guidelines. This phase was central to transforming the corpus into a scholarly resource that could support both humanistic inquiry and computational analysis. TEI encoding allowed for detailed representation of textual features such as chapter divisions, narrative perspective shifts, named entities, quotations, and paratextual elements (e.g., prefaces, footnotes). It also facilitated the inclusion of bibliographic metadata, authorial information, and historical publication context. Balancing descriptive accuracy with encoding efficiency required the development of project-specific schemas and tagging conventions, as well as the use of both automated tagging scripts and manual interventions. This section delves into the rationale behind the encoding strategy, the challenges of modeling 19th- and early 20th-century Slovak prose, and the tools and workflows adopted to ensure consistency and interpretive flexibility.

Presentation: Publishing with TEI Publisher

With the corpus fully encoded, the final phase focused on making the material available through a web-based interface that preserved the richness of the TEI markup while offering a smooth, intuitive user experience. For this, TEI Publisher served as the central platform, chosen for its native support of TEI-XML and its flexibility in presenting complex textual structures. The platform enabled not only the display of texts but also faceted browsing, full-text search, and customizable views tailored to different user groups—whether scholars, educators, or general readers. TEI Publisher's reliance on standards-based technologies like XSLT and REST APIs also allowed for future integration with visualization tools or external datasets. This section discusses the implementation of TEI Publisher in the context of the corpus, detailing how its configuration and extensions were used to bridge the gap between encoded data and accessible digital editions.

Bibliografia

Brown, Gordon, ed. *The Universal Declaration of Human Rights in the 21st Century, a Living Document in a Changing World.* Cambridge, [New York]: Open Book Publishers ; NYU Global Institute for Advanced Study, 2016.

"Extensible Markup Language (XML) 1.0 (Fifth Edition)." Accessed April 15, 2025. https://www.w3.org/TR/REC-xml/.

Fennell, Philip. "Extremes of XML." In *XML London 2013 Conference Proceedings*, 80–86. XML London, 2013. https://doi.org/10.14337/XMLLondon13.Fennell01.

"Unicode Standard." Accessed April 15, 2025. https://www.unicode.org/standard/standard.html.