

Univerzita Karlova

Filozofická fakulta

Ústav informačních studií a knihovnictví

Studijní obor: Informační studia a knihovnictví

BAKALÁŘSKÁ PRÁCE

Filip Kersch

Standards používané pro kódování předmoderních textů
a způsoby jejich prezentace v digitálních knihovnách

Standards used for encoding of pre-modern texts
and ways of their presentation in digital libraries

Děkuji za pomoc, přínosné podněty, vstřícnost a trpělivost svému vedoucímu práce
Mgr. Jindřichu Markovi, Ph.D.

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu

V Praze dne

.....
Podpis

Abstrakt

Bakalářská práce podává přehled o standardech a metodách používaných pro kódování předmoderních textů a způsobech prezentace těchto textů v digitálních knihovnách. Literárně přehledová část práce popisuje zásady pro kódování elektronických textů vytvořené konsorciem TEI a proces přípravy kódovaného textu. Představena je struktura zásad TEI, jejich modulárnost a možnosti popisu textových i netextových prvků obsažených v dokumentech. Dále se tato část zabývá způsoby prezentace textů v digitálních knihovnách, a to s přihlédnutím k současným trendům a především pravidlům webové použitelnosti. V praktické části práce jsou na základě kontrolního seznamu vytvořeného na podkladech teoretické části analyzovány vybrané digitální knihovny ve třech oblastech: způsobech prezentace textů včetně jejich kódování, aspektů interoperability a dodržování pravidel použitelnosti. V závěru práce jsou zjištěné informace vyhodnoceny a zasazeny do kontextu teorie.

Klíčová slova:

digitální knihovna, kódování textu, TEI, webová použitelnost

Abstract

The bachelor thesis provides an overview of standards and methods used for encoding of pre-modern texts and ways of presenting these texts in digital libraries. The literature review part of the thesis describes the principles for encoding electronic texts created by the TEI consortium and the process of preparing an encoded text. The structure of the TEI Guidelines, its modularity, and the possibilities of describing textual and non-textual elements contained in documents are presented. Furthermore, this part also investigates different ways of presentation of texts in digital libraries, considering current trends and especially the web usability principles. In the practical part of the thesis, selected digital libraries are analysed according to a checklist created on the basis of the theoretical part. Three main areas are examined: ways of presenting texts, including their encoding, interoperability aspects and compliance with usability principles. The information obtained is evaluated and put in the context of the theory at the end of the thesis.

Keywords:

digital library, text encoding, TEI, web usability

Obsah

1	Úvod.....	9
2	Standardy používané pro kódování textů.....	10
2.1	Text Encoding Initiative.....	12
3	TEI P5 – Zásady pro kódování elektronických textů	14
3.1	Úvod do XML.....	15
3.2	Struktura schématu TEI.....	17
3.3	Minimální struktura TEI dokumentu	20
3.4	Popis textových prvků.....	23
3.4.1	Strukturální jevy	23
3.4.2	Zvýraznění	25
3.4.3	Sémanticko-logické jevy.....	27
3.4.4	Analytické poznámky	28
3.5	Popis netextových prvků.....	28
3.6	Vykreslování	29
3.7	Možnosti úpravy schématu TEI	29
4	Proces přípravy textů	31
4.1	Úrovně kódování.....	31
4.1.1	Úroveň 1	31
4.1.2	Úroveň 2 – minimální kódování	32
4.1.3	Úroveň 3 – strukturální analýza.....	32
4.1.4	Úroveň 4 – základní obsahová analýza.....	32
4.1.5	Úroveň 5 – vědecké kódování	33
4.2	Validace dokumentu.....	33
4.3	Produkty digitalizace.....	34
5	Prezentace textů v digitálních knihovnách	35
5.1	Definice a funkce digitálních knihoven	35

5.2	Trendy v oblasti digitálních knihoven.....	37
5.2.1	Interoperabilita a agregace.....	37
5.2.2	Udržitelnost, dlouhodobá ochrana a identifikátory.....	41
5.2.3	Otevřenost.....	43
5.2.4	Zapojení uživatelů a personalizace.....	44
5.3	Způsoby prezentace textů.....	46
5.3.1	Prostý text.....	46
5.3.2	Skeny stránek.....	48
5.3.3	Kombinace skenů a transkripce.....	49
5.3.4	Vizualizace dokumentů kódovaných podle zásad TEI.....	52
5.4	Pravidla použitelnosti webu.....	55
5.4.1	Klíčové principy webové použitelnosti.....	56
5.4.2	Metody hodnocení použitelnosti.....	60
6	Analýza vybraných digitálních knihoven.....	62
6.1	Deutsches Textarchiv.....	64
6.1.1	Prezentace textu.....	64
6.1.2	Interoperabilita.....	67
6.1.3	Webová použitelnost.....	67
6.2	e-codices – Virtual Manuscript Library of Switzerland.....	71
6.2.1	Prezentace textu.....	71
6.2.2	Interoperabilita.....	72
6.2.3	Webová použitelnost.....	72
6.3	Manuscriptorium.....	76
6.3.1	Prezentace dokumentů.....	76
6.3.2	Interoperabilita.....	77
6.3.3	Webová použitelnost.....	79
6.4	Medieval Nordic Text Archive.....	81
6.4.1	Prezentace textu.....	81

6. 4. 2 Interoperabilita.....	84
6. 4. 3 Webová použitelnost.....	84
6. 5 Oxford Text Archive	89
6. 5. 1 Prezentace dokumentů	89
6. 5. 2 Interoperabilita.....	91
6. 5. 3 Webová použitelnost.....	91
6. 6 Projekt Gutenberg	94
6. 6. 1 Prezentace textů	94
6. 6. 2 Interoperabilita.....	94
6. 6. 3 Webová použitelnost.....	95
6. 7 Religiöse Friedenswahrung und Friedensstiftung in Europa	100
6. 7. 1 Prezentace textů	100
6. 7. 2 Interoperabilita.....	102
6. 7. 3 Webová použitelnost.....	102
7 Shrnutí a vyhodnocení analýzy.....	106
8 Závěr	112
9 Seznam použité literatury	113
10 Seznam obrázků.....	125
11 Seznam tabulek	126
12 Seznam příkladů	126
13 Seznam příloh	126
Příloha č. 1 – Kontrolní seznam.....	127
Kontrolní seznam pro analýzu vybraných digitálních knihoven	127

1 Úvod

Digitalizace dokumentů je v současnosti poměrně běžnou záležitostí. Samotný digitalizovaný text ovšem představuje jen velmi omezený zdroj pro další badatelskou činnost, a to zejména vezmeme-li v úvahu i možnosti, které se ve vědeckých činnostech otevřely s rozvojem informačních a komunikačních technologií. Rozšíření digitální reprezentace dokumentů o strojově čitelné informace pomocí kódování může výrazně rozšířit potenciál digitalizovaných textů, a to nejen v rámci akademické sféry.

Cílem bakalářské práce je podat přehled o standardech používaných pro kódování předmoderních textů, tedy textů vydaných či napsaných do roku 1800, a způsobu jejich prezentace veřejnosti prostřednictvím digitálních knihoven. Práce obsahuje literárně přehledovou část a část praktickou.

V literárně přehledové části budou představeny zásady pro kódování elektronických textů vytvořené a udržované iniciativou TEI (Text Encoding Initiative), ale také proces přípravy kódovaného textu a různé způsoby prezentace textů v digitálních knihovnách. Pozornost zde bude věnována i současným trendům v této oblasti zahrnujícím problematiku interoperability, trvalé identifikace dokumentů, otevřenosti a zapojení uživatelů. Tato část také poukáže na základní principy použitelnosti webu, neboť současné digitální knihovny jsou ve své podstatě typem webových stránek.

Praktická část práce se bude zabývat rozbořem vybraných digitálních knihoven. Na základě podkladů poskytnutých literárně přehledovou částí bude sestaven kontrolní seznam, podle kterého budou zjišťovány informace týkající způsobu kódování textů včetně použitých schémat a hloubky kódování. Analyzovány budou i přístupy digitálních knihoven k formě prezentace textů, aspektům interoperability a pravidlům webové použitelnosti

Celkový počet znaků práce (od úvodu až po závěr) je 180 454 včetně mezer (tj.100 normostran).

2 Standardy používané pro kódování textů

Předmoderní texty představují velmi rozsáhlou a nepříliš sourodou kategorii zahrnující rukopisy, inkunábule (prvotisky pocházející z druhé poloviny 15. století) a takzvané staré tisky, tedy veškeré knihy vytištěné v období mezi lety 1500-1800 [1]. Každá z těchto skupin textů je velmi specifická, ale zároveň mají všechny některé společné vlastnosti – jednou z nich je jejich dostupnost, respektive nedostupnost. Stará tištěná díla se sice často na rozdíl od rukopisů dochovala ve více exemplářích, není jich ale nepřehledné množství, a zvláště v případě nejstarších tisků a prvotisků není výjimkou dochování jen několika či pouze jediného exempláře. Tisky mohou být také vzhledem ke svému stáří poškozené či neúplné. Relativně častým problémem mohou být chybějící okrajové listy, které nesou z hlediska knihovědy významné informace o autorovi, názvu knihy, místě tisku a podobně. Porušený nebo chybějící text lze rekonstruovat, pouze pokud disponuje badatel více exempláři téhož díla. V případě rukopisů jsou pochopitelně všechny zmíněné problémy ještě mnohokrát výraznější. Získat pro badatelskou činnost samotný materiál, případně jeho dostatečné množství, ovšem není jen problém logistický, jak by se mohlo na první pohled zdát. Vzhledem k unikátnosti těchto objektů je totiž často nemožné poskytnout je odborníkům k dlouhodobějšímu prezenčnímu studiu a pokud už to možné je, přístup k danému objektu může mít vždy v jednu chvíli jen jeden badatel. Představa, že by s materiály mohli pracovat například vysokoškolští studenti nebo dokonce široká veřejnost je bohužel nereálná. Tento stav nijak nepomáhá rozvoji vědních oborů spojených s knihou, ale ani rozšiřování kulturního povědomí veřejnosti a dlouhodobé ochraně těchto kulturních artefaktů. V podstatě všechny zmíněné překážky jsou ovšem do jisté míry řešitelné současnými technickými možnostmi, zejména digitalizací a zpřístupňováním na internetu. [2, 3]

Proces digitalizace textu zahrnuje ve své běžné podobě naskenování fyzické předlohy, zpracování získaných obrazů a uložení nebo případně publikování digitalizované verze textu. Digitalizovaný text sám o sobě ale poskytuje pouze velmi omezené možnosti pro jeho další zkoumání v rámci digitálních humanitních věd¹. Aby bylo s textem možné dále vědecky pracovat, je nezbytné, aby byla jeho digitální reprezentace doplněná dalšími strojově čitelnými informacemi o jeho struktuře, sémantice, interpretaci či procesech a postupech, kterými byl

¹ Pojem digitální humanitní vědy (digital humanities) představuje označení pro „*nové modely bádání a institucionálních subjektů zaměřených na spolupráci a interdisciplinaritu v rámci výpočetně založeného výzkumu, výuky a publikační činnosti. Digitální humanitní vědy nejsou ani tak jednotným oborem, jako souborem souvisejících praktik, kterými se zkoumá prostředí, v němž tisk již není primárním médiem pro produkci a šíření vědění*“. [5]

zpracován. Vytváření přidané informační vrstvy se nazývá kódování textu, v angličtině text encoding. [4, 5, 6]

Právě ona metadatová vrstva představuje zásadní rozdíl mezi textem, který je pouze zpřístupněn elektronicky a elektronickým textem určeným pro vědecké bádání. Zatímco elektronicky zveřejněný text představuje dokument s předem vybranými a zafixovanými vlastnostmi s cílem pouhého publikování, digitální vědecký text představuje data extrahovaná z primárního dokumentu, která mohou být navíc obohacena o znalosti výzkumníka. Tato data potom ve spojení s vhodným nástrojem slouží ke zkoumání textu, jeho smyslu i konotací, ale také analýze napříč jednotlivými texty v rámci určité kolekce. Zásluhou kódování je možné provádět sofistikovanou textovou analýzu nebo vytvářet kritická či variantní vydání, která mohou být kombinována do ucelených textových databází, ve kterých je možné pokročile vyhledávat. Současné směřování digital humanities se dokonce snaží překročit hranici primárnosti textu a klade důraz na grafické metody výstavby a uspořádání vědění, na design, transmedialitu a výstavbu nástrojů a platforem, které umožní výzkum nejen textů a prvků v nich obsažených takovými způsoby, které by bez použití počítačů byly velmi nepraktické nebo úplně nemožné. Na počátku takto ambiciózních projektů ovšem stojí kvalitně připravená data a metadata o nich. [4, 5, 6]

Tvorba strojově čitelných textů pro použití v humanitním výzkumu má svůj počátek na přelomu druhé poloviny 20. století. Pro první projekty zahrnující kódování literárního textu se využívalo technologie děrných štítků, nicméně s rozvojem digitálních technologií se velmi brzy prosadil jediný přístup pro vytváření metadatové vrstvy, a to používání značkovacích jazyků (markup languages). Ty umožňují přidat dodatečné informace přímo do textu pomocí značek (tags), které jsou formálně odlišné od znaků digitálního přepisu textu. Značky slouží k identifikaci logických i fyzických aspektů textu, ale v nijak neformátovaném pohledu na takto označený, zakódovaný text se jeví pouze jako těžko pochopitelné shluky symbolů smíchané se slovy původního textu. [6]

2.1 Text Encoding Initiative

Jednotná praxe používání značkovacích jazyků ovšem sama o sobě neznamenal usnadnění interoperability a interaktivity, kterou by ideálně měly digitální humanitní texty umožňovat. Velké množství různých přístupů, zájmů a názorů totiž vedlo k tvorbě velmi specializovaných, rozmanitých a vzájemně těžko kombinovatelných kódovacích schémat [6]. Přelom v tomto směru znamenala schůze humanitních vědců zabývajících se různými přístupy ke kódování, která byla uspořádána v roce 1987 na Vassar College v americkém městě Poughkeepsie. Na tomto setkání byly formulovány takzvané Poughkeepské principy, které definovaly intelektuální základy Text Encoding Initiative – Iniciativy pro kódování textu. Iniciativa byla vytvořena proto, aby vyvinula soubor univerzálních zásad pro kódování textu. Zmíněné principy byly později komplexněji formulovány a publikovány v roce 1988 v dokumentu s názvem Design Principles for Text Encoding Guidelines, v překladu: Zásady návrhu pro pravidla kódování textu [7, 8].

Práce na vývoji zásad zaštitovaly tři organizace: The Association for Computers in the Humanities (Asociace pro počítače v humanitních oborech), Association for Literary and Linguistic Computing (Asociace pro literární a lingvistické výpočty, dnes Evropská asociace pro Digital Humanities) a Association for Computational Linguistics (Asociace pro výpočetní lingvistiku). Byly vytvořeny čtyři pracovní výbory a koncem roku 1989 již bylo do práce na zásadách zapojeno více než 50 vědců. První návrh podoby zásad označovaný jako P1 byl vydán v roce 1990. V následujících letech pracovalo 15 různých pracovních skupin na revizích a rozšířeních tohoto návrhu, přičemž zohledňovaly i komentáře, opravy a doplnění přicházející od vědců z celého světa, kteří již současnou verzí zásad používali ve svých projektech. První oficiální verze zásad označená jako P3 byla vydána v roce 1994. [4, 7]

V roce 1999 byl výboru TEI předložen návrh univerzit ve Virginii a Bergenu na vytvoření mezinárodní členské organizace nazvané TEI Consortium, která by měla udržovat a pokračovat v rozvoji TEI jako demokraticky konstituovaná a akademicky i ekonomicky nezávislá nezisková organizace. Tento návrh byl pozitivně přijat a konsorcium začalo oficiálně fungovat na přelomu let 2000 a 2001. Po jeho vytvoření se stalo jasnou prioritou přetvořit verzi zásad P3 tak, aby její uživatelé mohli pracovat se stále populárnějším značkovacím jazykem XML – do té doby byl využíván jazyk SGML. Tato verze (P4) byla vydána v roce 2002. Následovala důkladná revize, která zohledňovala vývoj v řadě klíčových oblastí včetně užívaných kódování znaků pro počítačové soubory, grafiky nebo popisu rukopisů. Nová verze zásad označená jako P5 byla vydána v listopadu 2007 a je stále aktuální [7]. Zásady TEI P5 nejsou zpětně

kompatibilní se staršími verzemi, nicméně verze P4 byla průběžně aktualizována až do roku 2012. TEI Konsorcium dále zásady rozvíjí a doplňuje skrze aktualizace vydávané každých několik měsíců. Od listopadu 2007 do dubna 2021 tak bylo vydáno 37 dílčích verzí TEI P5 [9, 10].

Současná podoba Zásad TEI je výsledkem velmi rozsáhlé mezinárodní spolupráce mnoha vědců, profesí a institucí, zásluhou čehož jsou v nich zpracovaná různá hlediska, zájmy a způsoby práce s textem. I proto jsou dnes Zásady TEI v podstatě jediným standardem používaným pro kódování textů a jsou používány po celém světě [11]. Konsorcium TEI dále pokračuje v jejich vývoji, ale mezi své cíle dále řadí jejich rozšiřování (zásady jsou dostupné v mnoha formátech a jazycích), pořádání školení a kultivování komunity, která okolo TEI vznikla [12].

3 TEI P5 – Zásady pro kódování elektronických textů

Než bude možné představit zásady TEI ke kódování textu a samotné kódovací schéma, je nejdříve potřeba vysvětlit některé terminologické problémy panující v této oblasti. Termín „Zásady TEI“ (TEI Guidelines) používá totiž konsorcium ve svých dokumentech jak k označení formální dokumentace popisující kódovací systém, tak k označení kódovacího systému samotného. Zkratka „TEI“ může odkazovat na samotné konsorcium, na Zásady TEI nebo také na kódovací schéma definované v zásadách. Zároveň je možné o dokumentu, který byl kódován pomocí TEI říct, že je vyjádřen v TEI. Pro větší srozumitelnost tedy bude v následujícím textu zkratka „TEI“ používána výhradně pro označení kódovacího schématu, „Zásady TEI“ budou odkazovat na dokumentaci ke schématu a samotné „TEI konsorcium“ bude vždy označeno jako na „konsorcium“.

Zásady TEI jsou definovány jako pravidla a doporučení, kterými je vhodné se řídit. Konsorcium záměrně neoznačuje zásady za standard, neboť zastává názor, že každý vědecký projekt a potažmo každý vědec má mít možnost postupovat při práci s textem svobodně a kódování používat pro takové aspekty textu, které považuje za důležité [10]. Zásady TEI P5 se skládají z kódovacího schématu, jeho dokumentace a zdrojového kódu TEI, který lze využít pro vytvoření svého vlastního kódovacího schématu [9] (více v kapitole 3.7).

Jak bylo zmíněno výše, kódovací schéma TEI využívá značkovacího jazyka XML, proto bude tento jazyk nejprve krátce představen, aby mohl následovat popis struktury schématu TEI, a především přestavení hlavních možností, které TEI v kódování textů nabízí spolu s jednoduchými příklady. Závěr kapitoly nastíní možnosti úpravy schématu pro potřeby specifických projektů.

3.1 Úvod do XML

Jazyk XML je otevřeným a neproprietárním standardem, který definovalo konsorcium W3C. Specifikuje formát pro přenos dat a obecných dokumentů, který vyvažuje potřeby lidí číst nebo zapisovat data s potřebami strojů číst nebo zapisovat data. Jazyk vychází ze staršího standardu SGML (Standard Generalized Markup Language – Standardní obecný značkovací jazyk), ze kterého vznikl například i formát dokumentů HTML. Zkratka XML znamenající eXtensible Markup Language, tedy rozšiřitelný značkovací jazyk, napovídá, jaká je hlavní výhoda a rozdíl XML a HTML. Zatímco sada značek formátu HTML je pevná, sada značek v XML může být definována různě pro různé sady dokumentů. XML tak vlastně není skutečně značkovacím jazykem, tvoří spíše rámec, ve kterém mohou vznikat specifické značkovací jazyky (jako je TEI) s vlastními definicemi. Definice značek jsou potom buď součástí XML dokumentu samotného nebo specifikované odkazem. [13, 14]

Značky slouží k označení jednotlivých částí dokumentu – celých odstavců, slov nebo i jednotlivých znaků. Rozlišují se značky otevírací (ty jsou umístěny před označovanou částí) a zavírací (ty jsou za ní).

Jednoduchý příklad může být:

```
<pozdrav>Dobrý den, sousede!</pozdrav>.
```

Značky lze nořit do sebe, musí být ale uzavřeny ve správném pořadí, tedy nesmí se křížit:

```
<pozdrav>Dobrý den, <oslovení>sousede!</oslovení></pozdrav>.
```

Zmíněné příklady jsou příklady elementů. Elementy mají obvykle obsah, který se nachází mezi otevírací a zavírací značkou – tak jako v příkladu má element „oslovení“ obsah „sousede!“. Kromě samotného obsahu mohou mít elementy ještě atributy a hodnoty [13, 14]. V následujícím příkladu je atributem „typVěty“ a hodnotou tohoto atributu „tázací“:

```
<věta typVěty="tázací"> Jak se máte?</věta>.
```

Ačkoliv může XML dokument na první pohled vypadat jen jako dlouhá sekvence smíchaných značek a textu, ve skutečnosti vytváří hierarchii, která je definována zanořováním elementů do sebe. Tato hierarchie je přirovnávána ke stromové struktuře, protože se skládá se z kořenového (root) elementu, který obsahuje všechny ostatní elementy. Pro označení elementů se podle jejich umístění ve struktuře užívají termíny jako rodič, předek a potomek. V příkladu č. 1 začíná kořenový element značkou `<bakalarskaPrace>`, elementy označené `<kapitola>` jsou jeho potomci, zároveň je `<bakalarskaPrace>` nejen kořenovým elementem ale také rodičem pro jednotlivé kapitoly (`<kapitola>`). Stromová struktura je výhodná pro práci s digitálním textem, protože text v dokumentu lze často pojmut jako hierarchii (bývá členěn na kapitoly, odstavce a podobně) a zároveň tato struktura umožňuje efektivně zpracovat text počítači, které pracují výrazně rychleji zpracovávají-li data v hierarchii než data nestrukturovaná [15]. Ne vždy a ve všech kontextech je samozřejmě možné a vhodné dokumenty a jejich obsah za jednoduché hierarchie považovat [4], i přes to ale zůstává XML a TEI (alespoň zatím) jednoznačně převažujícím způsobem pro kódování textu.

```
<?xml version="1.0"?>
<bakalarskaPrace>
  <kapitola cislo="1">Úvod</kapitola>
  <kapitola cislo="2">Standardy používané pro kódování textů</kapitola>
  <kapitola cislo="3">TEI P5</kapitola>
</bakalarskaPrace>
```

Příklad č. 1: XML dokument

Aby byl XML dokument skutečně XML dokumentem, musí být takzvaně dobře vytvořen (well formed). To znamená, že má jediný kořenový element, který obsahuje všechny ostatní elementy a ty jsou do sebe řádně vnořeny. Pokud v XML používáme konkrétní značkovací jazyk je vhodné také ověřit platnost (validitu) vytvořeného dokumentu podle pravidel, která daný jazyk definují. Tato pravidla, někdy nazývaná také gramatikou XML dokumentu, se označují jako schéma. Schéma běžně určuje, které elementy se mohou nořit do ostatních a v jakém pořadí, kolikrát je povoleno je opakovat nebo zda mohou úplně chybět. Schéma také popisuje, jaké atributy patří ke kterým elementům a jakých hodnot mohou různé atributy nabývat [13]. Více informací k validaci dokumentu podle TEI schématu přináší kapitola 4.2.

3.2 Struktura schématu TEI

Protože schéma TEI je koncipováno tak, aby mohlo sloužit ke kódování jakéhokoli žánru textu z různých časových období v různých jazycích, je extrémně robustní – obsahuje více než 500 elementů. V praxi ale uživatelé používají vždy jen část z nich, podle povahy projektu, na kterém pracují. Aby bylo použití schématu a výběr potřebných elementů pro dané účely co nejsnadnější, skládá se kódovací schéma z modulů, přičemž každý z těchto modulů definuje určitou sadu XML elementů a jejich atributů. Finální schéma pro kódování v rámci specifického projektu je tak možné sestavit kombinací těchto modulů, případně ho i doplnit vlastními elementy (více o úpravách schématu v kapitole 3.7). [16]

Některé moduly jsou nicméně obzvláště důležité a měly by být součástí schématu vždy. Je to modul *tei*, který definuje třídy, makra a datové typy, s nimiž pracují ostatní moduly, modul *core* (jádro) definující elementy a atributy tak časté, že se téměř jistě vyskytují ve všech dokumentech a modul *header* (záhlaví), který poskytuje metadatové elementy a atributy konstruuující povinný TEI Header (záhlavím se zabývá následující podkapitola). Všechna schémata budou také velmi pravděpodobně obsahovat modul *textstructure*, který definuje základní strukturální elementy potřebné pro kódování textu v knihách a všech objektech podobných knihám.

Celkem pracuje schéma TEI s 21 moduly. Každý z těchto modulů je v zásadách TEI podrobně rozebrán, jsou zde ukázány příklady a vysvětleny myšlenky, které za definovanými elementy a zamýšleným použitím modulů stojí. Zde bude nastíněn jen velmi stručný přehled dostupných modulů, každý bod obsahuje oficiální označení modulu, překlad jeho názvu a krátký popis [16, 17]:

- Modul *analysis* – Jednoduchý analytický mechanismus
obsahuje definice elementů a atributů umožňujících přidružení jednoduchých analýz a interpretací k textovým prvkům
- Modul *certainty* – Jistota, přesnost a odpovědnost
obsahuje definice elementů umožňujících označit aspekty kódovaného textu za problematické či nejisté a identifikovat odpovědnost za toto kódování
- Modul *core* – Společné jádro
obsahuje definice elementů a atributů, které je možné použít ve všech textech, bez ohledu na žánr nebo povahu textu. Tyto prvky zahrnují textové jevy (odstavce, citace,

zvýraznění), datové struktury (jména, adresy, data, čísla), ale i prvky pro přidávání křížových odkazů, poznámek, a grafických součástí

- Modul *corpus* – Metadata pro jazykové korpusy
obsahuje definice elementů a atributů pro kódování korpusů textů shromážděných podle konkrétních kritérií – elementy se vztahují zejména k dokumentaci těchto kritérií a kontextovým informacím
- Modul *dictionaries* – Slovníky
definuje prvky pro kódování slovníků, které mohou být strukturované pomocí elementů označujících homonyma, smysl a tvar slov, gramatické informace, definice a způsoby použití konkrétních slov.
- Modul *drama* – Dramatické texty
definuje elementy a prvky pro kódování a analýzu dramatických materiálů (tištěných scénářů, ale i obecně písemných prepisů jakékoli formy představení). Elementy umožňují kódování seznamů herců, jejich projevů, scénických poznámek a podobně. Pro multimediální tvorbu jsou k dispozici prvky pro popis úhlů kamery, titulků a zvuku
- Modul *figures* – Tabulky, vzorce, grafika a obrázky
obsahuje definice pro podrobné znázornění grafických prvků v textech, jako jsou tabulky, vzorce a obrázky
- Modul *gaiji* – Reprezentace nestandardních znaků a glyfů
Definice způsobů pro reprezentaci znaků, jejichž vyjádření není standardizováno
- Modul *header* – TEI Header (záhlaví dokumentu TEI)
obsahuje definice elementů tvořících záhlaví TEI dokumentů. Slouží k zakódování metadat o bibliografických aspektech textů a jejich vztahů ke zdrojovým materiálům.
- Modul *iso-fs* – Struktury prvků
definuje elementy a atributy pro datové struktury (struktury prvků), které identifikují a seskupují jednotlivé prvky, z nichž každý spojuje označení s jednou nebo více hodnotami. Tyto struktury jsou relativně obecné a lze je použít pro konstrukci analytických rámců, které lze použít k reprezentaci konkrétních analýz v textech.
- Modul *linking* – Propojování a členění
definuje prvky a atributy pro reprezentaci systémů křížových odkazů mezi jednotlivými prvky v textu. Umožňuje tak analyzovat strukturu textu, která není lineární nebo hierarchická

- Modul *msdescription* – Popis rukopisů
definuje speciální prvky pro popis podrobných informací o ručně psaných primárních zdrojích a teoreticky i dalších objektech obsahujících text, u kterých je důležité zaznamenat fyzický popis, historii a další vlastnosti objektu. Mezi definované elementy patří například i elementy pro popis heraldiky, rozměrů nebo vodoznaků
- Modul *namesdates* – Jména, data, lidé a místa
definuje elementy a atributy pro kódování jmen a frází popisujících osoby, místa a organizace, a to podrobněji, než umožňuje elementy dostupné ve všech TEI dokumentech (viz modul core).
- Modul *nets* – Grafy a sítě
definuje elementy a atributy pro analytické znázornění schematických vztahů mezi uzly označenými v textu.
- Modul *spoken* – Přepis řeči
definuje elementy a atributy pro kódování široké škály přepsaných mluvených materiálů. Elementy slouží mimo jiné k popisu například promluv, pauz, zvuků nebo gest.
- Modul *tagdocs* – Elementy dokumentace
obsahuje definice prvků a atributů pro dokumentaci kódovacího schématu
- Modul *tei* – Infrastruktura TEI
obsahuje definice datových typů, tříd a maker pro použití v ostatních modulech
- Modul *textcrit* – Textová kritika
obsahuje definice elementů a atributů pro reprezentaci různých verzí textů a jejich změn
- Modul *textstructure* – Výchozí struktura textu
obsahuje definice elementů a atributů popisujících strukturu textu kódovaného v TEI. Obsahuje například prvky pro označení desek, přideštů, titulní strany a podobně.
- Modul *transcr* – Reprezentace primárních zdrojů
obsahuje definice prvků a atributů použitelných pro reprezentaci primárních zdrojů jako jsou rukopisy nebo tištěné knihy. Elementy pokrývají mimo jiné způsoby kódování faksimile, doplňků, škrťů či poškození zdrojového materiálu nebo jeho nečitelnost
- Modul *verse* – Verše
definuje elementy a atributy pro kódování textů, které jsou úplně nebo převážně psány ve verších. Umožňuje například kódovat cézury a rýmovací schémata

Následující text se nebude věnovat popisu všech modulů, ale pouze relativně stručně představí strukturu TEI dokumentů a ukáže a vysvětlí základní možnosti, které TEI pro kódování textů poskytuje. Postupně budou představeny elementy a jejich atributy sloužící k označení textových i netextových prvků a budou ukázány také možnosti popisu vzhledu textů, které mohou sloužit k jejich následné vizuální prezentaci.

3.3 Minimální struktura TEI dokumentu

TEI dokument se skládá z kořenového elementu `<TEI>` a dvou hlavních komponentů: Elementu pro záhlaví `<teiHeader>`, který obsahuje všechna popisná metadata dokumentu a elementu `<text>`, který obsahuje samotný dokument. Tato struktura je povinná a vypadá následovně:

```
<?xml version="1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!--...-->
  </teiHeader>
  <text>
    <!--...-->
  </text>
</TEI>
```

Příklad č. 2: Minimální struktura TEI dokumentu

TEI Header, záhlaví TEI dokumentu, může být velmi komplexní objekt, ale také může být poměrně krátký a jednoduchý. Ve své minimální podobě obsahuje popis elektronického souboru v elementu `<fileDesc>`. Ten obsahuje co nejpodrobnější bibliografický popis elektronického textu, takže je možné ho použít pro vytvoření bibliografické citace nebo katalogizačního záznamu. `<fileDesc>` povinně obsahuje tři elementy:

- Prohlášení o názvu (`<titleStmt>`), kde jsou popsány informace o názvu (`<title>`) a autorovi (`<author>`)
- Prohlášení o publikaci (`<publicationStmt>`), které popisuje podrobnosti o publikaci elektronického textu, a to buď strukturovaně nebo jako prostý text

- Popis zdroje (<sourceDesc>), kde jsou dokumentovány bibliografické informace o zdrojovém materiálním dokumentu, ze kterého elektronická verze pochází. Tyto informace mohou být opět popsány strukturovaně nebo prostým textem v odstavcích (<p>).

Kromě popisu souboru, jehož přítomnost je povinná, je možné do záhlaví umístit i poměrně velké množství dalších metadat. Element <encodingDesc> umožňuje vložení popisu kódování, kde je vysvětlován vztah mezi elektronickým textem a jeho zdrojem a jsou zde uvedeny informace o tom, zda a jak byl text při kódování normalizován, do jaké úrovně kódování probíhalo, jak byly řešeny nejasnosti v textu a podobné záležitosti. Také je možné v záhlaví popsat (v elementu <profileDesc>) klasifikační a kontextové informace o textu, jako například jeho předmět, v jaké situaci vznikl a jací jednotlivci se na jeho vzniku podíleli. Takové informace mohou být užitečné například při kódování korpusů. Záhlaví může také obsahovat popis provedených revizí v průběhu tvorby elektronického textu (<revisionDesc>) a do elementu <xenData> mohou být vložena metadata jiných schémat, než je TEI – nejčastěji MARCXML, MODS nebo Dublin Core. [18, 19]

Druhá základní část TEI dokumentu – element <text> – obsahuje skutečný text dokumentu a jeho kódování. Minimálně musí mít tento element jeden podřízený prvek: tělo dokumentu (<body>). To obsahuje textové struktury nižší úrovně jako jsou odstavce <p>, případně řádky <l> pro poezii nebo promluvy <sp> pro dramata, které se sdružují do oddílů. Před samotným textem může být zařazen element <front>, který slouží pro kódování titulních stran, tiráže, předmluv a podobně. Jednotlivé komponenty této přední části se odlišují pomocí vnořeného tagu pro oddíl <div> se specifickým atributem. Podobně může za tělem textu následovat oddíl pro kódování například dodatků, glosáře či poznámek. Pro něj se používá značka <back> a její obsah se opět upřesňuje oddílem s atributem. [18]

```

1. <?xml version="1.0"?>
2. <TEI xmlns="http://www.tei-c.org/ns/1.0">
3.   <teiHeader>
4.     <fileDesc>
5.       <titleStmt>
6.         <title>The History Of the Bohemian Persecution</title>
7.         <author>Jan Amos Komensky</author>
8.       </titleStmt>
9.       <publicationStmt>
10.        <p>Not for distribution.</p>
11.      </publicationStmt>
12.      <sourceDesc>
13.        <bibl>
14.          <title>The History Of the Bohemian Persecution</title>
15.          <author>Jan Amos Komensky</author>
16.          <title>The Penny Histories</title>
17.          <publisher>Printed by B. A. [=Bernard Alsop] for Iohn Walker at the
18.            Star in Popes-Head-Ally</publisher>
19.          <date>1650</date>
20.        </bibl>
21.      </sourceDesc>
22.    </fileDesc>
23.  </teiHeader>
24.  <text>
25.    <front>
26.      <div type="preface">
27.        <p>To the Reader. God, who chastises his church for the faults, and
28.          in the enemy who seeks to destroy it, therefore we lawfully say, with
29.          that martyr mentioned - We suffer for our sins.</p>
30.      </div>
31.    </front>
32.    <body>
33.      <head>The History of the Bohemian Persecution</head>
34.      <div1 type="chapter" n="1">
35.        <head>Chap. I
36.          The conversion of the Bohemians to Christianity and the persecution
37.          that followed
38.        </head>
39.        <p>The Bohemians - Being Idolaters lived without the knowledge of the
40.          true God, and of Religion, till the year of Chist 894. In that year,
41.          their Capraine, being entertained by the King of Moravia, received
42.          the knowledge of Christ.</p>
43.      </div1>
44.    </body>
45.  </text>
46. </TEI>

```

Všechny elementy definované TEI mohou stejně jako jakékoliv elementy XML nést jeden nebo více atributů, které fungují jako jejich kvalifikátory a kvantifikátory. Většina atributů je specifická a jejich použití je povoleno jen ve spojení s pevně danými elementy (o některých z nich bude řeč dále). Několik atributů je ale globálních a mohou se pojit s jakýmkoli TEI elementem. Těmito globálními atributy jsou například `@xml:id` tvořící unikátní identifikátor elementu, `@n` představující číslo elementu, `@xml:lang` určující jazyk obsahu elementu. Speciální kategorií globálních atributů jsou atributy odpovědnosti, kam patří například `@resp` označující agenta odpovědného za intervenci nebo interpretaci a globální atributy popisující vykreslování (tém je věnována podkapitola 3.6). [18]

3. 4 Popis textových prvků

Pravidla a doporučení pro označování a popis široké škály fenoménů vyskytujících se v textech, která zásady TEI definují, lze obecně rozdělit na 4 kategorie, podle toho, jaké jevy popisují. Jsou to jevy strukturální, zvýrazňovací, logicko-sémantické a analytické [18].

3. 4. 1 Strukturální jevy

Pravidla týkající se strukturálních fenoménů umožňují především kódování skladby textu. V nejjednodušším případě tím máme na mysli označení jednotlivých odstavců v próze, promluv a scénických poznámek v dramatech anebo jednotlivých veršů či slok v poezii. Většina textů ovšem tyto základní textové komponenty seskupuje do oddílů a pododdílů. Názvy používané pro tyto oddíly jsou přitom různé s ohledem na žánr textů i období, kdy vznikly – lze se tak setkat například s kapitolami (typicky v próze), knihami (členění eposů či Bible), akty a scénami (u dramatu), čísly a sekcemi (u novin) a mnohými dalšími. Pro zjednodušení této problematiky a větší všestrannost pracují Zásady TEI s univerzálními elementy pro kódování struktury, které mohou být doplněny atributy pro jejich specifikaci (`@type` a `@subtype`). Tyto univerzální elementy mohou být buď číslované, pak se pro ně používají značky `<div1>`, `<div2>`, ... až `<div7>`, přičemž číslo značí hloubku zanoření v rámci hierarchie, nebo nečíslované, kdy se používá pouze značka `<div>`. Tyto dva systémy ale nelze v rámci těla dokumentu kombinovat. [20]

```

<body>
  <div1 type="část" n="1">
    <div2 type="kapitola" n="1">
      <!-- text první kapitoly první části -->
    </div2>
    <div2 type="kapitola" n="2">
      <!-- text druhé kapitoly první části -->
    </div2>
  </div1>
  <div1 type="část" n="2">
    <div2 type="kapitola" n="1">
      <!-- text první kapitoly druhé části -->
    </div2>
  </div1>
</body>

```

Příklad č. 4: Popis struktury s použitím číslovaných oddílů

```

<body>
  <div type="část" n="1">
    <div type="kapitola" n="1.1">
      <!-- text první kapitoly první části -->
    </div>
    <div type="kapitola" n="1.2">
      <!-- text druhé kapitoly první části -->
    </div>
  </div>
  <div type="část" n="2">
    <div type="kapitola" n="2.1">
      <!-- text první kapitoly druhé části -->
    </div>
  </div>
</body>

```

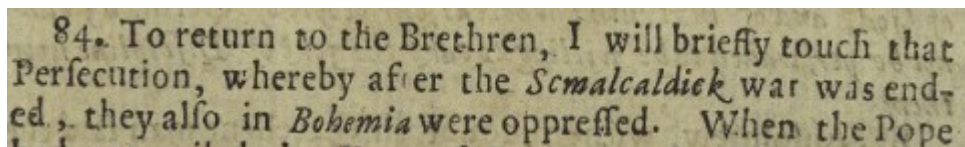
Příklad č. 5: Popis struktury s použitím nečíslovaných oddílů

Jak bylo zmíněno výše, samotný text jednotlivých oddílů se kóduje lehce odlišně pro prozaické texty, poezii a dramata. V případě prózy se pro jednotlivé odstavce používá element `<p>`, případně žánrově neutrální `<ab>` a pro případné seznamy položek element `<list>` pro seznam jako celek a elementy `<item>` pro jednotlivé položky seznamu. Poezie se obvykle skládá z veršů – řádků, pro jejichž popis se používá element `` a jednotlivé řádky se sdružují do skupin pomocí elementu `<lg>`. Pro promluvy v dramatech se používá element `<sp>`, který lze doplnit atributem `@who` nebo elementem `<speaker>`, aby bylo možné identifikovat aktéra, který mluví. Element `<sp>` slouží pouze jako obálka a samotný text promluvy musí být ještě označen (podle potřeby) elementy `<p>`, `<l>`, nebo `<ab>`. Scénické poznámky se kódují pomocí elementu `<stage>`. [21, 22]

Mezi strukturální fenomény patří také element `<head>` pro označení nadpisů nebo element `<titlePage>`, kterým se popisuje titulní stránka. Je důležité tento element nezaměňovat s elementem `<titleStmt>` ve `<fileDesc>` v záhlaví TEI dokumentu – ten slouží k uložení popisných metadat, zatímco `<titlePage>` se používá pro kódování skutečné fyzické titulní strany. [18]

3. 4. 2 Zvýraznění

Ve srovnání se současnými texty, jsou některé rukopisy specifické svými hustě popsanými stranami. Podoba rukopisu výrazně závisela na jeho funkci a s ní spojenými náklady, zejména cena psacího materiálu (nejprve pergamenu, později papíru) byla totiž vysoká. K orientaci na hustě popsaných stranách často sloužily iluminované prvky jako barevné iniciály, nadpisy, stínování nebo podtržení písmen [2]. Nejen proto jsou možnosti kódování těchto typografických prvků v textech velmi podstatné. Zásady TEI nabízí několik způsobů, jak k této problematice přistupovat: Pro pouhé zaznamenání zvýraznění je možné využít značku `<hi>` (z anglického highlight), nicméně protože většinou stojí za zvýrazněním nějaký důvod, často se pro jeho popis používají logické nebo sémantické značky. Pro jejich správné použití je ovšem nutné důkladné porozumění textu a jazyku, ve kterém je napsán. Mezi značky používané pro označení zvýrazněných pasáží textu včetně jejich významu patří například `<emph>` pro zdůraznění, `<foreign>` pro označení slov cizího jazyka a `<distinct>` pro označení jakéhokoli slova nebo slovního spojení, které lze považovat za jazykově odlišné (například je archaické, technické nebo patří do nářečí) [18, 21].



Obrázek č. 1: Kódování zvýraznění²

```
<p>
84. To return to the Brethren, I will briefly touch that Persecution, whereby
after the <emph>Schmalcaldick</emph> war was ended, they also in
<emph>Bohemia</emph> were oppressed. When the Pope
</p>
```

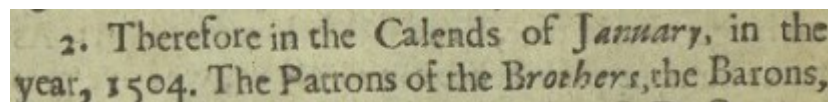
Příklad č. 6: Kódování zvýraznění

V písemných i tištěných textech se poměrně často vyskytuje také zvýraznění pomocí uvozovek. Uvozovky mohou odlišovat část textu od okolí například proto, že se jedná o přímou řeč nebo myšlenku, o technický výraz, žargon, případně si chce autor touto formou udržet určitý odstup. Do uvozovek se samozřejmě vkládají i citace. Není-li potřeba rozlišovat, o který z výše vyjmenovaných případů se jedná, lze text v uvozovkách kódovat jen za pomoci univerzálního tagu `<q>`. Chceme-li opět místo pouhého označení uvozovek zachytit i úmysly skryté za nimi, je možné využít například značky `<said>` pro označení pasáže textu, které jsou myšlenkami nebo vyřčené nahlas, `<quote>`, která se používá pro označení textu přiřčeného vypravěčem agentovy stojícímu mimo samotný text, `<cit>` pro označení citace jiného dokumentu spolu s bibliografickým odkazem na zdroj či elementu `<soCalled>`, kterým je možné označit slovo nebo frázi, u kterých se chce autor zřeknout odpovědnosti za ně (v angličtině se takové uvozovky označují jako scare quotes). [21]

² Zdroj obrázku: KOMENSKÝ, Jan Amos. An Exhortation Of The Churches of Bohemia To The Church of England In: *Česká digitální knihovna* [online digitální knihovna]. Dostupné z: <https://cdk.lib.cas.cz/uuid/uuid:72c1551b-ff98-4fe4-b2d6-2994bd72a123>. s. 52, upraveno

3. 4. 3 Sémanticko-logické jevy

Slova nebo fráze nesoucí sémantické a logické informace se mohou v textu vyskytovat i bez typografického zvýraznění. TEI umožňuje kódovat i takové údaje – například tituly (<title>), jména (<name>), čísla (<num>), míry (<measure>), data (<date>), časy (<time>) adresy (<address>) nebo zkratky (<abbr>), které je možné rozepsat v tagu <expan> (v takovém případě se oba posledně zmíněné elementy uzavírají do elementu <choice>). Všechny tyto elementy jsou dále rozvíjeny a jejich smysl upřesňován použitím atributů. Je tak možné popsat, zda je kódováno jméno člověka, místa, organizace nebo nějakého objektu. Data a časy je možné za pomoci atributu @when přepsat do standardní formy, případně lze atributy @from a @to popsat časový rozsah dané události. Stejně tak se atributy upřesňují informace o kódovaných číslech a mírách – specifikuje se tak například že je číslo procentem, nebo že je řadovou číslovkou. [18, 21]



Obrázek č. 2: Kódování dat³

```
<p>
Therefore in the <date when="1504-01-01">Calends of <emph>January</emph>, in the
year, 1504</date>. The patrons of the <emph>Brothers</emph>, the Barons
</p>
```

Příklad č. 7: Kódování dat

³ Zdroj obrázku: KOMENSKÝ, Jan Amos. The History Of the Bohemian Persecution.... In: *Česká digitální knihovna* [online digitální knihovna]. Dostupné také z: <https://cdk.lib.cas.cz/uuid/uuid:d9a6bcbb-3cda-44e2-bef6-c206481c5281>. s. 67, upraveno

3. 4. 4 Analytické poznámky

TEI umožňuje přidávat do textu ve formě metadat i poznatky, které vznikají během kódování. Příklady takových poznatků představují například opravy překlepů, sjednocení pravopisného úzu nebo označení textu pro indexování. Poznámky k textu je možné přidávat pomocí elementu `<note>` a jeho atributů, které upřesňují typ poznámky (`@type`) nebo osobu, která ji vytvořila (`@resp`). Za účelem vytváření indexu se jednotlivé termíny označují značkou `<term>` a celý element vkládá do elementu `<index>`. Dalšími editorskými zásahy do textu mohou být tvorby dodatků, výmazů nebo opomenutí. Označit je možné také zjevné chyby v textu, pro které se používá element `<sic>`. Chyby je případně možné opravit a v textu označit pomocí `<corr>`, nebo je možné v zakódovaném textu nechat obě varianty, tedy jak chybnou, tak opravenou, a obě uzavřít do elementu `<choice>` [18, 21].

3. 5 Popis netextových prvků

Netextové prvky jako jsou obrázky, fotky, ilustrace, grafy a podobně mohou mít různý vztah k textu a různou důležitost. Například zatímco některé rukopisy sloužily pouze k předávání informací, jiné obsahují kromě samotného textu i výzdobu a ilustrace (často přes celé strany), které mají vysokou výpovědní hodnotu [2]. Je proto na rozhodnutí kodéra, jak se k jejich označení postaví. Nejjednodušším způsobem pro označení grafiky je použití značky `<graphic>`, která se umístí na odpovídající místo v textu. Značku většinou doprovází atribut `@url`, který odkazuje na digitální reprezentaci objektu. Místo odkazu je také možné grafiku vložit přímo do dokumentu v podobě binárních dat (například Base64), které se vkládají do elementu `<binaryObject>`. Grafický prvek je možné samozřejmě doprovodit textem. Ten může být převzatý z popisku (v takovém případě se použije značka `<head>`) nebo se může jednat o komentáře k obrázku (`<p>`), případně o přesný popis (`<figDesc>`). Netextové prvky spolu se svými popisy se sdružují do elementu `<figure>` [18, 21].

3.6 Vykreslování

Vykreslováním (v angličtině rendition) se v kontextu TEI myslí způsoby, kterými jsou elementy formátovány pro prezentaci dokumentů. Jak vyplývá z textu výše, Zásady TEI se věnují především kódování textu s důrazem na sémantiku, definování jeho vzhledu není poskytnut velký prostor. Vzhled je možné popsat především pro zachování těchto informací pro případné další analýzy, nicméně je možné tyto zakódované informace použít i pro prezentaci zakódovaného textu pomocí jiné technologie.

Nejjednodušším způsobem pro zakódování vzhledu textu je použití globálního atributu `@rend`, který umožňuje zaznamenat informaci vztahující se k vykreslení elementu a jeho potomků. Hodnoty atributu `@rend` ovšem nejsou nikde formálně definovány a lze s nimi tedy jen těžko dále pracovat. Další možností je použití atributu `@rendition`, jehož hodnota slouží jako odkaz na definici vzhledu umístěnou jinde v daném dokumentu. K definici stylů slouží kódovací jazyky používané za tímto účelem v oblastech softwarového vývoje – typicky nejpoužívanější variantou je jazyk CSS (Cascading Style Sheets), který se používá pro formátování dokumentů napsaných v HTML. Formátování vyjádřené v CSS je také možné uvést přímo u daného elementu pomocí atributu `@style`. [16]

```
<p>
  Therefore in the <date when="1504-01-01">Calends of
  <emph style="font-weight:italic;">January</emph>, in the year, 1504</date>.
</p>
```

Příklad č. 8: Formátování pomocí CSS přímo v zakódovaném textu

3.7 Možnosti úpravy schématu TEI

Schéma TEI je velmi komplexní a umožňuje jak velmi povrchní a jednoduchý popis dokumentu, tak i vysoce specifickou analýzu. Současná verze obsahuje 589 elementů [23] nicméně zásady TEI nepředpokládají, že by všechny tyto elementy byly používány v rámci jednoho dokumentu současně. Jejich množství se pouze snaží vyhovět co nejširším možnostem použití. Schéma TEI je proto vytvořeno jako velmi modulární systém a umožňuje uživatelům vybrat si pouze ty součásti, které potřebují a ostatními se nezabývat. Jednotlivé moduly byly popsány v kapitole 3.2. Na druhou stranu se ale počítá také s tím, že se může objevit situace,

kteře se zásady vřbec nebo v dostatečné mřře nevěnují. Schéma je proto připraveno tak, že je možné jej rozšřřovat o nové vlastní prvky.

Konsorcium TEI poskytuje řadu vlastních variant TEI, které mohou posloužit pro různé účely. Jedním z nejvíce využívaných je TEI Lite, který obsahuje základní prvky pro popis jednotlivých dokumentů. Předpřipravená jsou ale také schémata například pro popis rukopisů (MS), tvorbu komplexních lingvistických korpusů (Corpus) nebo pro popis raně novověkých tištěných materiálů (TEI simplePrint). Pokud ovšem specifickým potřebám projektu neodpovídá žádné z předpřipravených schémat, je možné si vytvořit na základech TEI vlastní, a to dvěma procesy: výběrem nebo omezením stávajících prvků a atributů TEI a rozšřřením TEI o vlastní, nové prvky a atributy [24]. Dokonce je možné začlenit do schématu další XML jazyky, jako například ontologický jazyk RDF (Resource Description Framework) [25].

Schopnost upravit si schéma podle vlastních potřeb je zásluhou toho, že TEI schéma samo je definováno pomocí systému ODD („One document does it all“ – jeden dokument umí všechno). Systém ODD si lze představit jako specializovaný slovník, který obsahuje vše potřebné k definování a popsání XML schématu. Tento integrovaný dokument potom umožňuje generovat vlastní pracovní schémata a zároveň i jejich dokumentaci. Aby byla možnost upravit si TEI jednodušší, není nutné ručně upravovat ODD, ale je možné využít nástroj konsorcia TEI – Roma, který v jednoduchém a uživatelsky přívětivém rozhraní umožňuje vytvoření vlastního TEI ODD a jeho stažení, případně exportování schématu nebo dokumentace v různých formátech. [26, 27]

Jako příklad lze uvést úpravu, která umožňuje začlenit do TEI podrobný popis papíru a vodoznaků v rukopisech a raných tištěných knihách. Podle autorky úpravy disponuje TEI-P5 rozsáhlými možnostmi pro popis rukopisů, ale definuje jen základní prvky pro popis papíru a vodoznaků. Vytvořila proto dva nové moduly – WatermarkDesc a PaperMoldDesc – které umožňují uživatelům popsat papír a vodoznaky podle pokynů IPHN (Mezinárodní standard pro popis papíru, vodoznaků a forem na papír v relačních databázích) a tvořit popisy papírových forem včetně historických informací o papírnách, které je používaly. [28]

4 Proces přípravy textů

První fází každého projektu, který zahrnuje kódování textu, by mělo být kromě vymezení kolekce dokumentů, se kterými se bude pracovat, také jejich relativně podrobná analýza. Ta umožní rozhodnout, které aspekty textu je žádoucí zakódovat. Jak je zřejmé z předešlých kapitol, možností jak a do jaké hloubky kódovat text je obrovské množství, a to i pokud uvažujeme jen použití oficiálních Zásad TEI. V kombinaci s tím, že TEI schéma je možné upravovat a doplňovat podle vlastních potřeb můžeme říct, že v podstatě každý projekt zahrnující kódování textu, ačkoli by vždy využíval Zásady TEI, bude svým způsobem jedinečný. Právě analýza a definice kódovacího schématu by měly být prvními kroky v takových projektech. Nutné je ovšem počítat i s tím, že až začne samo kódování, pravděpodobně se objeví větší či menší nedostatky navrženého schématu a bude potřeba provést určité úpravy. Jak kódování, tak definování schématu, tak i samotnou prvotní analýzu textů je proto nutné chápat jako iterativní proces. [15, 29]

4.1 Úrovně kódování

Definice schématu, tedy elementů, atributů a pravidel jejich použití, výrazně závisí na typu digitalizačního projektu a jeho účelu. S ohledem na to navrhlo konsorcium TEI ve svém dokumentu shrnujícím osvědčené postupy při používání TEI pět základních úrovní kódování [30]. Tyto úrovně budou krátce představeny:

4.1.1 Úroveň 1

První úroveň je úroveň základní, která počítá s tím, že transkripce textu i následné kódování bude provedeno plně automaticky. Přepis textu se v takovém případě ponechá plně na technologii OCR (Optical Character Recognition), která rozpozná písmena nebo celá slova na naskenovaných stránkách a uloží je buď do jednoduchého textového souboru bez dalších informací, případně vytvoří XML soubory, které mohou obsahovat i údaje o poloze jednotlivých slov na stránce. OCR většinou není úplně dokonalé a například hůře čitelný text často nebývá přepsán správně. V této úrovni kódování se ovšem výstupy OCR nijak neupravují, neboť text kódovaný na této úrovni není zamýšlen pro použití při analýze textu nebo pro samostatné použití – nemůže tedy nahradit skeny dokumentu.

V zásadě jediný přínos kódování na této úrovni je vytvoření elektronického dokumentu umožňujícího vyhledávání v jeho plném textu. Hlavním důvodem pro použití TEI je možnost

připojit k dokumentu záhlaví TEI (TEI Header). První úroveň kódování je tak vhodné použít pro případy, kdy je potřeba rychle publikovat velké množství materiálu, u kterého je možné uživatelům prezentovat digitální obrázky každé stránky, ale zároveň není nutné nabízet sofistikované vyhledávání a zobrazování založené na struktuře textu. Použití TEI zároveň znamená ponechání si eventuality později přidat vyšší úroveň kódování.

4. 1. 2 Úroveň 2 – minimální kódování

Na této úrovni je, na rozdíl od předešlé, vytvářen elektronický text umožňující nejen fulltextové vyhledávání, ale také identifikaci jednoduché strukturální hierarchie pro zlepšení navigace. Transkripce je opět prováděna automaticky pomocí OCR, nicméně to je alespoň základně opraveno a doplněno o metadata umožňující rozeznat nadpisy či bloky textu. Díky tomuto kódování lze například později generovat obsah dokumentu. Použití této úrovně může být vhodné pro případy, kdy není možné provádět analýzu textových prvků nebo podrobné korektury automatického OCR, ale zároveň je žádoucí poskytnout vyhledávání a zobrazování založené na velkých strukturách textu a možnost pozdějšího rozšíření kódování. Ani dokumenty zpracované na této úrovni ovšem nejsou primárně určeny pro prezentaci odděleně od skenů stránek.

4. 1. 3 Úroveň 3 – strukturální analýza

Cílem třetí úrovně kódování je vytvořit elektronický text, ve kterém je relativně podrobně popsána jeho logická struktura. Text je vytvářen buď pomocí OCR, které prochází následnou korekcí, nebo doslovným přepisem primárního dokumentu. V textu jsou identifikovány odstavce, obrázky, poznámky i použití různých typografických prvků. To umožňuje vytvořit řadu různých zobrazení nebo poskytnout podrobné vyhledávání. Text může být prezentován samostatně, ale vzhledem k absenci specializované analýzy obsahu může být vhodné kombinovat ho i se skeny stránek. Tato úroveň tak vyžaduje určité lidské úsilí, nicméně protože obsahová analýza textu stále není primárním cílem, není její použití příliš náročné. Třetí kódovací úroveň je vhodná pro případy, kdy je potřebné zajistit složitější možnosti vyhledávání či zobrazování založené na struktuře textu, ale zároveň není možné využít odborné znalosti pro obsahovou analýzu. Tu je na druhou stranu samozřejmě možné doplnit později.

4. 1. 4 Úroveň 4 – základní obsahová analýza

Čtvrtá úroveň kódování umožňuje vytvořit samostatný elektronický text, ve kterém je popsána struktura textu, typografie, smysl a funkce textových a strukturálních prvků a zároveň i povaha obsahu. Nejsou zde ovšem kódovány všechny strukturální a sémantické prvky. Kódování na

této úrovni zaručuje zobrazování a vyhledávání na velmi sofistikované úrovni (například je možné vyhledat a zobrazit pouze takové promluvy z divadelní hry, které pronáší pouze ženské postavy, a to šeptem). Toto kódování umožňuje textovou analýzu. Text pro tuto úroveň je vytvářen buď pomocí OCR a korekce, nebo přímo přepisem dokumentu. Vzhledem k obsahové analýze a kódování je ovšem tato úroveň poměrně časově i odborně náročná [31].

4. 1. 5 Úroveň 5 – vědecké kódování

Elektronický text vytvořený kódováním na úrovni 5 je schopen stát samostatně, ale na rozdíl od čtvrté úrovně disponuje popisem sémantických, jazykových, prozodických a dalších prvků a využívá tak plně všech možností, které TEI nabízí – součástí textu mohou být také editorské poznámky, kritické nebo analytické doplňky, překlady nebo popisy rukopisů. Cílem kódování na této úrovni je vytvoření hluboce analyticky zakódovaného textu, který je vhodný pro konkrétní výzkumné účely. To také znamená, že k jeho vytvoření je zapotřebí značné zapojení kodérů se znalostí daného předmětu.

4. 2 Validace dokumentu

Vytvořené dokumenty je vhodné systematicky kontrolovat, aby bylo zajištěno, že odpovídají stanoveným pravidlům a schématům. Kontrola je důležitá vždy, zejména ale pokud se mají dokumenty vyměňovat mezi různými výzkumnými infrastrukturami nebo do nich přímo integrovat a také pokud mají být dokumenty použity pro strojové zpracování či jsou-li určeny k dlouhodobé archivaci. Zásady TEI definují v rámci validace dokumentů pojem „TEI Conformance“, který lze volně přeložit jako „vyhovění TEI“. Dokument lze označit za vyhovující TEI, pokud splňuje několik kritérií [32]:

- Je to dobře vytvořený XML dokument
- Pracuje s TEI namespace a použití prvků v rámci tohoto jmenného prostoru je kompatibilní s jejich zamýšlenou funkcí definovanou v Zásadách TEI
- Lze ho ověřit oproti schématu TEI, které je odvozeno ze Zásad TEI
- Je dokumentován pomocí souboru ODD vztahujícího se k Zásadám TEI

Dobře vytvořený XML dokument byl popsán v kapitole 3.1. Všechny prvky, které Zásady TEI vymezují, patří do jmenného prostoru (namespace) TEI, který je definován pomocí URI <http://www.tei-c.org/ns/1.0>. Aby dokument vyhovoval TEI, je nutné, aby byl tento jmenný prostor pro všechny prvky správně deklarován. Běžně se tak činí v kořenové značce, například takto: `<TEI xmlns="http://www.tei-c.org/ns/1.0">`. Dokument nelze

považovat za vyhovující TEI, pokud umísťuje prvky nebo atributy nepatřící TEI, do jmenného prostoru TEI, nebo pokud naopak prvky TEI přiřadí jinému prostoru.

Tyto i další náležitosti je možné do určité míry kontrolovat automatizovaně – je-li pro kódování místo obyčejných textových editorů použit XML editor, ideálně přímo s podporou TEI. Konsorcium TEI udržuje na svých webových stránkách [33] přehled významných editorů XML a základních informací o nich (zda jsou placené, jaké platformy podporují, zda mají explicitní podporu pro XML, TEI a podobně). Jedním z nejpoužívanějších editorů v projektech kódování textu je program <Oxygen/>, který je placený, dostupný na všech platformách a umožňuje nejen validaci vytvořeného dokumentu proti zvolenému schématu, ale napomáhá také s vkládáním TEI elementů a v reálném čase upozorňuje kodéra na případné chyby [29].

Zásady TEI nepožadují, aby bylo pro kódování použito určité konkrétní schéma, konečně přímo umožňují odvození vlastních schémat. Nicméně aby vzniklý dokument vyhovoval TEI, je potřeba, aby bylo kódovací schéma vygenerováno z TEI ODD. To slouží také k dokumentaci použitých modulů a definic prvků, atributů a tříd použitých pro kódování. Aby dokument vyhovoval TEI, je nutné, aby k němu bylo TEI ODD buď přiloženo, nebo na něj bylo náležitě odkazováno [32].

4.3 Produkty digitalizace

Završením digitalizačních procesů bývá prezentace dokumentů v digitálních knihovnách. Způsoby prezentace mohou být velmi různé a věnovat se jim budou další kapitoly. Je ovšem nutné upozornit, že nejdůležitějšími výstupy digitalizace zahrnující kódování textů není často jejich finální podoba dostupná online, ale samotné TEI dokumenty a prostředky spojené s jejich vznikem – tedy analýza textu, výběr či vytvoření kódovacího schématu a dalších pokynů pro kódování a jejich následná aplikace. Právě tyto prvky umožňují jak další výzkum, tak dlouhodobé uchování textů i přidáných odborných znalostí o nich. Prezentace textů v konkrétní aplikaci je sice velmi důležitá, je potřeba ji ovšem chápat jen jako jeden z mnoha možných pohledů na data. Navíc je třeba počítat s tím, že s vývojem technologií a formami digitálního publikování je prezentace textů online relativně pomíjívá a rychle zastarává. Přesto ne vždy zveřejňují digitální knihovny vedle skenů nebo různých vizualizací i zdrojové TEI dokumenty. [34]

5 Prezentace textů v digitálních knihovnách

5.1 Definice a funkce digitálních knihoven

Digitální knihovna je v České terminologické databázi knihovnictví a informační vědy definována jako „*integrováný systém zahrnující soubor elektronických informačních zdrojů a služeb umožňující získávání, zpracovávání, vyhledávání a využívání informací v tomto systému uložených*“ Její účel je v rámci této definice spatřován v „*poskytování možností jednotného přístupu k digitálním anebo digitalizovaným dokumentům*“ [35]. Ačkoliv vychází tato definice z více než dvě desetiletí starých zdrojů [36, 37], je v podstatě stále aktuální. Právě umožnění přístupu k digitálním objektům a vyhledávání v nich je i dnes nejčastěji uváděným důvodem pro budování digitálních knihoven. Jako další důležitý argument je také často zmiňováno dlouhodobé uchovávání významných kulturních artefaktů v digitální podobě. Současní autoři také ale stále častěji definici digitální knihovny rozšiřují takovým způsobem, aby zahrnovala nejen webovou stránku digitální knihovny, ale i veškeré další záležitosti s ní spojené včetně personálního zajištění [38, 39].

Přestože digitální knihovny jsou relativně nové entity, lze je také chápat jako elektronickou tvář tradiční knihovny. Funkce tradičních knihoven po celém světě jsou do jisté míry postaveny na práci indického matematika a knihovníka S. R. Ranganathana, který formuloval pět základních výroků o knihovnictví. Ty vešly ve známost jako Pět zákonů knihovní vědy. Tyto zákony zní následovně [40]:

1. Knihy jsou k užívání
2. Každému čtenáři jeho knihu
3. Každé knize jejího čtenáře
4. Šetřete čas čtenáře
5. Knihovna je rostoucí organismus

Ačkoli byla tato pravidla poprvé publikována v roce 1931 a původně se vztahovala jen k indickému kontextu, inspirovala se jimi většina knihoven po celém světě. S. R. Ranganathan znění svých zákonů později lehce změnil, aby měly univerzálnější podobu – slovo kniha nahradil slovem dokument, čímž docílil toho, že pravidla zahrnula širší spektrum informačních zdrojů [41]. Myšlenky stojící za těmito zákony bezpochyby stojí i v dnešním světě, důkazem může být například i fakt, že se těmito zákony inspirovaly nejen knihovny a dnes tak existuje například varianta Ranganathanových pěti zákonů i pro webové prostředí [42]. Není tedy důvod

z těchto historicky úspěšných pravidel nevycházet i při charakteristice a hodnocení digitálních knihoven.

Pátý zákon říkající, že knihovna je rostoucí organismus nelze hodnotit jinak než jako pravdivý. Důkazem, že tento zákon platí, je samotná existence digitálních knihoven. I digitální knihovny lze ovšem chápat jako rostoucí organismy, které se neustále vyvíjí. Nástin trendů a směřování digitálních knihoven proto přináší následující podkapitola 5.2.

V kontextu digitálních knihoven jsou nepochybně splněna i pravidla č. 2 a č. 3 („každému čtenáři jeho knihu“ a „každé knize jejího čtenáře“). Mezi základní charakteristiky a vlastnosti digitálních knihoven totiž patří fakt, že nemají žádné fyzické hranice a jsou dostupné 24 hodin denně. Navíc může mít stejný zdroj v jednu chvíli vypůjčeno větší (teoreticky neomezené) množství čtenářů. Lze tedy říct, že prostředí digitálních knihoven umožňuje získat každému čtenáři jeho knihu (potažmo informaci) velmi snadno. Smyslem pravidla „Každé knize jejího čtenáře“ je, že každá kniha má mít své místo v knihovně, a to i pokud by si ji vybralo ke čtení jen velmi malé množství čtenářů. Splnění tohoto pravidla je ideálním příkladem, proč digitální knihovny budovat. Zatímco tradiční knihovny jsou omezeny prostorem, do kterého svůj fond umisťují a musí tak vážit, zda danou publikaci zařadit do sbírky, přesunout do skladu (což snižuje její dostupnost), nebo úplně vyřadit, digitální knihovny se potýkají pouze s místem pro uložení dat. To má nesrovnatelně menší náklady, zásluhou čehož lze do digitálních sbírek zařazovat i dokumenty, které zaujmou jen malé množství čtenářů. [40, 39]

Zbývající dva zákony, č. 4: „šetřete čas čtenáře“ a především č. 1: „knihy jsou k užívání“, jsou zásadní z hlediska problematiky prezentace textů v digitálních knihovnách. Digitální knihovna totiž není jen kolekce elektronických materiálů, velmi důležité je i uživatelské rozhraní, které stojí mezi čtenářem a touto kolekcí. Není přitom výjimkou, že při navrhování těchto rozhraní se uživatelské zkušenosti a potřeby opomíjí [38]. Podkapitola 5.3 proto představí jednotlivé možnosti prezentace textů ve virtuálním prostředí a dále budou v podkapitole 5.4 charakterizována pravidla použitelnosti webu ve vztahu k digitálním knihovnám, jejichž dodržování zajišťuje, že použitý způsob prezentace dokumentů je uživatelsky dostatečně přívětivý a nezabraňuje či přímo neodrazuje uživatele od používání digitální knihovny.

5.2 Trendy v oblasti digitálních knihoven

V oblasti trendů týkajících se digitálních knihoven je možné v současnosti sledovat čtyři klíčová témata [43]:

- Agregace a s ní spojená interoperabilita dat
- Udržitelnost, dlouhodobá ochrana a identifikátory
- Otevřenost
- Zapojení uživatelů a personalizace

5.2.1 Interoperabilita a agregace

Prezentace dokumentů v digitálních knihovnách je důležitým způsobem pro rozšiřování přístupu ke kulturnímu dědictví jak mezi vědci a profesionály, tak i širokou veřejností. Digitální knihovny se nicméně potýkají se současným stavem informačního chování uživatelů internetu, který spočívá v získávání informací převážně skrze velké vyhledavače, jakým je například Google. Toto chování může představovat z hlediska digitálních knihoven problém, protože optimalizovat webový portál digitální knihovny pro indexování vyhledavači není snadné a může vyžadovat i poměrně nákladné investice, na které nemusejí mít menší digitalizační projekty zdroje. Druhým problémem je, že Google a další vyhledavače nedokáží indexovat takzvaný hluboký web (deep web) - tedy například takové webové stránky, které jsou dynamicky generované, nebo k jejichž zobrazení je třeba přihlášení uživatelským účtem. Hluboký web je podle některých odhadů i 500krát větší než ten povrchový, který je indexován. Velkou část informací tak nelze nalézt skrze běžné vyhledavače, což samozřejmě zahrnuje mimo jiné i materiály z digitálních knihoven. [43]

Jedním ze stále více používaných řešení této situace je agregace dat nebo metadat z jednotlivých projektových nebo institucionálních digitálních knihoven do větších, často nadnárodních portálů, které mohou uživatelům nahradit komerční vyhledavače a sloužit jako vstupní brána k více zdrojům. Takovou platformu představuje na evropské úrovni portál Europeana spuštěný v roce 2008, z dalších projektů potom například Digital Public Library of America, the National Digital Library of India, nebo Australská Trove. [44]

Agregace a interoperabilita dat pocházejících z různých digitálních knihoven je zajištěna používáním stejných standardů pro vytváření a publikování dat a metadat. O této oblasti může být napsáno mnoho, v této práci budou pouze představeny tři nejpoužívanější nebo nejdiskutovanější přístupy používané při agregaci dokumentů nebo metadat o nich

z jednotlivých digitálních knihoven. Mezi tyto tři vybrané přístupy patří protokol OAI-PMH, aplikační rozhraní IIIF a oblast propojených dat (Linked Data).

OAI-PMH

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) je jeden z nejstarších a nejběžněji používaných protokolů určených ke sklizení metadat, tedy k automatickému agregování metadatových záznamů z různých zdrojů. Tento protokol vytvořila organizace Open Archive Initiative v roce 2000. Protokol je navržen jako jednoduchý mechanismus pro snadnou interoperabilitu mezi repozitáři a je založen na běžných standardech používaných na webu – konkrétně protokolu HTTP a jazyku XML. [45, 46]

V rámci fungování OAI-PMH jsou rozeznávány dva druhy účastníků: Data Providers (zprostředkovatelé dat) tedy repozitáře, které zpřístupňují metadata, a Service Providers (zprostředkovatele služeb), kteří tato metadata sklízí, přenáší do své databáze a používají pro vytvoření vlastní služby. OAI-PMH může přenášet metadata v libovolném formátu, pokud je možné je zapsat v XML, nicméně je vyžadováno, aby součástí přenášených dat byla i data zapsána v Dublin Core [45]. Protokol OAI-PMH používá k agregaci dat do své sbírky například výše zmíněný projekt Europeana, do kterého tímto způsobem dodávají data mimo jiné Česká digitální knihovna [47] nebo Manuscriptorium [48].

Stále častěji se zvláště v kontextu digitálních knihoven objevuje názor, že OAI-PMH je zastaralou technologií, která není dostatečně efektivní, neboť spoléhá na kopírování dat na několik míst, není optimálně škálovatelná a zároveň je relativně náročná a drahá – instituce často musejí provozovat samostatný server pouze za účelem sklizení dat. V současnosti jsou proto zkoumány modernější a jednodušší přístupy k agregaci dat. [44]

IIIF

Zatímco OAI-PMH je protokol orientovaný na přenos metadat, IIIF⁴ (International Image Interoperability Framework) se zabývá přímo přenosem dat (konkrétně obrazů a souvisejících metadat) v rámci webového prostředí. Autoři IIIF považovali dosavadní možnosti využívané pro doručování obrazů v oblasti kulturního dědictví za velmi pomalé a příliš složité. Vzhledem k tomu, že přístup k obrazovým zdrojům pokládali za zásadní pro výzkum a studium, rozhodli se vytvořit komunitu, která by tento problém vyřešila. Od svého založení v roce 2011 se IIIF

⁴ Zkratka IIIF bývá vyslovována jako „Triple-Eye-Eff“

stalo globální sítí, ke které se připojily a na jejímž vývoji dále spolupracují renomované organizace po celém světě.

IIIF představuje sadu otevřených standardů, které pomáhají archivům, knihovnám a muzeím využívat jejich digitalizované sbírky, sdílet tyto sbírky nebo jednotlivé dokumenty a dále s nimi pracovat (například pomocí možnosti obrázky jednoduše zvětšovat, ořezávat nebo anotovat). Standardy jsou vyjádřeny v podobě specifikací API (Application Programming Interface, rozhraní pro programování aplikací). To umožňuje na jedné straně jednotlivým digitálním knihovnám vystavit své dokumenty nebo informace o nich a na straně druhé jednoduchou práci s těmito dokumenty mimo původní digitální knihovnu, včetně možnosti zařazení těchto dokumentů do své vlastní sbírky. Informace jsou v rámci IIIF API předávány ve výměnném datovém formátu JSON (JavaScript Object Notation). V současnosti udržuje konsorcium IIIF čtyři stabilní specifikace [49, 50]:

- *Image API*, které umožňuje interoperabilní doručování digitálních obrazů na základě vytvoření parametrizovaných identifikátorů (URI) sloužících k získání konkrétní oblasti anebo velikosti daného obrazu. Toto API také umožňuje získávat informace o obrazu a informace o tom, jaké manipulace je s ním povolené provádět.
- *Presentation API* umožňující přenos popisných metadat, která nejsou součástí dat přenášených v rámci Image API. Metadata jsou přenášena v rámci JSON dokumentu nazvaného „presentation manifest“, který zahrnuje všechny informace potřebné k představení celého dokumentu uživateli, včetně informací o souvisejícím Image API.
- *Authentication API* umožňuje použití IIIF pro zdroje s omezeným přístupem, zabezpečuje tedy fungování zmíněných API na základě propojení s existujícími autentizačními mechanismy dané digitální knihovny.
- *Content Search API* umožňuje prohledávání přepisů digitálních objektů generovaných pomocí OCR nebo například i po určité úpravě přepisů vytvořených v TEI.

Zároveň jsou již v beta testování další dvě specifikace: *Change Discovery API* zlepšující prohledatelnost zdrojových dokumentů a jejich indexaci velkými vyhledávači typu Google, a *Content State Transfer API*, jehož cílem je poskytnout standardizovaný formát pro sdílení konkrétního pohledu na jeden nebo více zdrojů IIIF Presentation API.

Standardy IIIF představují jednoduchý, flexibilní a nepříliš nákladný způsob přenosu a sdílení dat, přičemž se silně zaměřují na poskytování pokročilých interaktivních funkcionalit přímo koncovému uživateli – ten může díky IIIF API získávat rychle kvalitní obrázky, se kterými může

dále pracovat [49]. IIF také umožňuje rekombinovat různé zdroje a sestavovat tak nové kolekce bez ohledu na původní zdroj – z tohoto pohledu je tak možné IIF API použít i pro agregaci dat do větších platforem, ačkoli původně k tomu přímo nebyly tyto standardy navrženy [51].

Linked Data

Jiným stále častěji využívaným přístupem k interoperabilitě a svým způsobem i k agregaci jsou propojená data (Linked Data). Současnou podobu webu, včetně webových stránek digitálních knihoven a jejich obsahu, si lze představit jako síť propojených dokumentů. Problémem těchto dokumentů je, že z velké části jsou čitelné pouze pro člověka. Počítače je umí pouze zobrazit, ale ne číst. Tyto dokumenty ovšem popisují objekty skutečného světa a imaginární koncepty, mezi kterými existují vztahy. Pokud by byly tyto dokumenty obohaceny o data ve strojově čitelné podobě, mohly by být ony vztahy mezi objekty dokumentů využity. Tato myšlenka je známá jako „sémantický web“, ve kterém se vytvářejí vazby mezi informacemi takovým způsobem, kterému mohou počítače porozumět a interpretovat je. Výsledkem je, že informace o dokumentech a v dokumentech nejsou izolovány na jednotlivých webových stránkách, ale jsou propojeny s dalšími informacemi rozptýlenými na webu, ze kterých mohou počítače vytvořit komplexní záznam o daném objektu. Zjednodušeně lze říct, že použití propojených dat obohatí prezentované dokumenty v digitálních knihovnách a umožní uživatelům využít tento zdroj jako počáteční bod jejich bádání, neboť dokument bude propojen s dalšími informacemi mimo danou digitální knihovnu, přičemž stejně tak může jiný uživatel dospět k materiálům v knihovně přes propojená data z jiné webové stránky. [43, 52]

Propojování dat v zásadě spočívá ve vytváření identifikátorů věcí nebo prostředků na webu (pro identifikaci se používají URI, viz následující podkapitola 5.2.2) a následné propojení těchto zdrojů pomocí standardizovaného formátu RDF (Resource Description Framework). RDF popisuje vztah (predikát) mezi subjektem a objektem [53]. Například Osoba „Komenský“ může být pomocí predikátu „deathPlace“ propojen s objektem „Amsterdam“. RDF slouží pouze jako datový model, ne výměnný formát. Subjekty, predikáty a objekty tak mohou být v rámci RDF vyjádřeny například v XML, JSON nebo i HTML. Propojená data ve formátu JSON (takzvaném JSON-LD) se používají v rámci výše popsaných IIF API. Tripletty v RDF jsou dále specifikovány a propojovány pomocí ontologií [54], nicméně další rozbor této problematiky sahá, bohužel, mimo téma tohoto textu. Je ovšem nutné upozornit, že používání propojených dat k obohacování dokumentů prezentovaných v digitálních knihovnách, a dokonce i úvahy nad možnostmi jejich použití v rámci agregace sbírek, jsou velmi diskutovaným tématem [44].

5. 2. 2 Udržitelnost, dlouhodobá ochrana a identifikátory

Jak bylo předesláno v kapitole 4.3, digitální knihovny ve smyslu webových aplikací jsou svým způsobem pouze dočasné, neboť stejně jako veškeré další technologie podléhají neustálému zastarávání a proti němu jdoucímu vývoji. Nicméně vzhledem k tomu, že schraňují často unikátní artefakty kulturního dědictví je potřeba věnovat náležitou pozornost udržitelnosti a ochraně jak jejich obsahu, tak digitálních knihoven samotných. Udržitelnost digitálních knihoven zahrnuje několik oblastí – přes ekonomickou stránku věci (potřeba trvalého financování, vytvoření udržitelného obchodního modelu, sledování metrik) a etické a sociální aspekty (rovný přístup, problematika autorského práva, hodnotové ukotvení a relevance) až po problematiku týkající se uchovávání samotných digitálních materiálů. [43]

Dlouhodobé uchovávání digitálních objektů zahrnuje procesy, které zajišťují ideálně nekonečnou životnost elektronických dokumentů, ať již vznikly jako digitální, nebo byly vytvořeny v rámci digitalizace. Digitální informace mohou být snadno ztraceny kvůli celé řadě problémů, například technické poruše, nemožnosti přístupu k fyzickému médiu, na kterém jsou informace uloženy, nebo ztrátě softwaru, který informace interpretuje. Přístupy dlouhodobé ochrany lze v zásadě rozdělit na dvě části: jednu týkající se informační infrastruktury a druhou, týkající se přímo digitálních objektů. Na straně infrastruktury je ochrana zajišťována strategiemi jako jsou automatické, geograficky redundantní zálohy, kryptografické zabezpečení, audity a certifikace. Ochrana samotných digitálních objektů se potom zabývá problematikou jednoznačné identifikace, fixity (zajištění, že objekt neprodělal žádnou změnu, většinou poškození), autenticity (objekt je tím, čím tvrdí, že je), jeho ontologií (porozuměním jeho významu a povaze) a dalšími. Tato část textu se bude dále věnovat pouze problematice unikátních trvalých identifikátorů, které jsou významné jak z hlediska dlouhodobého uchování dokumentů, tak i výše zmíněné oblasti interoperability, agregace a propojování dat. [43, 55]

Obecně slouží identifikátory k jednoznačnému určení objektu v konkrétním prostředí. Uvažujeme-li digitální dokumenty v digitálních knihovnách, pak je tímto prostředím internetová síť. Nejčastějším způsobem identifikace dokumentů v rámci webu je jejich URL adresa. Ta neurčuje přímo daný objekt, ale pouze jeho umístění v rámci internetu (proto také název Uniform Resource Locator – jednotný lokátor zdroje). Vzhledem k tomu, že dokumenty mohou být v rámci internetu odstraňovány, přesouvány nebo měněny, není možné považovat URL za perzistentní identifikátor. Pro trvalou identifikaci dokumentů v rámci webu je nutný systém založený na globálním a trvalém identifikátoru označující přímo daný dokument, nikoliv pouze jeho umístění. Tuto funkci plní jiný internetový standard – identifikátor URN

(Uniform Resource Name). Ten byl navržen jako identifikátor, který má zůstat globálně jedinečný a trvalý, a to i ve chvíli, kdy prostředek, který označuje, přestane existovat nebo se stane nedostupným. Obecným pojmem zahrnujícím jak URL, tak URN je termín URI (Uniform Resource Identifier), který byl zmíněn v předchozí kapitole.

Podobu URL pravděpodobně není potřeba představovat, neboť se s ní setkal a každý den setkává jakýkoli uživatel webu. URN má odlišnou, přesně určenou syntaxi, danou zápisem: `<URN> ::= "urn:" <NID> ":" <NSS>`, kde `<NID>` označuje identifikátor jmenného prostoru (Namespace Identifier) a `<NSS>` (Namespace Specific string) definuje část identifikátoru jedinečnou v rámci daného jmenného prostoru. URN funguje tak, že je považováno za neprůhlednou URL adresu, která je interpretována tzv. resolversy – aplikacemi, které po zadání URN přesměrují uživatele na správnou URL nebo zobrazí informace o daném zdroji. V českém prostředí je takovým resolverem ČIDLO provozované Národní knihovnou ČR. ČIDLO slouží k přidělování a správě URN pro identifikaci digitalizovaných dokumentů, a to v rámci jmenného prostoru NBN, který je schváleným jmenným prostorem pro potřeby systémů národních bibliografií. Konkrétní příklad URN spravovaného v tomto systému může vypadat následovně: `urn:nbn:cz:aba007-0007cj`. [56, 57, 58, 59]

Vedle URI existuje několik dalších systémů pro trvalou identifikaci dokumentů, přičemž je možné je do jisté míry kombinovat (použít pro jeden dokument více identifikátorů, umožní-li to jejich metodika). Mezi nejčastěji používané trvalé identifikátory patří systém Handle provozovaný CNRI (Corporation for National Research Initiatives) a na něm založený systém DOI, který umožňuje identifikovat různé typy digitálních objektů a v současnosti je využíván především pro identifikaci vědeckých článků. Oba tyto systémy jsou zpoplatněné a pro registraci dokumentu vyžadují relativně velké množství metadat, tato (mimo jiné i časová) náročnost je častým důvodem jejich ne příliš rozšířeného využívání digitálními knihovnami. Jako další významný identifikátor lze jmenovat PURL, který v současnosti provozuje Internet Archive, a který funguje na bázi URL – registrovaný uživatel si nejprve vytvoří URL v systému PURL, a na tu následně naváže URL odkazovaného dokumentu. Při změně umístění je třeba tuto informaci v rámci PURL aktualizovat, nicméně původní PURL URL zůstává stejná. [43]

5. 2. 3 Otevřenost

S rozvojem informačních a komunikačních technologií a zpřístupněním obrovského množství různorodého obsahu online došlo k zásadním změnám v komunikaci a interakci lidí s počítači a v možnostech výměny, propojení a opakovatelného používání obsahu dostupného na internetu. Toto prostředí těží, jak bylo ukázáno i v textu o interoperabilitě a propojených datech, z otevřených systémů a snadného přístupu k digitálnímu obsahu bez překážek. Mezi aspekty otevřenosti je možné zařadit otevřené standardy, otevření zdrojových kódů (open source), otevření obsahu (open content) a otevření služeb (open services).

Důležitost otevřených standardů je všeobecně dobře známa. Umožňují maximalizovat přístupy ke zdrojům a službám, poskytují nezávislost na konkrétních aplikacích nebo zařízeních a umožňují tak dlouhodobý přístup ke zdrojům a službám, což je zejména v oblasti vědeckého bádání a kulturního dědictví zvláště významné [60]. Otevřené standardy jsou také (a především) základem interoperability, která byla popsána výše. Většina standardů, formátů a protokolů zmíněných v této práci jsou otevřené.

Vedle standardů se stává stále větším trendem i otevřený software, přesněji Free and Open Source Software (FOSS), tedy svobodný software s otevřeným kódem. Na rozdíl od softwaru s uzavřeným zdrojovým kódem umožňuje FOSS provádění úprav a přizpůsobení, která odpovídají specifickým potřebám uživatele nebo určité skupiny. Je zvykem, že autoři úprav v rámci určitého softwaru svůj upravený zdrojový kód dále zveřejňují, respektive že na úpravách spolupracují ve veřejném (virtuálním) prostoru. Právě vytvoření komunity okolo softwaru je považováno za jednu z výhod práce s otevřeným softwarem. Spolupráce většího množství lidí nebo celých institucí potom také vede k větší transparentnosti, spolehlivosti a bezpečnosti daného softwaru. Vývoj a používání softwaru s otevřeným kódem také zabraňuje tzv. vendor lock-in, tedy uzamčení zákazníka v digitálním ekosystému okolo proprietárního softwaru. FOSS bývá pochopitelně budován na základech otevřených standardů zmíněných výše. Je nutné podotknout, že slovo „Free“ v názvu Free and Open Source Software neznamena zdarma, nýbrž „svobodný“ – s vývojem, implementací a provozem FOSS jsou samozřejmě spojeny určité náklady, i tak ale bývá jeho pořízení a provozování výrazně levnější než komerční řešení [60, 61, 62]. Mezi FOSS řešení v oblasti digitálních knihoven patří například Greenstone [63], Omeka [64], eXist-db [65] nebo v Česku nejrozšířenější software pro digitální knihovny – Kramerius [66]. Otevřený software kráčí často ruku v ruce s otevřeným obsahem a službami, mezi které můžeme zařadit například uvolňování celých datasetů, nebo

zpřístupňování dat pomocí protokolů a API diskutovaných výše, jako jsou OAI-PMH, IIIF nebo RSS⁵ [60].

Představa otevřenosti, ať již standardů, softwaru, nebo i obsahu a služeb ovšem naráží na problematiku autorského práva, se kterým musí nejen (digitální) knihovny počítat a aspekty otevřenosti pečlivě vyvažovat. Jedním ze stále více akcentovaných přístupů řešení této situace, který se rozšiřuje zejména v akademické sféře, jsou licence Creative Commons [67]. Ty představují soubor veřejných licencí posilujících roli autora při rozhodování, za jakých podmínek má být jeho dílo zveřejněno, případně jak chce, aby bylo s jeho dílem dále nakládáno (zda je při jeho použití nutné uvádět jeho původ, zda a za jakých podmínek je možné ho dále zpracovávat, využívat komerčně a podobně). Použití těchto licencí podporuje šíření a další použití vytvořeného obsahu [68]. Právě v kontextu digitální vědecké práce, jako například s kolekcemi TEI dokumentů, je otevřené licencování důležitým nástrojem, který umožňuje opětovné použití dat a další výzkum, ale také zajišťuje udržitelnost výstupů výzkumných procesů [69].

5. 2. 4 Zapojení uživatelů a personalizace

Jak je patrné z předchozího textu, digitální knihovny dosáhly obrovského technického pokroku a stále pracují na lepších možnostech přístupu, sdílení a použití svého obsahu. S tímto pokrokem není zcela v rovině schopnost digitálních knihoven spolupracovat s komunitou, které slouží. Právě zapojení komunit přitom může významně zvýšit hodnotu digitální knihovny, a pomoci v koncipování knihovny jako sbírky online obsahu užitečného pro čtenáře a pomoci tak nasměrovat digitální knihovnu spíše k cíli soustředit se více na potřeby svých uživatelů než na potřeby dokumentů samotných. Tato oblast je v současnosti považována za klíčovou výzvu v tomto oboru.

V souladu s funkcemi a smyslem Webu 2.0⁶ jsou uživatelé digitálních knihoven stále častěji proměňováni z pouhých konzumentů obsahu na jeho (spolu)tvůrce. Jedním z možných přístupů je například využití pomoci a ochoty uživatelů digitálních knihoven ke korekci přepisů textových dokumentů získaných procesem rozpoznávání znaků – OCR (více o této technologii v podkapitole 5.3.1). Protože takto získaný přepis často není přesný, může tato dobrovolná

⁵ RSS (Really Simple Syndication) je webový zdroj, který umožňuje uživatelům a aplikacím přístup k aktualizacím na webových stránkách ve formátu XML. V digitálních knihovnách není v současnosti ve větší míře využíván. [44, 117]

⁶ Web 2.0 je označení pro etapu vývoje webových stránek počínající po roce 2000, nebo přímo pro označení takových webových stránek, které kladou důraz na obsah vytvářený uživateli, participaci, interoperabilitu a uživatelskou přívětivost [118]

lidská práce zlepšit kvalitu prezentovaných dat a zároveň zvýšit zájem o digitální knihovnu skrze aktivní participaci na jejím obsahu [70]. Uživatele lze zapojit nejen do oprav prostého textu, ale i například do anotace TEI dokumentů, ve kterých mohou pomoci mimo jiné s označováním jmen a názvů nebo číselných údajů [71]. Vedle oprav textů se nabízí využití čtenářů digitálních knihoven i k tvorbě dalších metadat, například k označování jednotlivých dokumentů pomocí značek (social tagging) určených k následnému indexování a vyhledávání [72].

Dalším trendem v oblasti potřeb uživatelů je personalizace služeb, tedy dynamické přizpůsobování systému digitální knihovny v závislosti na preferencích a cílech konkrétního uživatele. Typickým příkladem je personalizace obsahu: Digitální knihovny obvykle nabízí velké množství materiálů, ve kterých není snadné se zorientovat, nicméně za pomoci informací o preferencích, kompetencích a návštěvnické historii konkrétního čtenáře je možné dostupné materiály filtrovat a doporučovat uživateli v reálném čase takový obsah, který by ho mohl zajímat. Tímto mechanismem je možné pomoci s překlenutím propasti mezi množstvím dostupného obsahu a individuálními potřebami čtenářů. Nezbytnou podmínkou pro takovýto systém pro tvorbu návrhů (a personalizaci obecně) je ovšem shromažďování a uchovávání dat o každém uživateli, což je v určitém rozporu s jeho právy týkajícími se soukromí a anonymity. Je tedy nutné vždy pečlivě zvažovat benefity implementace takových systémů a dbát na obezřetné zacházení s uživatelskými daty. [73]

5.3 Způsoby prezentace textů

5.3.1 Prostý text

Velmi jednoduchým způsobem prezentace digitální verze knihy je zveřejnění prostého textu jejího přepisu. Přepis je nejčastěji vytvářen pomocí technologií OCR (Optical Character Recognition – Optické rozeznávání znaků) v případě tisků [74] anebo technologií HTR (Handwritten Text Recognition) pro rozpoznávání ručně psaného textu [75]. Tyto technologie dokáží převést obrazy obsahující text do formátu umožňujícího jeho editaci. Proces rozpoznávání znaků zahrnuje postupné algoritmy analyzující rozložení stránky, řádků a slov až po jednotlivé znaky a následně převedení obrazů každého znaku na příslušné kódy zvolené znakové sady. Protože rozpoznání obrazu některých znaků (například rozlišení „I“ a „l“ nebo „C a G“) může být složité, a to i vzhledem ke kvalitě výchozího materiálu, bývají získané výsledky ještě porovnávány se slovníkem.

Takto získaný přepis textu nebývá prezentován hierarchicky a ve webovém prohlížeči se nejčastěji čtenáři ukazuje jako dlouhý nestrukturovaný text. Elektronická kniha může být rozdělena na jednotlivé kapitoly, nicméně jemnější struktura většinou chybí a původní členění na stránky, kapitoly a podobně není dodrženo. Často bývá přítomno jen základní formátování – jsou odlišeny například nadpisy kapitol. V textu také mohou chybět obrázky, které obsahuje primární text. [38]

Prostý text nebo HTML dokument s jednoduchým formátováním používá jako své hlavní způsoby pro zveřejňování textů Projekt Gutenberg – příklad textu prezentovaného v této digitální knihovně ukazuje Obrázek č.3, více o digitální knihovně samotné přináší kapitola 6.6.

THE
FIRST BOOK OF FRANCIS BACON;
OF THE PROFICIENCE AND
ADVANCEMENT OF LEARNING,
DIVINE AND HUMAN.

To the King.

THERE were under the law, excellent King, both daily sacrifices and freewill offerings; the one proceeding upon ordinary observance, the other upon a devout cheerfulness: in like manner there belongeth to kings from their servants both tribute of duty and presents of affection. In the former of these I hope I shall not live to be wanting, according to my most humble duty and the good pleasure of your Majesty's employments: for the latter, I thought it more respective to make choice of some oblation which might rather refer to the propriety and excellency of your individual person, than to the business of your crown and state.

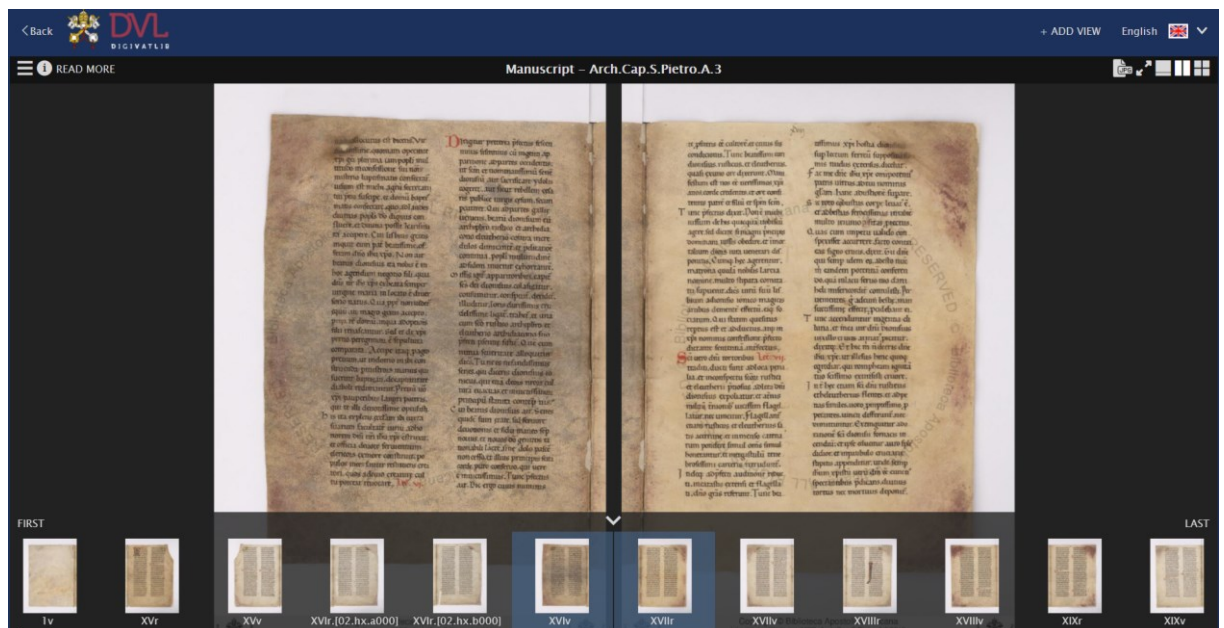
Wherefore, representing your Majesty many times unto my mind, and beholding you not with the inquisitive eye of presumption, to discover that which the Scripture telleth me is inscrutable, but with the observant eye of duty and admiration, leaving aside the other parts of your virtue and fortune, I have been touched—yea, and possessed—with an extreme wonder at those your virtues and faculties, which the philosophers call intellectual; the largeness of your capacity, the faithfulness of your memory, the swiftness of your apprehension, the penetration of your judgment, and the facility and order of your elocution: and I have often thought that of all the persons living that I have known, your Majesty were the best instance to make a man of Plato's opinion, that all knowledge is but remembrance, and that the mind of man by Nature knoweth all things, and hath but her own native and original notions (which by the strangeness and darkness of this tabernacle of the body are sequestered) again revived and restored: such a light of Nature I have observed in your Majesty, and such a readiness to take flame and blaze from the least occasion presented, or the least spark of another's knowledge delivered. And as the Scripture saith of the wisest king, "That his heart was as the sands of the sea;" which, though it be one of the largest bodies, yet it consisteth of the smallest and finest portions; so hath God given your Majesty a composition of understanding admirable, being able to compass and comprehend the greatest matters, and nevertheless to touch and apprehend the least; whereas it should seem an impossibility in Nature for the same instrument to make itself fit for great and small works. And for your gift of speech, I call to mind what Cornelius Tacitus saith of Augustus Cæsar: *Augusto profluens, et quæ principem deceret, eloquentia fuit.* For if we note it well, speech that is uttered with labour and difficulty, or speech that savoureth of the affectation of art and precepts, or speech that is framed after the imitation of some pattern of eloquence, though never so excellent; all this hath somewhat servile, and holding of the subject. But your Majesty's manner of speech is, indeed, prince-like, flowing as from a fountain, and yet streaming and branching itself into Nature's order, full of facility and felicity, imitating none, and inimitable by any. And as in your civil estate there appeareth to be an emulation and contention of your Majesty's virtue with your fortune; a virtuous disposition with a fortunate regiment; a virtuous expectation (when time was) of your greater fortune, with a prosperous possession thereof in the due time; a virtuous observation of the laws of marriage, with most blessed and happy fruit of marriage; a virtuous and most Christian desire of peace, with a fortunate inclination in your neighbour princes thereunto: so likewise in these intellectual matters there seemeth to be no less contention between the excellency of your Majesty's gifts of Nature and the universality and perfection of your learning. For I am well

Obrázek č. 3 Příklad textu prezentovaného v Projektu Gutenberg⁷

⁷ Zdroj obrázku: The Advancement of Learning, by Francis Bacon. In: *Project Gutenberg* [online digitální knihovna]. 1. 4. 2004 [cit. 2021-04-18]. Dostupné z: <https://www.gutenberg.org/files/5500/5500-h/5500-h.htm>

5. 3. 2 Skeny stránek

Druhou krajní možností prezentace elektronického textu v digitální knihovně je zveřejnění naskenovaných stránek originální tištěné verze. V podstatě je tak vytvořena a zveřejněna digitální faksimile – věrná reprodukce přibližující se co nejvíce obsahu, formě a vzhledu originálního dokumentu [76]. Tento přístup má několik výhod, ale přináší i komplikace. Zásadou zachování původní podoby a struktury působí text stejně organizovaně jako ve fyzickém originále, nijak se nemění jeho formátování a žádné části (především obrázky) nechybí. Jak bylo upozorněno výše, automaticky prováděný přepis textu pomocí OCR nebývá vždy přesný, a je-li elektronický text prezentován ve formě prostého textu, jsou vzniklé chyby předkládány i čtenářům. Zveřejnění skenů tento problém eliminuje. Na druhou stranu ale není v takovém textu možné jednoduše vyhledávat a čtenář je odkázán jen na listování podobně jako ve skutečné knize. Je také nutné počítat s tím, že obrázky stránek pořízené často ve velkém rozlišení zabírají na úložišti digitální knihovny výrazně větší místo než samotný text [38]. Pouhé skeny rukopisů jsou zveřejněny například digitální knihovně Vatikánské knihovny DigiVatLib – viz Obrázek č. 4.



Obrázek č. 4: Prezentace rukopisu v DigiVatLib⁸

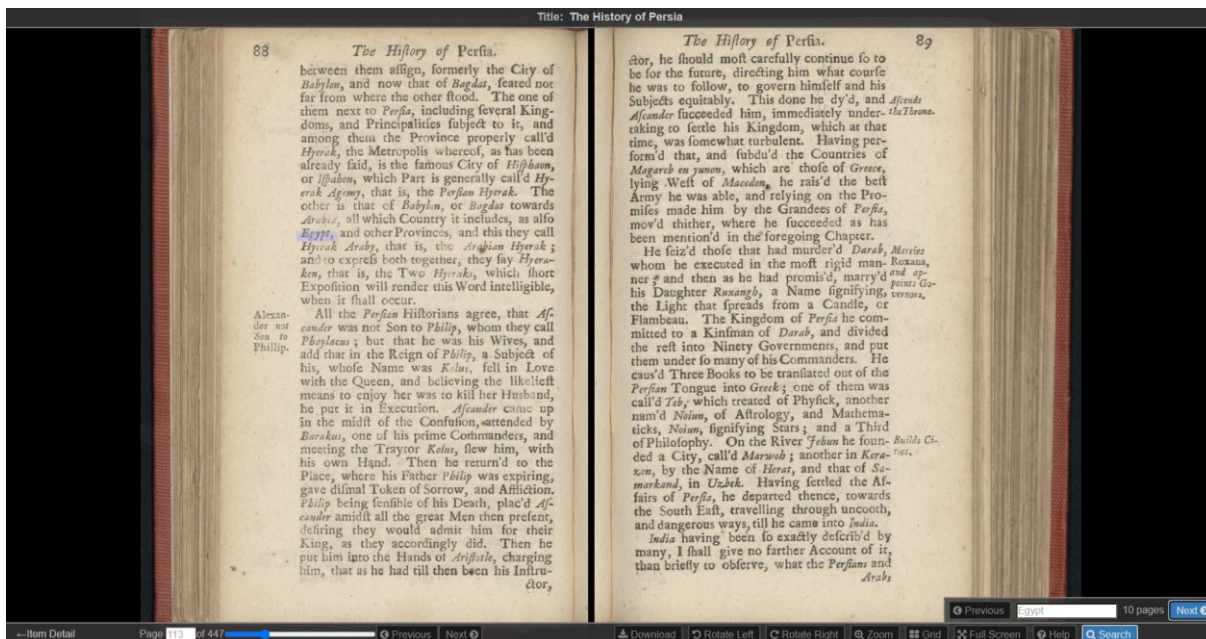
⁸ Zdroj obrázku: Manuscript - Arch.Cap.S.Pietro.A.3. In *DigiVatLib* [online digitální knihovna]. Vatican Library. [cit=2021-04-05]. Dostupné z: https://digi.vatlib.it/view/MSS_Arch.Cap.S.Pietro.A.3

5. 3. 3 Kombinace skenů a transkripce

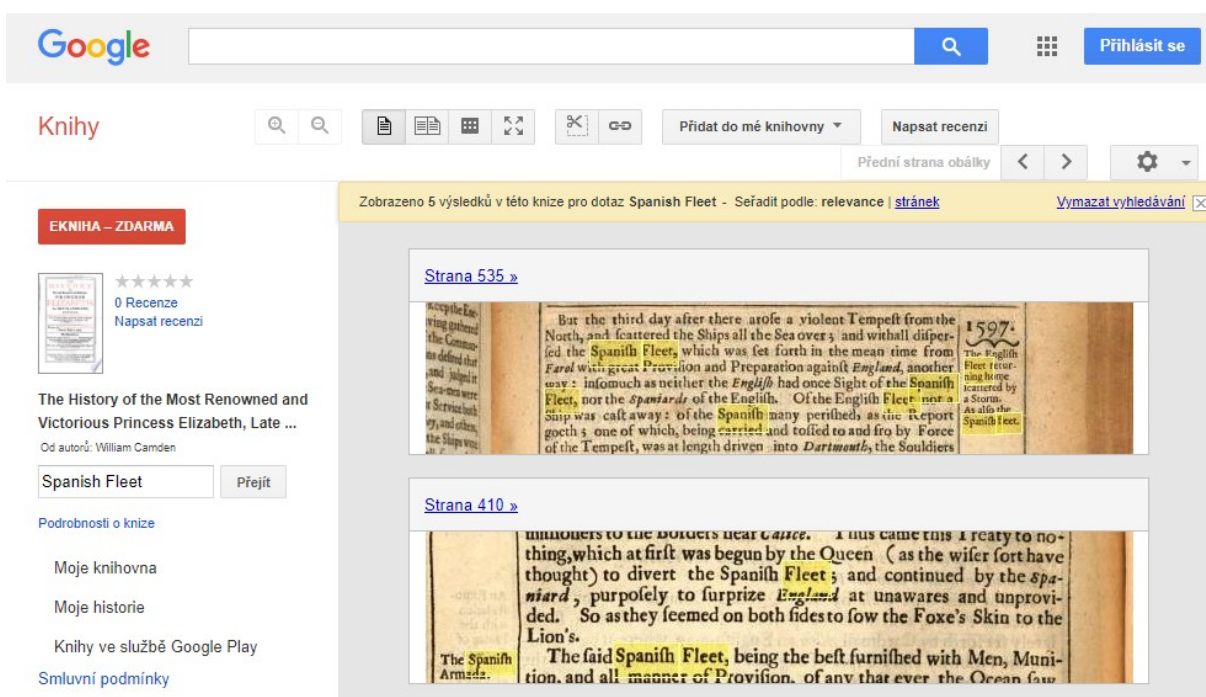
Řešení problému vyhledávání v textu, ale i některých pravidel použitelnosti a přístupnosti webových stránek, která budou zmíněna později, přináší zkombinování obou přístupů – tedy prezentace jak skenů, tak textového přepisu. Aby bylo v textu možné vyhledávat, je v procesu digitalizace vytvořen jeho přepis pomocí OCR nebo HTR, který ovšem není uživateli nijak prezentován a je použit pouze pro vytvoření indexu. Když později čtenář vyhledává určité slovo nebo frázi je tento index použit a daný termín je tak vyhledáván v plných textech všech dokumentů v digitální knihovně. Protože přepis textu většinou obsahuje kromě samotných slov i popis jejich umístění na stránce konkrétních knih, jsou uživateli na jeho dotaz zobrazeny konkrétní stránky, kde se jím hledané termíny vyskytují a hledaná slova jsou zvýrazněna. [38]

Tuto funkcionalitu nabízí například World Digital Library nebo Google Books. World Digital Library (WDL) je digitální knihovnou Knihovny Kongresu (národní knihovny Spojených států amerických), jejímž cílem je zveřejňovat zdarma v prostředí internetu digitální kopie dokumentů, které jsou významné z hlediska zkoumání různých kultur a národů. V současnosti obsahuje více než 3000 digitalizovaných knih a čísel časopisů a téměř 1700 rukopisů [77]. Pohled na její rozhraní nabízí Obrázek č. 5. Stejný způsob prezentace dokumentů používá i patrně největší digitální knihovna světa – Google Books, kde je v současnosti dostupných více než 40 milionů knih, přičemž všechny obsahují skrytý textový přepis použitelný pro plnotextové vyhledávání [78] (viz Obrázek č. 6).

Jiným způsobem řešení je připravení takového rozhraní digitální knihovny, které umožní uživateli přepínat mezi zobrazením skenů a samotného textu – uživatel si tak může sám vybrat, zda chce vidět sken dané stránky nebo její textový přepis. I tento způsob samozřejmě umožňuje používat transkripci textu pro vyhledávání. Tuto funkcionalitu nabízí například digitální knihovna HathiTrust. Obrázek č. 7 přináší srovnání obou pohledů na stejnou část textu. Podobnou možnost nabízí i Národní digitální knihovna používající systém Kramerius – na Obrázku č. 8 je vidět náhled textového přepisu stránky starého tisku.



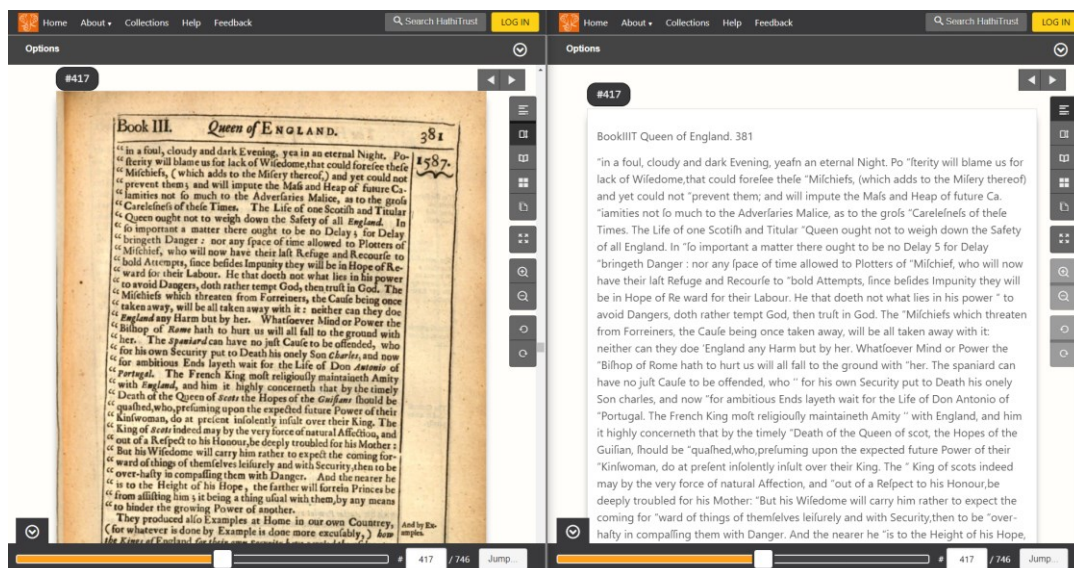
Obrázek č. 5: Pohled na prezentaci dokumentu ve World Digital Library⁹



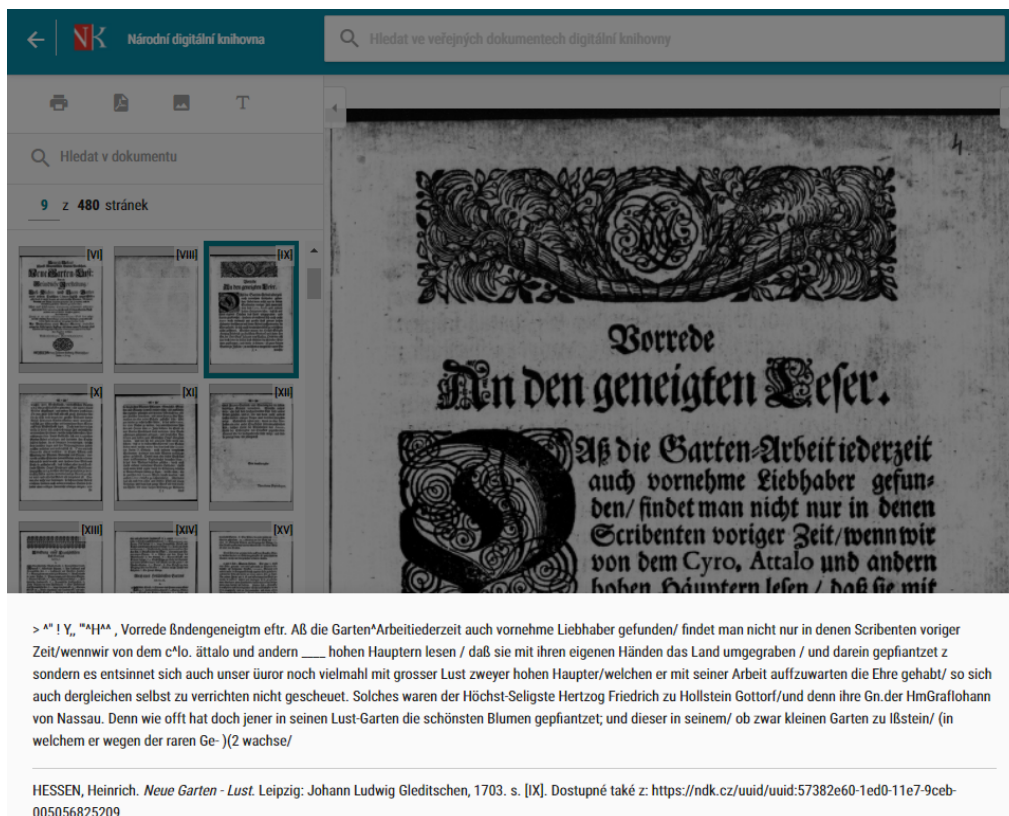
Obrázek č. 6: Pohled na vyhledávání v plném textu dokumentu v prostředí Google Books¹⁰

⁹ Zdroj obrázku: The History of Persia. In: *World Digital Library* [online digitální knihovna]. [cit. 2021-04-18]. Dostupné z: <https://www.wdl.org/en/item/2399/view/1/113/#q=Egypt>

¹⁰ Zdroj obrázku: The History of the Most Renowned and Victorious Princess Elizabeth, Late... In: *Google Books* [online]. [cit. 2021-04-18]. Dostupné z: https://books.google.cz/books?id=tayxm6f01kIC&printsec=frontcover&hl=cs&source=gbs_ge_summary_r&cad=0#v=onepage&q=Spanish%20Fleet&f=false



Obrázek č. 7: Prezentace dokumentu v digitální knihovně HathiTrust¹¹



Obrázek č. 8: Prezentace dokumentu v Národní digitální knihovně¹²

¹¹ The history of the most renowned and victorious Princess Elizabeth, ... In: *HathiTrust* [online digitální knihovna]. [cit. 2021-04-18]. Dostupné z:

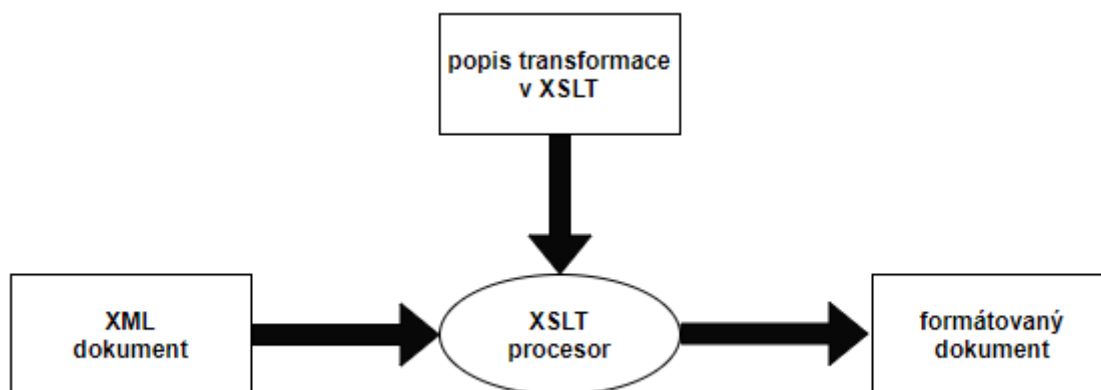
<https://hdl.handle.net/2027/uc1.31822042947556?urlappend=%3Bseq=411>

¹² Zdroj obrázku: HESSEN, Heinrich. *Neue Garten - Lust*. Leipzig: Johann Ludwig Gleditschen, 1703. s. [IX]. In: *Národní digitální knihovna* [online digitální knihovna]. Národní knihovna ČR. Dostupné z: <https://ndk.cz/uuid/uuid:57382e60-1ed0-11e7-9ceb-005056825209>

5. 3. 4 Vizualizace dokumentů kódovaných podle zásad TEI

Jak bylo zmíněno v kapitole 3.6, samotné TEI dokumenty mohou být doplněny o informace, popisující jejich vzhled. Je to možné učinit buď přidáním pravidel jazyka CSS přímo do dokumentu nebo odkazováním na dokument obsahující informace sloužící k vykreslování dokumentu. Takto vytvořený dokument může být následně relativně jednoduše prezentován ve webovém prohlížeči jako běžná webová stránka. Ačkoli dokument kaskádového stylu umožňuje relativně dobře řídit formátování obsahu všech elementů dokumentu XML, neumožňuje sofistikovanější úpravy zobrazení ani filtrování nebo přeskupení obsahu či případné dodání dalších informací, které mají být uživateli prezentovány. Proto se pro vizualizaci XML dokumentů nejčastěji používá výrazně komplexnější, ale zároveň flexibilnější způsob, a to jejich transformace pomocí XSL. [79]

XSL je nástroj vyvinutý konsorciem W3C, který umožňuje sofistikovanou práci se strukturou a obsahem XML dokumentů. Z jeho celého názvu (eXtensible Stylesheet Language, česky rozšiřitelný jazyk šablony stylů) je zřejmé, že jeho původním smyslem bylo sloužit k vizualizaci, nicméně v současnosti je kromě toho používán pro obecné transformace XML dokumentů do libovolného formátu nebo k extrakci dat. Transformace XML dokumentu jsou v XSL definovány pomocí pravidel v jazyce XSLT (eXtensible Stylesheet Language Transformation). Tato pravidla, nazývaná šablony (templates), se skládají ze vzoru a akce. Akce se provádí nad částí dokumentu, který vyhovuje vzoru pravidla. Vzorem je výraz v jazyce XPath, který slouží k výběru uzlu v dokumentu. Akce definuje požadované úkony, které mají být v rámci vybraného uzlu aplikovány – obecně vzato může akce obsahovat text, který má být výstupem, elementy výsledného dokumentu nebo další konstrukty definované v jazyce XSLT. Proces XSL zpracování má dva kroky: 1) XML dokument je pomocí XSLT zpracován a transformován na výsledný dokument, 2) dokument je interpretován – interpretem v případě transformace do HTML dokumentu je webový prohlížeč, v případě transformace v textový dokument je jím textový procesor a podobně. Výsledkem je formátovaný výstup předkládaný uživateli [80]. Tento proces znázorňuje Obrázek č. 9, Příklad č. 9 potom ukazuje, jak lze z jednoduchého TEI dokumentu (viz Příklad č. 3 na str. 22) vytvořit HTML dokument, přičemž v tomto příkladu je ze zmíněného TEI dokumentu úmyslně zobrazena jen první kapitola včetně jejího nadpisu.



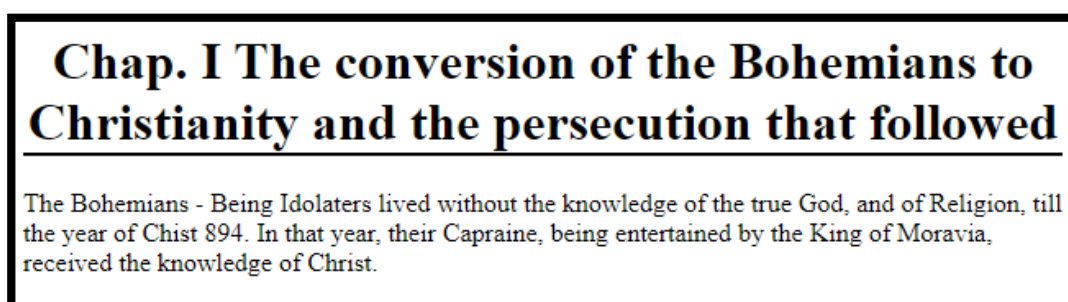
Obrázek č. 9: XSL transformace, [80], upraveno

```

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
version="1.0">

  <!-- toto je vzor pravidla -->
  <xsl:template match="/">
    <!-- po něm následuje definování akce -->
    <html>
      <body>
        <h1 style="text-align:center; border-bottom:2px solid black;
margin-bottom:20px">
          <xsl:value-of select="TEI/text/body/div1/head" />
        </h1>
        <p>
          <xsl:value-of select="TEI/text/body/div1/p" />
        </p>
      </body>
    </html>
  </xsl:template>
  
```

Příklad č. 9: kód XSL transformace¹³, jejíž výsledek zobrazuje Obrázek č. 10

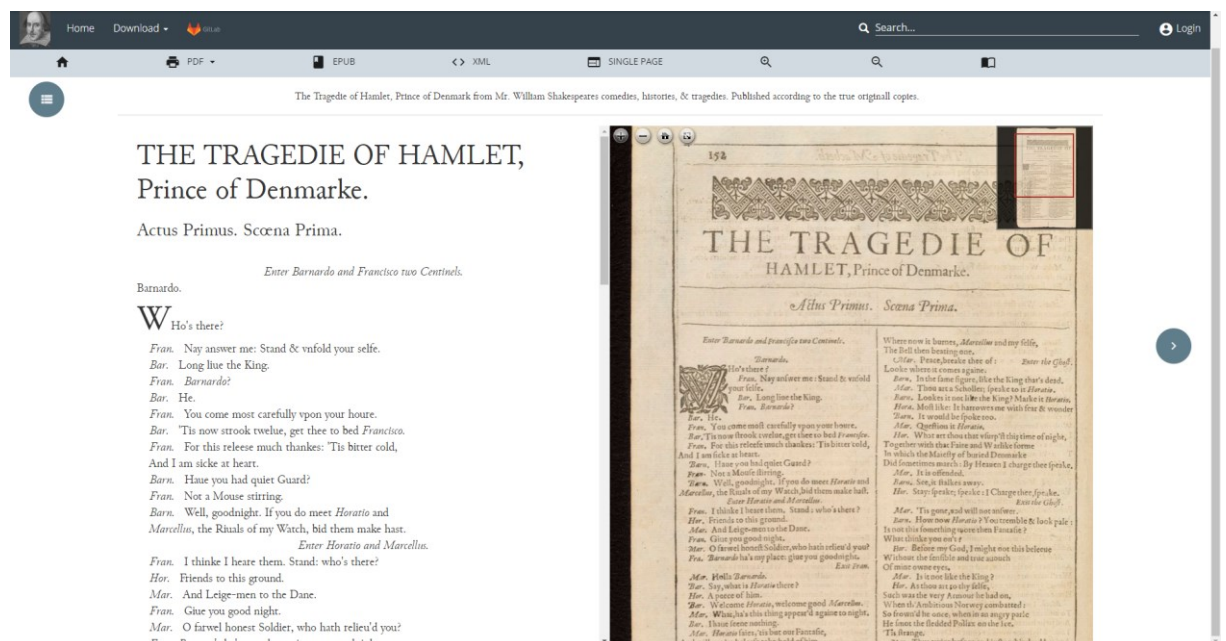


Obrázek č. 10: TEI dokument vizualizovaný pomocí XSL transformace

¹³ Jedná se o velmi jednoduchý příklad, i proto jsou CSS pravidla pro formátování textu definována přímo v rámci HTML elementů, v praxi by byly součástí samostatného dokumentu, na který by se náležitě odkazovalo.

Konsorcium TEI nabízí poměrně velké množství předpřipravených XSL transformací, které umožňují převody TEI dokumentů do nejpoužívanějších formátů: HTML, DOCX, ODT, Markdown, LaTeX, PDF, formát pro použití v systémech media wiki a další. Tyto šablony jsou udržovány ve veřejném repozitáři konsorcia na GitHubu [81].

Stejně jako v předešlých případech je často žádaným způsobem prezentace kombinace TEI se skeny primárního dokumentu. Za tímto účelem je možné využít velmi jednoduchých řešení, jako je například šablona pro XSL transformaci vyvinutá na katedře informačních studií a knihovnických věd Indriánské univerzity – TEI Boilerplate, která umožňuje převést styly definované v TEI dokumentu (viz kapitola 3.6 a atribut `@rendition`) na HTML dokumenty doplněné o skeny. Také je ale možné využít velmi sofistikovaných a komplexních systémů jako je eXist-db – otevřený dokumentový NoSQL XML databázový systém, který kromě transformací XSL a zobrazování obrázků vedle textu umožňuje i dotazování v XQuery (dotazovací jazyk XML) a mnoho dalšího [82, 83, 84]. Příklad dokumentu prezentovaného za použití eXist-db ukazuje Obrázek č. 11.



Obrázek č. 11: Příklad prezentace dokumentu za použití nástroje eXist-db¹⁴

¹⁴ Zdroj obrázku: The Tragedie of Hamlet, Prince of Denmark. In: *Shakespeare's Plays - TEI PM Edition* [online digitální knihovna]. [cit. 2021-04-18]. Dostupné z: <http://showcases.exist-db.org/exist/apps/shakespeare-pm/works/F-ham.xml?odd=shakespeare.odd>

5.4 Pravidla použitelnosti webu

Digitální knihovna, stejně jako ta klasická, nemá sloužit dokumentům, ale lidem. Proto je důležité, aby s ní čtenáři dokázali pracovat bez velkých problémů. V oblasti softwaru se jako faktor kvality, ukazující jak jednoduché je danou aplikaci používat, definuje takzvaná použitelnost. Přesněji ukazuje použitelnost to, jak rychle se lidé dokáží naučit používat konkrétní aplikaci, jak efektivní při tom jsou, jak zapamatovatelné jsou pro ně prováděné akce, jak často se dopouští chyb, jak vzniklé chyby řeší a také jak moc se uživatelům dané prostředí líbí a jak rádi ho používají. Půžitelnost tedy v podstatě popisuje kvalitu systému z pohledu jeho uživatelů, přičemž problémy s použitelností odkazují na aspekty, jejichž zásluhou je aplikace neúčinná, nefunkční nebo se s ní obtížně pracuje. [85, 86]

Vedle použitelnosti existují i další důležité kvalitativní atributy, které je potřeba brát v úvahu. Jedním z klíčových je například užitek – ten se vztahuje k funkčnosti designu a odpovídá na otázku, zda daná věc dělá to, co uživatel potřebuje. Půžitelnost a užitek jsou stejně důležité a společně rozhodují o tom, zda lze danou věc jako celek posoudit jako užitečnou. [87]

Ve webovém prostředí, jehož součástí jsou dnes i digitální knihovny, je použitelnost nutnou podmínkou. Uživatelé jsou zvyklí na určitý komfort a ve chvíli, kdy jim konkrétní webovou stránku přijde příliš složité používat, ji prostě opustí. S trochou nadsázky lze říct, že pokud nějakou funkci na webu lidé nechtějí nebo nemohou používat, nemusí vůbec existovat [87] – a bude-li tedy složité používat digitální knihovnu, je otázkou, zda plní svoji funkci sloužit čtenářům. Důraz kladený na použitelnost umožňuje lidem jednoduše a příjemně pracovat s technologiemi a zároveň napomáhá i dosahování obchodních cílů, čímž může podporovat dlouhodobou udržitelnost digitálních knihoven [86].

5. 4. 1 Klíčové principy webové použitelnosti

Použitelností se zabývá část mezinárodní normy ISO 9241 (Ergonomic Requirements for Office Work with Visual Display Terminals) [88] nicméně de facto standardem v této oblasti se stala práce Jakoba Nielsena [87], který definuje 5 základních komponent použitelnosti:

- **Naučitelnost** (*learnability*): to, jak snadné je pro uživatele uspět v provádění nějaké akce ve chvíli, kdy tuto akci provádí v rámci daného systému poprvé
- **Efektivita** (*efficiency*): rychlost a přesnost, se kterou dokáže uživatel dané úkony provádět ve chvíli, kdy je se systémem obeznámen
- **Zapamatovatelnost** (*memorability*): snadnost zapamatování funkcionalit systému a způsobu jejich používání tak, aby se příležitostný uživatel mohl do systému po určité době nepoužívání vrátit a nemusel se zacházení se systémem znovu učit
- **Chybovost** (*errors*): počet a frekvence chyb, které uživatel dělá, včetně toho, jak vážné jsou, jak rychle je možné je napravit a jak s tím systémem pomáhá
- **Spokojenost** (*satisfaction*): jak příjemné je používání daného systému

Tyto základní komponenty je možné dále rozkládat na jemnější kritéria a následně je ověřovat. Zároveň je nutné je blíže upřesňovat pro použití ve specifické doméně – například ve vztahu k webové použitelnosti je třeba vzít v úvahu hlavní úkoly webu, mezi které patří hledání požadovaných informací a služeb, a to jak přímým vyhledáváním, tak procházením obsahu, dále porozumění předkládaným informacím a vyvolávání a provádění služeb specifických pro určité webové aplikace [85]. V případě digitálních knihoven může být touto specifickou funkcí například prohlížení přepisu textu zobrazeného skenu stránky nebo stažení obrázku.

V rámci hodnocení webového uživatelského prostředí dále Nielsen definuje 10 heuristik, které budou krátce představeny [89]:

1) Viditelnost stavu systému

Uživatelské prostředí aplikace by mělo neustále uživatele informovat o tom, co se děje, a to bez zbytečného časového prodlení. Znalost aktuálního stavu systému může uživatelům pomoci pochopit výsledek jejich předchozích interakcí a určit další kroky. Zároveň je tím výrazně snižována frustrace uživatelů, kteří si nemusí být jistí, jestli systém jejich akce registruje. Je proto žádoucí, aby systém poskytoval zpětnou vazbu tak rychle, jak to jen jde, a to i ve chvílích, kdy nelze požadovanou akci dokončit okamžitě. Dobrým příkladem zpětné vazby je indikátor průběhu ukazující procenta.

2) Shoda mezi systémem a reálným světem

Design uživatelského prostředí by měl mluvit jazykem uživatele – použitá slova, fráze a koncepty by měly být uživateli známé. Jazykové prostředky je tedy nutné volit tak, aby odpovídaly specifitě komunity, pro kterou je systém vytvořen. Zároveň se doporučuje postupovat podle konvencí v reálném světě. Do tohoto rámce zapadá například i používání vhodných symbolů pro označení určitých akcí – například symbol lupy pro označení vyhledávání

3) Uživatel má mít kontrolu a svobodu

Uživatelé provádějí relativně často akce omylem, je proto nutné vždy nabízet a jasně označit „nouzový východ“. Tím se zajistí, že uživatel může nežádoucí akce snadno zanechat a vrátit se do předchozího stavu systému. To umožňuje uživateli mít aplikaci neustále pod kontrolou, vyvarovat se zaseknutí a pocitu frustrace, a naopak posílit v uživateli pocit svobody. Jednoduchou implementací tohoto konceptu je vždy přítomné tlačítko „zpět“.

4) Konzistence a standardy

Vzhledem k tomu, že lidé tráví většinu času používáním jiných produktů, než je ten, který uvažujeme (například konkrétní digitální knihovna), je nutné vždy uplatňovat oborové konvence. Uživatelé by neměli přemýšlet, zda různá slova a označení v různých situacích znamenají totéž. Konzistence je důležitá jak ve vztahu k jiným platformám, tak v rámci jedné aplikace. Nedodržení konzistence zvyšuje kognitivní zátěž uživatelů a nutí je učit se něco nového.

5) Předcházení chybám

Ideální je takové uživatelské prostředí, které aktivně zabraňuje vznikům chyb a problémů. Chyby jsou v zásadě dvojího druhu: Vědomé chyby (mistakes) založené na nesouladu mezi mentálním modelem uživatele a designu a nevědomé chyby (slips) způsobené nepozorností. Vědomým chybám lze předcházet odstraněním kognitivní zátěže a včasným varováním uživatelů. Řešit je může například tlačítko „zpět“. Nevědomým chybám je možné se vyhnout použitím dostatečných omezení a dobrými výchozími hodnotami.

6) Rozpoznávání před rozpomináním

Zviditelněním prvků, akcí a možností se minimalizuje zatížení paměti uživatelů. Uživatel by si neměl muset pamatovat informace z jedné části rozhraní do druhé – informace potřebné k použití by měly být vždy viditelné, případně alespoň snadno vyhledatelné. Vhodné je nabízení pomoci v kontextu právě vykonávaných akcí, místo poskytování dlouhého návodu k zapamatování. Dobrým příkladem je vyhledávací rozhraní, které prezentuje výsledky vyhledávání spolu se zadaným vyhledávacím dotazem.

7) Flexibilita a efektivita

System by měl umožňovat uživatelům provádět stejné akce různými způsoby. Například poskytování zkratk, které jsou začínajícím uživatelům skryty, mohou urychlit interakci zkušeného uživatele a systém jako celek tak může uspokojovat jak zkušené, tak začínající uživatele. Prostředí by mělo také disponovat možnostmi ovládní pomocí klávesových zkratk nebo gest v případě dotykových obrazovek. Je také vhodné, aby systém disponoval určitou mírou personalizace, a to jak obsahu, tak funkcionality. Kromě automatické personalizace je také důležité umožňovat taková přizpůsobení, aby si uživatelé mohli vybrat, jak chtějí, aby produkt fungoval.

8) Minimalistický design

Vizuální prvky uživatelského rozhraní by měly podporovat primární cíle uživatele. Prostředí by nemělo obsahovat informace, které jsou irelevantní nebo jen zřídka potřeby, protože každá další informace v rozhraní snižuje relativní viditelnost těch, které sledují hlavní obsah a funkci. Špatně organizované nebo chaotické rozhraní zvyšuje náklady uživatelů na interakci s obsahem.

9) Rozpoznávání a opravování chyb

Dojde-li k chybě, měly by být všechny chybové zprávy psány prostým jazykem a nikdy by neměly obsahovat kódy chyb či částí programového kódu. Zpráva by měla přesně označit problém a navrhnout jeho řešení. Doporučené je používat tradiční vizuální styl chybových zpráv – například červený text. Typickým příkladem správného řešení je poskytnutí užitečné pomoci ve chvíli, kdy nejsou na uživatelský dotaz nalezeny žádné výsledky – vhodnou pomocí může být navržení jiných vyhledávacích slov nebo oprava překlepů ve vyhledávacím dotazu.

10) Pomoc a dokumentace

Nejlepší variantou je, když uživatelské prostředí žádné vysvětlení nepotřebuje. U složitějších aplikací ale bývá vhodné poskytovat dokumentaci, která pomůže uživatelům pochopit, jak se systémem pracovat, a především jak efektivně dokončit své úkoly. Obsah dokumentace by měl být zaměřený na uživatele a jeho možné akce, měl by být stručný a strukturovaný do jednotlivých kroků, které jsou pro dokončení úkonu nutné. Dokumentace by měla být snadno prohledávatelná a ideálně prezentována (respektive její vhodné části) v okamžiku, kdy ji uživatel může potřebovat. Dobrým řešením jsou také seznamy s často kladenými otázkami a odpověďmi na ně (FAQ).

Vedle Nielsenových pravidel jsou často citována také pravidla pro použitelnost podle Schneidermana a Plaisant [90], ta zní následovně:

- Usilujte o konzistenci (odpovídá Nielsenovu 4. pravidlu)
- Zajistěte univerzální použití (odpovídá Nielsenovu 4. pravidlu)
- Nabízejte informativní zpětnou vazbu (odpovídá Nielsenovu 1. a 9. pravidlu)
- Navrhněte postupy v rámci úkolů tak, aby bylo snadné dosáhnout uzavření (odpovídá Nielsenovu 7. a 8. pravidlu)
- Předcházejte chybám (odpovídá Nielsenovu 5. pravidlu)
- Povolte snadné vrácení provedených akcí (odpovídá Nielsenovu 3. pravidlu)
- Dodejte uživatelům pocit, že mají vše pod kontrolou (odpovídá Nielsenovu 3. pravidlu)
- Minimalizujte zatížení krátkodobé paměti (odpovídá Nielsenovu 6. pravidlu)

Dále bývá ve vztahu k použitelnosti zmiňována práce Dona Normana, který definoval několik univerzálních principů designu: viditelnost (visibility), zpětnou vazbu (feedback), omezení (constraints), mapování (mapping), konzistenci (consistency) a využitelnost (affordance) [91].

Všechna zmíněná pravidla a principy se z velké části překrývají, což není náhoda – všechna totiž do jisté míry vychází z lidské psychologie, konkrétně toho, jak lidé vnímají, učí se, uvažují, pamatují si a převádějí záměry na činy. Mnoho autorů věnujících se tvorbě pravidel pro použitelnost mají své původní akademické zázemí právě v psychologii [90].

Použitelnost je také velmi úzce spojena s konceptem webové přístupnosti (web accessibility). Přístupný web je takový web, který je navržen a vyvinut tak, aby jej mohly plnohodnotně používat osoby se zdravotním postižením. Přístupnost se věnuje všem postižením, která nějakým způsobem ovlivňují přístup na web, včetně sluchových, kognitivních, neurologických, fyzických, zrakových i řečových. W3C Web Accessibility Initiative definuje konkrétní pravidla, která mají být splněna, aby byla daná webová stránka přístupná a umožnila tak všem bez rozdílu web vnímat, rozumět mu, procházet ho a pracovat s ním včetně přispívání na něj. Splnění těchto pravidel je prospěšné nejen pro hendikepované, neboť zlepšuje možnosti práce s webem, ale i pro uživatele, kteří používají pro navštěvování internetu starší počítače nebo mobilní telefony, mají špatné připojení k internetu, jsou důchodového věku, mají dočasný hendikep nebo jsou v nějak omezující situaci (nemohou momentálně poslouchat zvuk, jsou na přímém slunečním světle a podobně). [92]

5. 4. 2 Metody hodnocení použitelnosti

Úroveň použitelnosti webových aplikací a stránek bývá posuzována především uživatelským testováním. To je považováno za nejefektivnější způsob hodnocení použitelnosti z pohledu uživatelů, je ale časově i finančně poměrně nákladné. Vedle testování se proto využívají i evaluační metody, nejčastěji heuristická analýza a kognitivní průchod (cognitive walkthrough). Vedle těchto metod je možné některé aspekty použitelnosti analyzovat i za pomoci záznamů o přístupech na webovou stránku (logů), ve kterých je možné hledat například vzorce chování při navigování po stránce nebo odhalit, ve kterých chvílích uživatel web opustil. [85, 93, 94]

Uživatelské testování

Cílem uživatelského testování je zjistit chování uživatelů určitého produktu pozorováním reprezentativního vzorku skutečných uživatelů při provádění předem definovaných úkolů. Testovat je možné za použití již hotové aplikace ale i prototypů. V průběhu testování jsou zaznamenávána data zejména o času potřebném pro dokončení daných úkolů, počtu chyb a spokojenosti jednotlivých uživatelů. Sebraná data jsou následně analyzována a použita k vylepšení použitelnosti. Aby se předešlo nespolehlivým a zkresleným výsledkům, měl by být test pečlivě naplánován, včetně jasné definice cílů testu, vzorku uživatelů, úkolů a způsobů vyhodnocování. [85, 95]

Heuristická analýza

Heuristická analýza je metoda, při které odborník nebo ideálně skupina odborníků hodnotí danou aplikaci podle předem připraveného seznamu uznávaných principů použitelnosti – heuristik. Při analýze prochází vždy výzkumník uživatelské rozhraní dvakrát. Nejprve aby získal ucelený pohled na aplikaci, porozuměl jejímu rozsahu a možnostem interakce. V druhém kroku se již zaměřuje na funkčnosti a hodnotí návrh a implementaci podle seznamu heuristik. Výstupem hodnocení je seznam problémů, často prezentovaný formou tabulky. Heuristické analýzy je vhodné provádět zejména v situacích, kdy jsou výrazně limitované časové i finanční zdroje a je třeba získat relevantní výsledky bez zapojení skutečných uživatelů. Hodnocení může v zásadě provádět jen jeden výzkumník, ale výhodnější je, pokud se ho účastní několik nezávislých odborníků, jejichž zjištění jsou následně agregována. [85, 96]

Kognitivní průchod

Metoda kognitivního průchodu simuluje proces řešení problémů uživatele, tedy to, co uživatel dělá v konkrétních situacích, jak to dělá a proč. Hodnotitelé prochází rozhraní aplikace krok za krokem za pomoci scénáře úlohy, který popisuje cíle, jichž chce teoretický uživatel dosáhnout. V průběhu experimentu diskutují o problémech použitelnosti, na které v aplikačním rozhraní narazí. Tato metoda je zvláště vhodná k detekci problémů v oblasti naučitelnosti ovládnání aplikace. [85]

6 Analýza vybraných digitálních knihoven

V praktické části bakalářské práce budou analyzovány vybrané digitální knihovny, a to ve vztahu k metodám a principům popsaných v literárně přehledové části. Konkrétně bude zkoumán jejich přístup ke standardům používaných pro kódování textů, k současným trendům v oblasti digitálních knihoven a také základním principům použitelnosti webu.

K analýze bylo vybráno 7 digitálních knihoven, a to tak, aby ve vzorku byly zahrnuty jak knihovny obecné, obsahující velké množství digitalizovaných materiálů (nejen předmoderních textů), tak specifičtěji zaměřené platformy. Vybrané digitální knihovny jsou:

- Deutsches Textarchiv¹⁵
- e-codices – Virtual Manuscript Library of Switzerland¹⁶
- Manuscriptorium¹⁷
- Medieval Nordic Text Archive¹⁸
- Oxford Text Archive¹⁹
- Project Gutenberg²⁰
- Religiöse Friedenswahrung und Friedensstiftung in Europa 1500–1800²¹

Rozbor digitálních knihoven bude proveden na základě kontrolního seznamu. Ten je rozdělen na tři části:

První část se zabývá způsobem prezentace textů v digitálních knihovnách, tedy toho, zda jsou texty prezentovány ve formě skenů stránek, prostého textu, kombinací těchto přístupů nebo například vizualizací TEI dokumentu. Zároveň bude zjišťováno, zda je prezentovaný text kódován v TEI, do jaké hloubky je kódován a zda je v digitálních knihovnách dostupných více variant jednoho textu – například jedna varianta obsahující přesný přepis originálního

¹⁵ *Deutsches Textarchiv* [online digitální knihovna]. Berlin: Brandenburgische Akademie der Wissenschaften, 2007- [cit. 2021-5-2]. Dostupné z: <https://deustextarchiv.de>

¹⁶ E-codices: Virtual Manuscript Library of Switzerland [online digitální knihovna]. swissuniversities [cit. 2021-5-2]. Dostupné z: <https://www.e-codices.unifr.ch/en>

¹⁷ Manuscriptorium: Digital Library of Written Cultural Heritage [online digitální knihovna]. Národní knihovna České republiky [cit. 2021-5-2]. Dostupné z: <http://www.manuscriptorium.com/cs>

¹⁸ Medieval Nordic Text Archive: Menota [online digitální knihovna]. 2007- [cit. 2021-5-2]. Dostupné z: <https://www.menota.org/forside.xhtml>

¹⁹ Oxford Text Archive [online digitální knihovna]. University of Oxford, 2019 [cit. 2021-5-2]. Dostupné z: <https://ota.bodleian.ox.ac.uk>

²⁰ *Project Gutenberg* [online digitální knihovna]. [cit. 2021-5-2]. Dostupné z: <https://www.gutenberg.org/>

²¹ *Religiöse Friedenswahrung und Friedensstiftung in Europa 1500–1800: Digitale Quellenedition frühneuzeitlicher Religionsfrieden* [online digitální knihovna]. [cit. 2021-5-2]. Dostupné z: <https://www.gutenberg.org/>

textu a druhá s převedením některých znaků či slov tak, aby odpovídaly současnému jazyku a jeho gramatice.

Druhá část kontrolního seznamu zjišťuje základní informace z oblasti interoperability, konkrétně zda mají prezentované dokumenty přidělen perzistentní identifikátor, zda a jakým způsobem jsou dokumenty doplněny o metadata a také jestli je knihovna agregována nebo je sama agregátorem dokumentů, například prostřednictvím protokolu OAI-PMH.

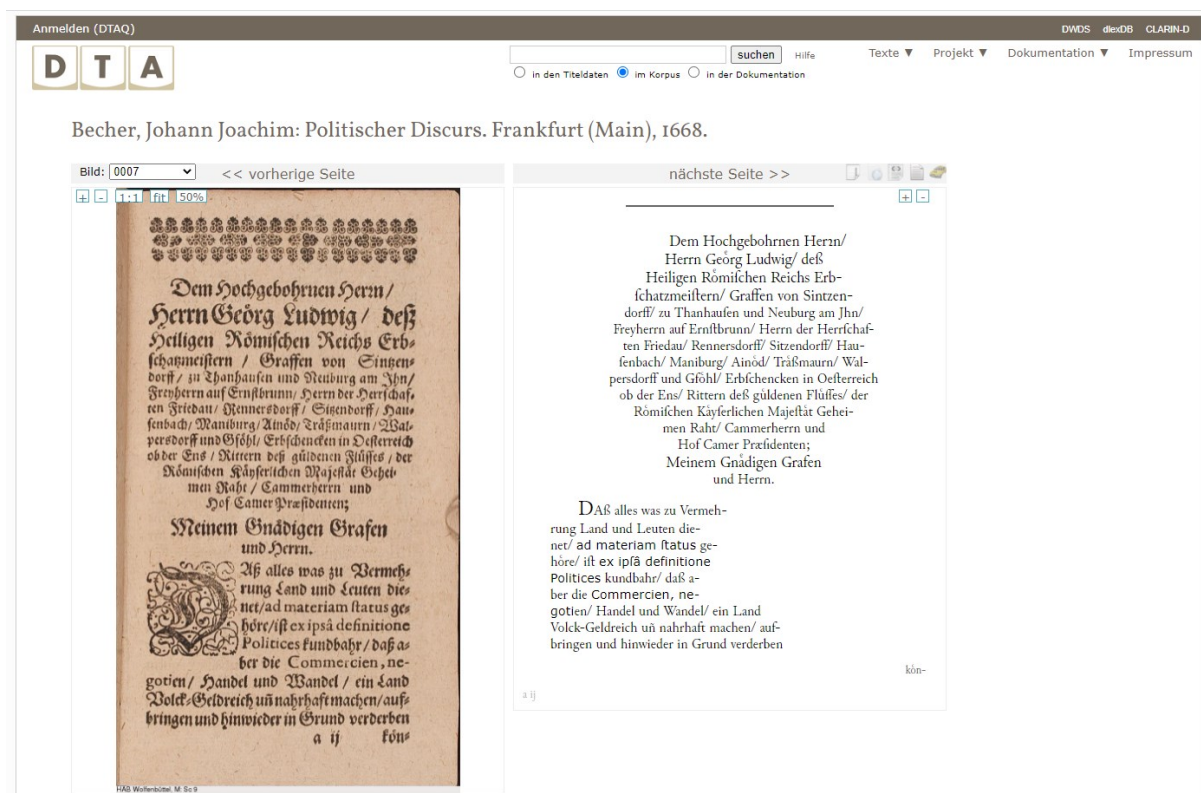
Poslední třetí část kontrolního seznamu se věnuje pravidlům webové použitelnosti, které budou posuzovány za pomoci heuristik J. Nielsena uvedených v kapitole 5.4.1. Výstupem této části bude v souladu s metodou heuristické analýzy představené tamtéž především tabulka shrnující objevené nedostatky. Předmětem zkoumání nebudou pravidla webové přístupnosti, jejichž splnění je předpokládáno. Celý kontrolní seznam je součástí Přílohy č. 1.

6.1 Deutsches Textarchiv

Deutsches Textarchiv (DTA) – Německý textový archiv – představuje historický referenční korpus pokrývající široké spektrum německého jazyka, oborů a žánrů. Jádrem digitální knihovny je korpus obsahující přibližně 1500 textů z období mezi lety 1600-1900, které mají mimořádný význam pro rozvoj německého jazyka anebo příslušných žánrů a akademických oborů. Vedle tohoto jádra vytvořeného v letech 2007-2016 obsahuje Deutsches Textarchiv také velké množství dalších významných textů z období od poloviny 15. století do 20. století. DTA má ambici se odlišovat od jiných digitalizačních projektů (jako je Project Gutenberg, nebo Google Books) pečlivým výběrem textů a edic, velmi vysokou přesností textových přepisů, jejich kódováním a spolehlivostí metadat. V současnosti obsahuje celkem téměř 6500 textů. [97]

6.1.1 Presentace textu

Dokumenty jsou uživatelům digitální knihovny prezentovány v prostředí kombinujícím naskenované stránky a textový přepis dokumentu (viz Obrázek č. 12). Pohled na přepis textu je možné přepínat mezi několika způsoby zobrazení (více dále). Možné je také zobrazit vedle skenu samotný TEI XML dokument, ze kterého jsou jednotlivé varianty zobrazení textu generovány. Přepisy vznikají procesem automatického OCR, ale jsou následně ručně kontrolovány a chyby v nich jsou opravovány. Při přípravě plných textů se také používají postupy automatické tokenizace a lemmatizace. Veškeré soubory – tedy TEI XML dokument, HTML dokument konkrétního pohledu na text, prostý text i samotný sken – je možné z digitální knihovny stáhnout a uložit do svého počítače.



Obrázek č. 12 Prostředí pro prohlížení textu, *Deutsches*²²

Kódování textů v DTA je provedeno pomocí vlastního kódovacího schématu nazvaného DTABf, které je postaveno na Zásadách TEI-P5. DTABf představuje podmnožinu značek dostupných v TEI-P5, sada kódovacích značek tedy nebyla nijak rozšířena, naopak jedná se vlastně o výběr vhodných elementů z TEI-P5 včetně definování přesných atributů a hodnot, kterých mohou nabývat. Schéma DTABf je tak plně kompatibilní se zásadami TEI a texty kódované v TEI mohou být proto snadno integrovány do infrastruktury DTA. Zároveň je možné texty v DTA opětovně použít v jiných textových archivech. Schéma DTABf navíc doplňují pokyny pro přepis textu, které zajišťují soudržnost strukturování textu v celém digitálním korpusu. TEI XML dokumenty jsou, jak bylo uvedeno výše, dostupné ke stažení, dostupný je i ODD soubor popisující DTABf schéma.

Kódování dokumentů odpovídá přibližně třetí základní úrovni kódování, tedy úrovni, kdy je text kódován na rovině logické struktury. Jsou označeny odstavce, řádky, poznámky, nadpisy, záhlaví, případně obrázky. Text je také doplněn elementy `<hi>` a atributy určujícími

²² Zdroj obrázku: Becher, Johann Joachim: Politischer Discurs. Frankfurt (Main), 1668. In: *Deutsches Textarchiv* [online digitální knihovna]. Berlin: Brandenburgische Akademie der Wissenschaften. [cit. 2021-05-05]. Dostupné z: https://www.deustextarchiv.de/book/view/becher_discurs_1668?p=7

vykreslování. Mimo to jsou opravovány evidentní chyby v textu, a to pomocí elementu `<choice>` obsahujícího vždy označenou chybu (`<sic>`) a její opravu (`<corr>`).

Pro kódování tisků se podle metodiky DTA používají všechny povinné moduly (*core*, *header*, *tei* a *textstructure*) a několik modulů volitelných, konkrétně *corpus*, *drama*, *figures*, *linking*, *msdescription*, *namesdates*, *textcrit*, *transcr*. Některé moduly jsou ovšem více či méně omezené DTABf schématem. Například z modulu *header* je takto rovnou vyškrtnuto používání elementu `<handNote>` pro popis konkrétního stylu nebo písma používaného v rukopisu, obdobně je element `<handDesc>` odebrán z modulu *msdesctiprion*. Dále jsou upraveny některé atributy, respektive jsou striktně omezeny hodnoty, kterých mohou nabývat. Například pro element `<note>` modulu *core* sloužícího pro zápis poznámek v textu, je definován uzavřený seznam pro popis místa, kde se poznámka v textu nachází (`@place`). Obdobně je uzavřený seznam definován u několika dalších atributů, například atributu `@type` elementu `<bibl>` pro bibliografické citace. Změněny jsou také některé definice elementů v ODD dokumentu, a to většinou na německé ekvivalenty anglických termínů, pravděpodobně aby kodéři lépe rozuměli významu značek. U poměrně velkého množství elementů (mimo jiné těch, které označují jména osob nebo míst) je odstraněn atribut `@resp` sloužící k označení konkrétní osoby, která je za použití tagu zodpovědná. Pro popis rukopisů existuje speciální verze DTABf schématu, její úpravy jsou ale víceméně podobné, pouze je jich méně a převážně se jedná opět o uzavření seznamů pro výběr hodnot atributů a především rozšíření a přeložení popisů elementů, atributů a jejich hodnot do němčiny. [98, 99]

Pohled na textový přepis je možné přepínat mezi různými způsoby zobrazení: Na výběr je prostý text, standardizované („normmierte“) a originální zobrazení („originalzeichen“). Originální zobrazení umožňuje prohlížení textu obsahujícího původní znaky (například malé „s“ ve slově „Wiffenschaft“), zatímco standardizované zobrazení používá znaky současné latinky (tedy například „Wissenschaft“). Nejedná se ovšem o dvě varianty transkripce textu, ale pouze o funkci vykreslování. V TEI XML dokumentu jsou vybrané znaky, jejichž zobrazení může být měněno, zapsány pomocí HTML hexadecimálních kódů, který je potom podle požadovaného zobrazení odlišně interpretován XSL transformací. Slova „Wiffenschaft“ a „Wissenschaft“ jsou tak v TEI XML dokumentu zapsána pouze jednou, a to jako `Wiſſenſchaft`.

6. 1. 2 Interoperabilita

Všem dokumentům je přiřazen trvalý identifikátor URN:NBN se jmenným prostorem přiděleným Berlínsko-Brandenburgské kooperativní síti knihoven - `urn:nbn:de:kobv`. Metadata jednotlivých titulů je možné kromě podoby kompletních TEI dokumentů získat jednoduše i jako XML soubory obsahující pouze záhlaví TEI, případně i ve formátech Dublin Core a CMDI²³. Metadata zahrnutá v záhlaví TEI je možné sklízet přes protokol OAI-PMH. [100]

6. 1. 3 Webová použitelnost


Uživatelské prostředí Deutsches Textarchiv je poměrně jednoduché a přehledné. Skládá se z úvodní obrazovky, která obsahuje poměrně široké informace o digitální knihovně, použitých řešeních a novinek týkajících se knihovny nebo souvisejících oblastí. Nahoře na úvodní obrazovce je vyhledávací řádek, obsah digitální knihovny je ale také možné procházet podle kategorií dokumentů nebo doby jejich vzniku. K tomu slouží pravé postranní menu. Z vyhledávání je možné přecházet na detaily jednotlivých titulů (Obrázek č. 13) nebo přímo do prohlížečky textů (Obrázek č. 12).

²³ CMDI je zkratka pro Component MetaData Infrastructure, metadatový formát vytvořený Evropskou výzkumnou infrastrukturou pro jazykové zdroje a technologie – CLARIN. CMDI poskytuje rámec pro popis a opětovné použití šablon metadat, díky čemuž je možné jednotně vyjádřit metadata jinak uzavřená v doménově specifickém standardu (jako je například záhlaví TEI pro texty a IMDI pro multimediální sbírky) [119]

Anmelden (DTAQ) DWDS [dlexDB](#) [CLARIN-D](#)

DTA Hilfe

In den Titeldaten im Korpus In der Dokumentation



Text-Bild-Ansicht öffnen ...

Adolph, Christian: Daktulion (he)pomnematikon. Breslau, 1641.

BIBLIOGRAPHISCHE ANGABEN

URN: urn:nbn:de:kobv:b4-200054-4
 Titel: Daktulion (he)pomnematikon (gr.) sive Annulus monitorius [...] Geistlicher Gedend-Ring/ mit zweyen zusammen-gefügeten Händen Da die Eine schlägt: Die Ander trägt. Zu wolverdientem Christlichem Ehrengedächtnuß Der Weiland Erbahren/ Woi- vnd Viel-Ehren-tugendreichen Fr. Mariae Steudnerin/ geborner Prellerin Christian Adolph (GND, Wikipedia, ADB/NDB)
 Autor/in:
 Erscheinungsjahr: 1641
 Verlag/Drucker: Georgius Baumann
 Ort: Breslau
 Auflage: 1. Auflage
 Bibliothek: Universitätsbibliothek Breslau
 Signatur: Universitätsbibliothek Breslau, 4 F 1622/23 / 386427

INFORMATIONEN ZUM WERK

Publikationstyp: Monographie
 Verfügbarkeit: Text (TEI-XML-, HTML-, TCF-, E-Book-Fassung): Dieses Werk steht unter der „Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz“ (CC BY-SA). Weitere Informationen: Nutzungsbedingungen.
 Schriftart: Fraktur
 Genre: Gebrauchsliteratur :: Leichenpredigt
 im DTA seit: 2014-04-03 13:17:28
 Korpus: DTAE (DTA-Erweiterungen) – AEDit

GRUNDLAGE DIESES DIGITALISATS

Dieses Werk wurde im Rahmen des Moduls DTA-Erweiterungen (DTAE) digitalisiert. Weitere Informationen ...

INHALTSVERZEICHNIS

- o [Titelseite]
- o In Luctuosam, Mariæ Suæ Fidelissimæ, Absentiam,
- o [Abschnitt]
- o Abdanckung/
- o [Abschnitt]

Suche im Werk

Hilfe

Ansichten für dieses Werk

- Text-Bild-Ansicht
- alle Faksimiles
- DTAQ (Qualitätssicherung)

Download

XML (TEI P5) · HTML · Text
 TCF (text annotation layer)
 TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)
 XML (TEI P5 inkl. att.linguistic)

Metadaten

TEI-Header · CMDI · Dublin Core

Statistiken

Scans: 84
 Zeichen: ca. 111962
 Tokens: ca. 16225
 Oberflächentypes: ca. 5783

Wortwolken

- Lemmata
- Lemmata (nur Nomen)
- Types
- Types (nur Nomen)

Voyant Tools

- transliterierter Text
- normalisierter Text
- lemmatisierter Text

Language Resource Switchboard

- Resource/URL übermitteln

Obrázek č. 13 Detail titulu v Deutsches Textarchiv²⁴

Z pohledu hodnocení webové použitelnosti nepodává uživatelské prostředí dostatečnou informaci o tom, kde se v něm uživatel nachází, a především jak se na určitou stránku dostal. Vzhledem k tomu, že na obrazovku prohlížení textů je možné se dostat z úvodní obrazovky, vyhledávání nebo i detailu záznamu, může být uživatel zmaten tím, jak hluboko se v systému nachází. Jeho zmatení a kontrola nad aplikací může být ještě zhoršena tím, že uživatelské prostředí neobsahuje žádné tlačítko „zpět“ nebo „zrušit“. Vhodným řešením by mohlo být přidání tlačítka „zpět“, případně tlačítek odkazující uživatele zpět na vyhledávání, zpět na detail titulu a podobně. Dalším možným řešením je přidání tzv. drobečkové navigace²⁵.

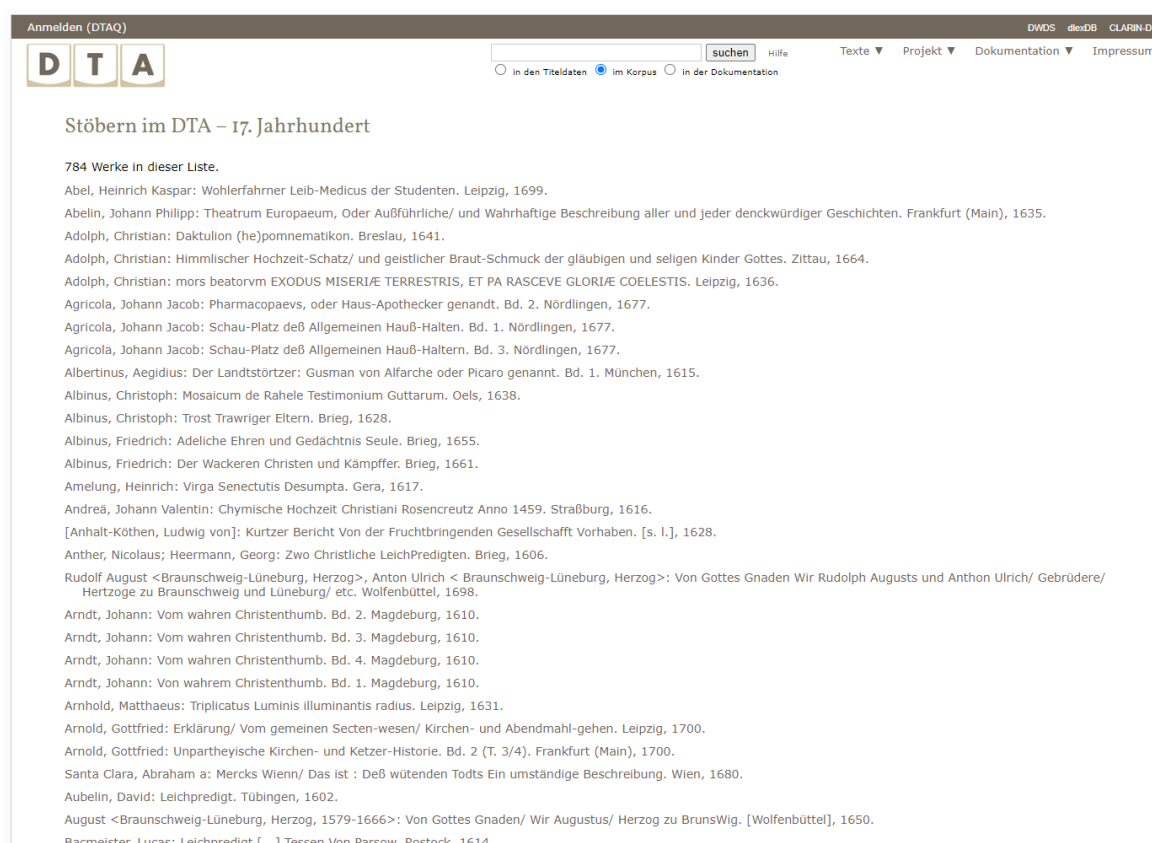
V rámci vyhledávání nejsou rozpoznávány zjevné překlepy, které by mohl systém automaticky opravovat nebo na ně upozorňovat. Samotné vyhledávání může být pro zkušeného uživatele velmi efektivní a flexibilní, neboť nabízí sofistikované vyhledávací funkce zahrnující

²⁴ Zdroj obrázku: Adolph, Christian: Daktulion (he)pomnematikon. Breslau, 1641. In: *Deutsches Textarchiv* [online digitální knihovna]. Berlin: Brandenburgische Akademie der Wissenschaften. [cit. 2021-05-05]. Dostupné z: <https://www.deutschestextarchiv.de/book/show/386427>

²⁵ Drobečková navigace, anglicky breadcrumb navigation, je forma kontextové navigace v podobě seznamu odkazů, které naznačují cestu, jež návštěvník prošel, než se na konkrétní stránku dostal, případně indikují umístění stránky v rámci struktury celého webu. [120]

práci s filtry, regulárními výrazy i možnostmi vyhledávání přímo v TEI dokumentu pomocí jazyka XPath. Všechny tyto funkce je ale možné používat jen pomocí kombinování speciálních znaků doprovázejících vyhledávané výrazy, které se zadávají do vyhledávacího pole (například výrazem `#has [author, / Arnim /]` je možné vyhledávat pouze tituly, jejichž autor má ve jméně slovo Arnim). Vzhledem k tomu, že tento způsob vyhledávání je použitelný pouze přes vyhledávací pole jsou méně zkušenému uživateli tyto funkce skryty a zneprístupněny. Řešením by mohlo být umožnit provádět alespoň některé skrze grafické rozhraní.

Jako problematický se může jevit i způsob pro procházení obsahu digitální knihovny, který umožňuje pouze zobrazit názvy titulů podle roku nebo typu dokumentu. Zobrazené tituly není možné nijak filtrovat nebo řadit, jedná se jen o dlouhý seznam odkazů, který působí nepřehledně (viz Obrázek č. 14). Prostředí pro procházení dokumentů by mělo být více strukturované a umožňovat třídění a filtrování zobrazených výsledků. Všechny nalezené problémy shrnuje tabulka č. 1.



Obrázek č. 14 Prostředí pro procházení dokumentů, Deutsches Textarchiv²⁶

²⁶ Zdroj obrázku: Stöbern im DTA – 17. Jahrhundert. In: *Deutsches Textarchiv* [online digitální knihovna]. Berlin: Brandenburgische Akademie der Wissenschaften. [cit. 2021-05-05]. Dostupné také z: <https://www.deustextarchiv.de/list/browse?century=17>

Nalezený problém	Odpovídající heuristika	Možné řešení
Chybějící informace o pozici uživatele v systému	Viditelnost stavu systému	Přidání drobečkové navigace
Chybějící tlačítko zpět	Uživatel má mít kontrolu a svobodu	Doplnit tlačítko zpět
Vyhledávání neopravuje překlepy.	Rozpoznávání a opravování chyb	Doplnění vyhledávání o návrhy na opravu možných překlepů
Pokročilé funkce vyhledávání jsou dostupné pouze přes vyhledávací řádek	Flexibilita a efektivita	Přidat možnost provést alespoň některé vyhledávací funkce pomocí grafického prostředí
Procházení obsahu má podobu dlouhého seznamu titulů	Předcházení chybám, Minimalistický design	Vytvořit více strukturované prostředí pro procházení dokumentů a možnosti jejich filtrování

Tabulka č. 1 Souhrn problémů s použitelností nalezených v Deutches Textarchiv

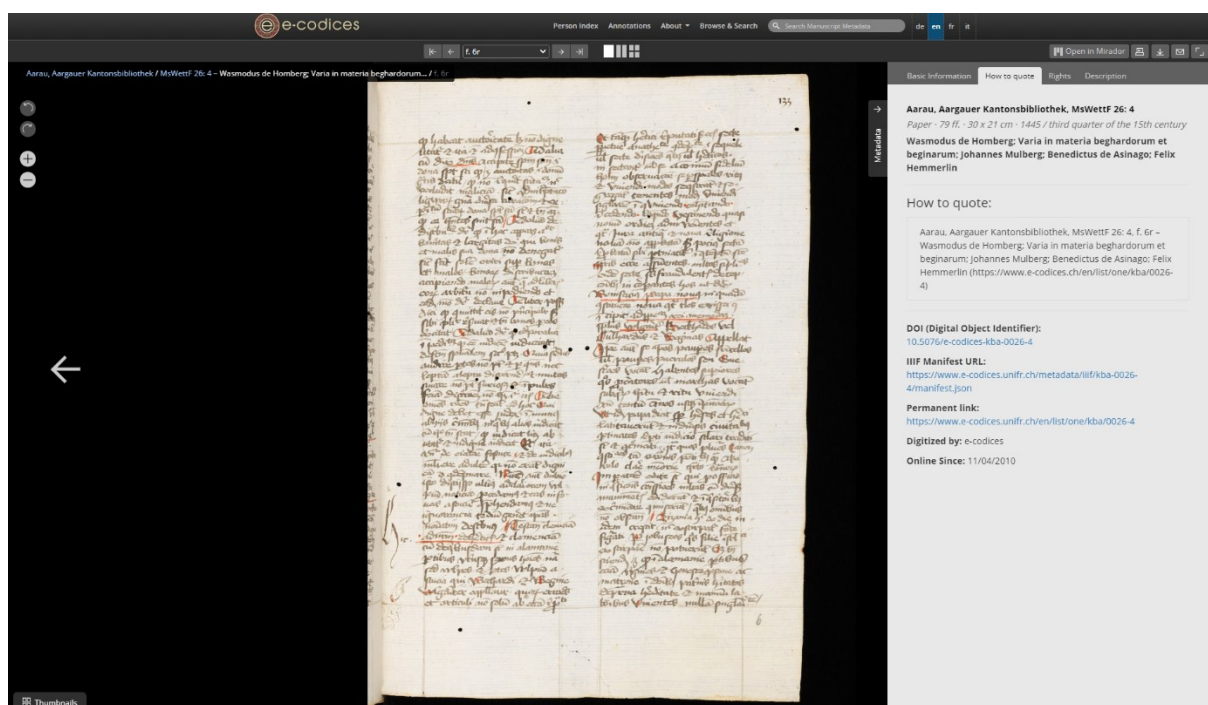
6.2 e-codices – Virtual Manuscript Library of Switzerland

e-codices je virtuální knihovna rukopisů Švýcarska, která funguje od roku 2008, nicméně staví na základech švýcarských digitalizačních projektů, které probíhaly již několik let před tím. V současnosti nabízí bezplatný online přístup k více než 2500 středověkých a raně novověkých rukopisů z veřejných, církevních i soukromých sbírek pocházejících z celého Švýcarska. Od roku 2013 jsou elektronické kodexy integrovány do národního programu švýcarských univerzit představujícího jednotnou digitální infrastrukturu nazvanou „Scientific information“. e-codices spolupracuje za účelem digitalizace a zveřejňování rukopisů se stovkami specialistů provádějících výzkum švýcarských rukopisů a provozuje dvě digitalizační centra. [101]

6.2.1 Prezentace textu

Dokumenty jsou uživatelům digitální knihovny prezentovány v podobě naskenovaných stránek. V prostředí pro prohlížení dokumentu je většina obrazovky věnována skenu stránky, který je možné otáčet, zvětšovat a zmenšovat. V pravém sloupci jsou dostupná metadata dokumentu – základní informace, jak dokument citovat, licence, pod kterou je dokument zpřístupněn, a podrobný popis dokumentu. Ten lze otevřít ve zvláštním okně pro větší přehlednost. Z prostředí je také možné daný sken uložit do počítače nebo otevřít v rámci webové aplikace Mirador. Ta představuje online pracovní prostředí, ve kterém je možné zkoumat obrázky z různých repozitářů. Fungování této aplikace je zajištěno díky využívání IIIF standardu.

Elektronické dokumenty dostupné v knihovně neobsahují textové přepisy. Jejich popisná metadata jsou ovšem kódována v TEI-P5 a tyto dokumenty jsou volně ke stažení. Text není kódován vůbec, tedy není splněna ani první úroveň zmíněná v kapitole 4.1. Obsahem TEI dokumentu je pouze element `<teiHeader>`, v němž jsou v rámci elementu `<fileDesc>` popsány informace o titulu (`<titleStmt>`), informace o edici (`<editionStmt>`), informace o publikování (`<publicationStmt>`) a především je velmi podrobně popsán zdrojový dokument (`<sourceDesc>`), pro jehož popis je využit modul *msdescription*. Ostatní moduly nelze přesně určit, protože schéma, na kterém je popis založen, není veřejně dostupné.



Obrázek č. 15 Prostředí pro prohlížení dokumentu, e-codices²⁷

6. 2. 2 Interoperabilita

Dokumenty obsažené v digitální knihovně jsou označeny perzistentním identifikátorem DOI, který je ještě doplněn permanentním URL odkazem založeným na syntaxi tohoto identifikátoru (například <https://www.e-codices.unifr.ch/en/list/one/bc/b-0026>). Metadata dokumentů jsou zapsána v TEI-P5 jak bylo popsáno výše. Jsou zobrazitelná v prostředí digitální knihovny nebo je možné je stáhnout jako dokument ve formátu XML.

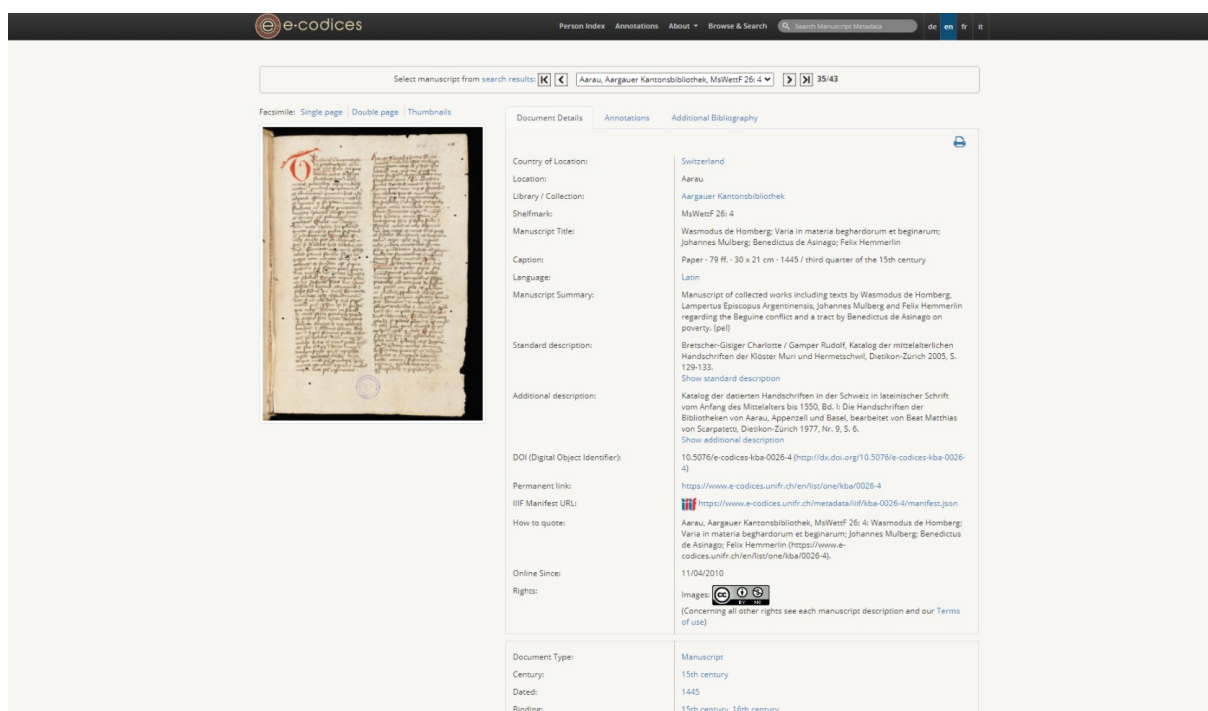
e-codices disponuje vedle manifestu IIIF také OAI-PMH rozhraním, skrz které je možné dokumenty, respektive metadata dokumentů, sklízet. Takto probíhá agregace záznamů do Europeany, ale například i do badatelského prostředí Manuscriptoria. [102, 103]

6. 2. 3 Webová použitelnost

Úvodní stránka digitální knihovny nabízí vedle vyhledávacího pole v záhlaví rovnou možnost procházet dostupné dokumenty podle knihoven a kolekcí, ze kterých pochází. Jednotlivé dokumenty jsou uživatelům prezentovány na souhrnné stránce obsahující přehled informací

²⁷ Zdroj obrázku: Aarau, Aargauer Kantonsbibliothek, MsWettF 26: 4. In: *e-codices – Virtual Manuscript Library of Switzerland* [online digitální knihovna]. [cit. 2021-05-05] Dostupné také z: <https://www.e-codices.unifr.ch/en/kba/0026-4/6r/0/>

o dokumentu (titul, jazyk, původ, popis, současné uložení a mnoho dalších, viz Obrázek č. 16), ze které se uživatel kliknutím na sken stránky rukopisu dostane do režimu prohlížení digitální faksimile popsaného výše. Odtud má možnost vrátit se na stránku s přehledem, na vyhledávání, nebo se může podívat na detail dokumentu, odkud je možné stáhnout TEI XML soubor.

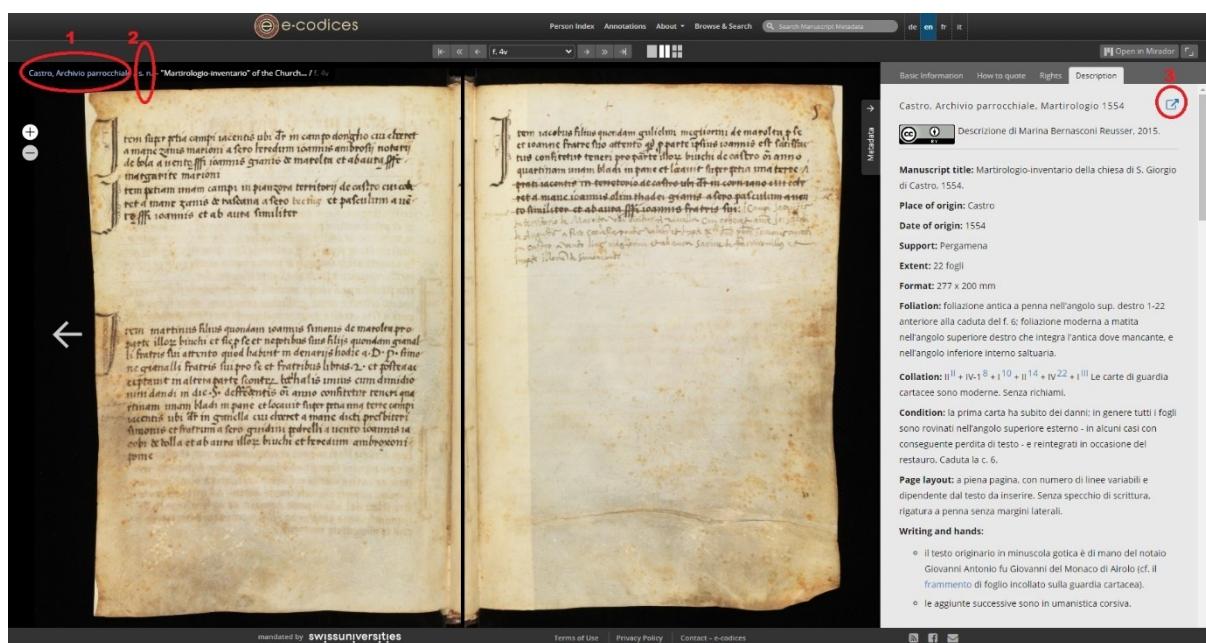


Obrázek č. 16 Souhrnná stránka titulu, e-codices²⁸

Právě obrazovka detailu dokumentu je z hlediska webové použitelnosti problematická, neboť není patrné, kde v hierarchické struktuře *úvodní obrazovka* → *vyhledávání* → *přehled dokumentu* → *digitální faksimile* stojí. Dostat se na stránku detailu dokumentu lze pouze z postranního sloupce při prohlížení skenů, a to kliknutím na ikonu „otevřít v novém okně“, z detailu se ovšem dá rovnou přejít na přehled dokumentu, odkud ale žádný zpětný odkaz nevede. Pro lepší navigaci by bylo vhodné umístit odkaz vedoucí na stránku detailu i na obrazovku přehledu dokumentu a záhlaví stránek doplnit drobečkovou navigací.

²⁸ Zdroj obrázku: Wasmodus de Homberg. In: *e-codices – Virtual Manuscript Library of Switzerland* [online digitální knihovna]. [cit. 2021-05-05]. Dostupné také z: <https://www.e-codices.unifr.ch/en/searchresult/list/one/kba/0026-4>

Nejasné jsou také ovládací prvky pro přecházení mezi jednotlivými obrazovkami. Z prostředí pro prohlížení skenů je možné dostat se, jak ukazuje Obrázek č. 17, zpět na vyhledávání (1), na přehled o dokumentu (2) a na detail dokumentu (3). Ani jedno z těchto tlačítek ale není jasně označeno a nezkušený uživatel může tápat v tom, jak nebo zda je použít. Jednoduchým řešením by bylo jejich přepracování tak, aby z jejich podoby byla jasná funkce. Pomocť uživateli by mohla také úprava vyhledávání tak, aby v případě, že na zadaný dotaz nejsou nalezeny žádné výsledky, poskytovala nápovědu, jak dále postupovat nebo navrhovala úpravy v dotazu. Zjištěné problémy shrnuje tabulka č. 2. Za zásadní nedostatek digitální knihovny v oblasti použitelnosti lze považovat úplnou absenci jakékoliv nápovědy a uživatelské dokumentace.



Obrázek č. 17 Prostedí pro prohlížení skenů, e-codices, upraveno²⁹

²⁹ Zdroj obrázku: Castro, Archivio parrocchiale, s. n. In: *e-codices – Virtual Manuscript Library of Switzerland* [online digitální knihovna]. [cit. 2021-05-05]. Dostupné také z: <https://www.e-codices.unifr.ch/en/doubleview/caap/Martirologio1/4v/>. Upraveno.

Nalezený problém	Odpovídající heuristika	Možné řešení
Chybějící informace o pozici uživatele v systému	Viditelnost stavu systému	Přidání drobečkové navigace, případně křížového odkazu na příslušná místa.
Nejasné ovládací prvky pro přecházení mezi obrazovkami	Uživatel má mít kontrolu a svobodu, Rozpoznávání před rozpomínáním	Přepracování tlačítek a doplnění jejich popisu
Pokud nejsou na vyhledávací dotaz nalezeny žádné výsledky, systém nenabízí řešení	Rozpoznávání a opravování chyb	Navrhovat uživateli jiné vyhledávací dotazy
Chybějící nápověda	Pomoc a dokumentace	Vytvoření uživatelské nápovědy a vhodné odkazování na ní

Tabulka č. 2 Souhrn problémů s použitelností nalezených v e-codices

6.3 Manuscriptorium

Manuscriptorium je digitální knihovna umožňující snadný přístup k soustředěným informacím o historických fondech – rukopisech, inkunábulích, raně novověkých tiscích a dalších. Uživatelům jsou nabízeny jak informace o fyzických dokumentech, tak přímo jejich digitální kopie. Jedním z cílů Manuscriptoria je zpřístupnění již existujícího digitálního obsahu pomocí jednotných nástrojů tak, aby byl co nejnáze dostupný. Knihovna proto agreguje dokumenty (nebo informace o nich) mnoha významných institucí ze zemí EU i mimo ně. Uživatelské prostředí Manuscriptoria je velmi sofistikované a obsahuje kromě možností vyhledávání a prohlížení dokumentů i vytváření vlastních uživatelských kolekcí a jejich sdílení [104]. V současnosti obsahuje tato digitální knihovna více než 370000 záznamů³⁰.

6.3.1 Prezentace dokumentů

Manuscriptorium funguje tak, že soustřeďuje popisná metadata o fyzických dokumentech ve formě tzv. evidenčních záznamů ve formátu XML a odkazuje uživatele na digitální dokumenty, které existují v Manuscriptoriu jako tzv. komplexní digitální dokumenty. Některé objekty dostupné v prostředí Manuscriptoria jsou tak jen záznamy o dokumentech, více než 150000 těchto záznamů je ale doplněno o digitální faksimile. Uživatelům jsou tak prezentovány skeny v kombinaci s popisnými metadaty. Manuscriptorium nicméně neobsahuje textové přepisy dokumentů. [105]

Záznamy obsažené v Manuscriptoriu jsou kódovány v TEI. TEI XML dokument je z prostředí digitální knihovny zobrazitelný. Dostupné je i schéma, na základě kterého jsou dokumenty kódovány. Jak bylo zmíněno výše, text samotný kódován není, TEI dokument se tak sestává pouze z hlavičky (`<teiHeader>`) a v případě, že je záznam doplněn i o faksimile, je zde tento fakt zmíněn. Manuscriptorium pracuje se schématem vyvinutým pro vlastní potřeby nazvané TEI P5 ENRICH schéma. Z názvu je patrné, že se jedná o schéma vycházející ze Zásad TEI-P5. Konkrétně je schéma pouze podmnožinou TEI-P5, tedy žádné moduly, elementy ani atributy nejsou do schématu dodány navíc, naopak, jedná se v podstatě o pouhá omezení.

³⁰ Údaj k 2.5. 2021 pocházející přímo z prostředí digitální knihovny

Schéma definuje tři rozdílné aspekty dokumentu:

- Metadata popisující dokument (<teiHeader>)
- Metadata popisující digitalizované obrazy originálního zdroje (<facsimile>)
- transkripce textu obsaženého v originálním zdroji (tento aspekt ale není povinný)

Metadata popisující dokument odpovídající záhlaví TEI obsahují:

- <fileDesc> obsahující bibliografický popis elektronického souboru
 - <titleStmt> s informacemi o názvu díla a osobách zodpovědných za jeho obsah
 - <publicationStmt> se souborem informací týkající se publikace nebo distribuce elektronického textu
 - <sourceDesc> s popisem pramene, ze kterého byl elektronický dokument vytvořen
- <revisionDesc> pro popis revizí

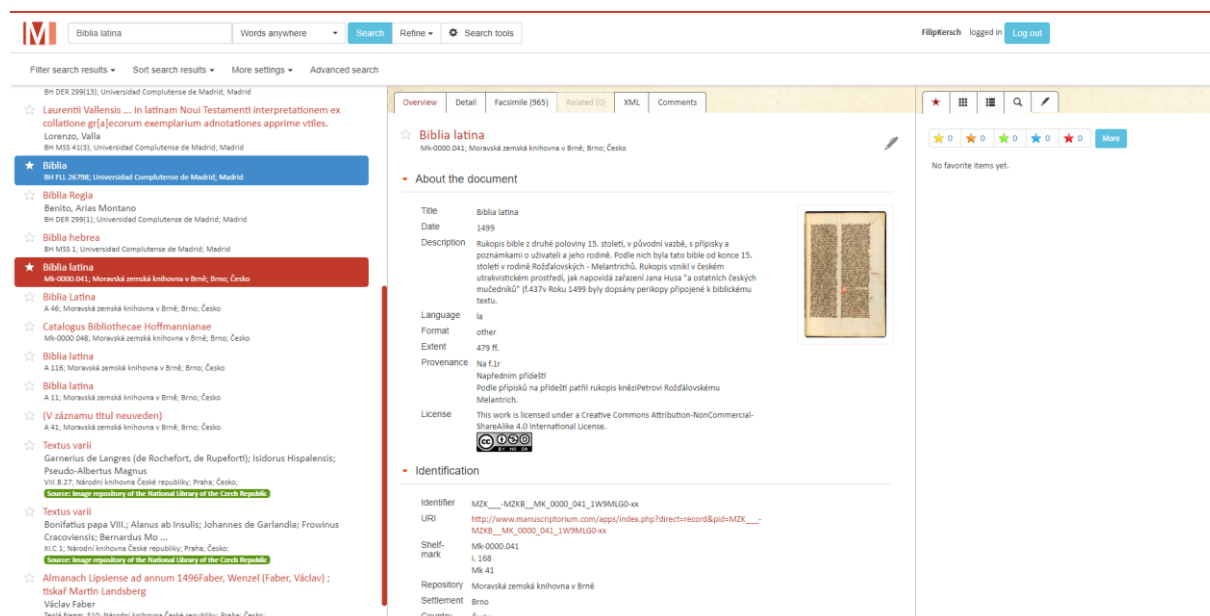
Schéma TEI-P5 ENRICH pracuje se čtyřmi základními moduly TEI: *header*, *core*, *tei* a *textStructure* a dále je zapojeno 5 specializovaných modulů: *msdescription*, *linking*, *namesdates*, *figures* a *transr*. Přibližně 60 elementů těchto modulů bylo ovšem pro potřeby schématu smazáno a některé atributy v TEI definované jako nepovinné, jsou v TEI P5 ENRICH povinné. Zároveň byly omezeny hodnoty, kterých mohou tyto atributy nabývat. Jedná se například o elementy <msDesc>, kde je nyní nutné uvádět atribut @xml:id a @xml:lang, u elementu <anem> je povinný atribut @type a smí nabývat pouze hodnot "person", "place", "org", "unknown". Změn v attributech je ale výrazně více. [106]

6.3.2 Interoperabilita

Dokumenty v Manuscriptoriu jsou označeny pomocí jednoznačného URI identifikátoru, který se jeví jako perzistentní, jedná se ovšem o interní identifikátor této digitální knihovny, a ne o jeden z nejběžněji používaných identifikátorů zmíněných v kapitole 5.2.2. Tento identifikátor má podobu například NKCR__-NKCR__22_B_000022_2CP7FGF-cs a není součástí TEI XML dokumentu.

Metadata dokumentů je možné zobrazit v prostředí digitální knihovny (Obrázek č. 18). Na kartě „souhrn“ jsou dostupné informace o dokumentu, jeho identifikaci a obsahu (rozsahu). Na kartě „detail“ jsou potom navíc informace o uložení, fyzickém popisu, historii (provenienci a vzniku). Tato metadata se nedají jednoduše stáhnout v žádném výměnném formátu (například Dublin Core). Jediný dostupný formát představuje TEI XML dokument.

Manuscriptorium je agregátorem pro evropskou digitální knihovnu Europeana. Skrze protokol OAI-PMH je možné do něj dokumenty přidávat z jiných digitálních knihoven a zároveň jsou skrze tento protokol data sdílena dál. V současnosti existují 3 profily pro sklizení Manuscriptoria: DC, QQualified DC a ESE (Europeana Semantic Elements v 3.2). [107]

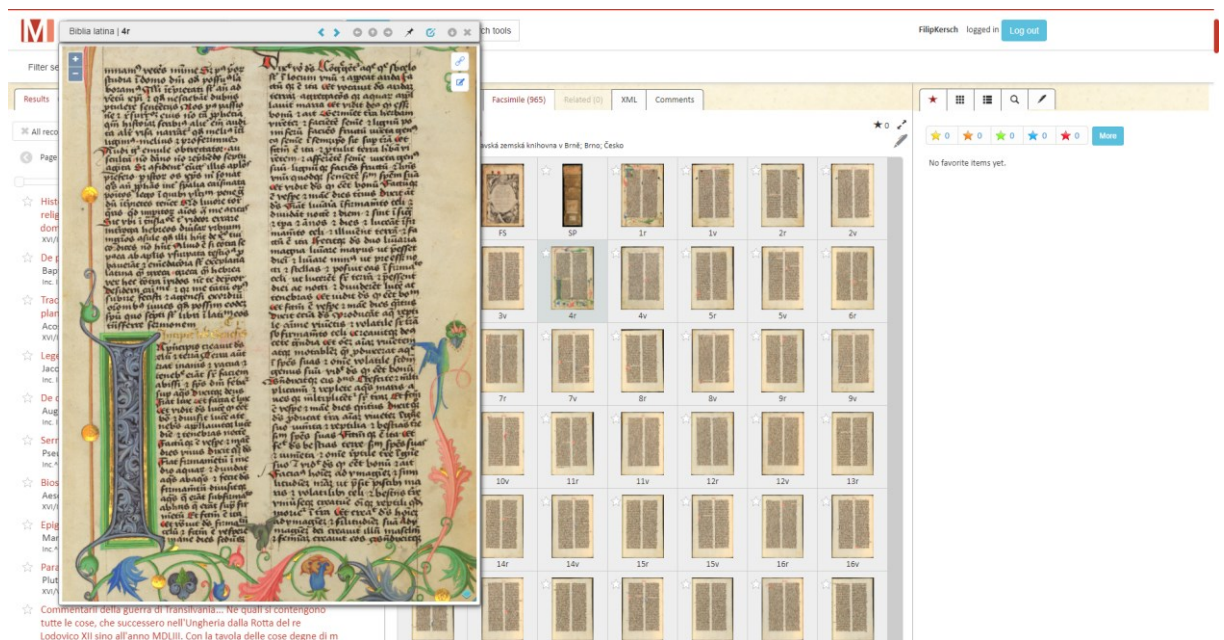


Obrázek č. 18 Pohled na uživatelské prostředí digitální knihovny Manuscriptorium³¹

³¹ Zdroj obrázku: *Manuscriptorium: Digital Library of Written Cultural Heritage* [online digitální knihovna]. Národní knihovna České republiky [cit. 2021-5-2]. Dostupné také z: <http://www.manuscriptorium.com/apps/index.php?envLang=en#search>

6. 3. 3 Webová použitelnost

Prohlížení dokumentů v digitální knihovně funguje v rámci systému Guliver. V horní části obrazovky se nachází vyhledávací pole a další nástroje vyhledávání, levé menu zobrazuje výsledky vyhledávání, které je možné řadit a filtrovat. V prostředním okně se nachází záznam dokumentu, přičemž kartami v záhlaví této části je možné přepínat mezi pohledy na souhrn záznamu, jeho detail, faksimile (pokud je dostupná), související dokumenty (pokud jsou dostupné), TEI XML dokument a poznámky. Po kliknutí na kartu „faksimile“ se zobrazí náhledy jednotlivých stránek, přičemž po kliknutí na vybranou stránku se sken otevře ve svém vlastním okně (v rámci aplikace, ne prohlížeče – viz Obrázek č. 19). Pravý sloupec potom poskytuje prvky personalizace – oblíbené položky, vlastní kolekce, uložené dotazy, poznámky a podobně.



Obrázek č. 19 Prohlížení skenu v digitální knihovně Manuscriptorium³²

Prostředí digitální knihovny je velmi komplexní a třísloupcová organizace informací může zejména méně zkušené uživatele informacemi přímo přehltit. Pro lepší orientaci a možnost soustředit se by mohlo být vhodné zvážit implementaci možnosti skrytí obou postranních sloupců. Orientaci uživatele v prostředí a funkcích Manuscriptoria nepomáhá ani fakt, že

³² Zdroj obrázku: Biblia latina: Mk-0000.041 Moravská zemská knihovna v Brně; Brno Česko. In: *Manuscriptorium: Digital Library of Written Cultural Heritage* [online digitální knihovna]. Národní knihovna České republiky [cit. 2021-5-2]. Dostupné také z: <http://www.manuscriptorium.com/apps/index.php#search>

nápověda k ovládnání je dostupná pouze skrze tři krátká videa, na která navíc není odkazováno z digitální knihovny samotné, ale pouze z úvodní stany. Drobností, která ovšem může pomoci předcházet chybám, je pak přidat omezení do vyhledávání v rámci některých polí – například při vyhledávání v poli „datum“ je možné zadat dlouhý textový řetězec, který téměř jistě nepovede k nalezení shody. Stejně tak by systém mohl uživateli napovídat, jak řešit situaci, kdy není nalezen žádný výsledek jeho vyhledávání a opravovat zjevné překlepy. Všechny nalezené problémy s použitelností ukazuje tabulka č. 3.

Nalezený problém	Odpovídající heuristika	Možné řešení
Vyhledávání v některých polích (například datum) neomezuje zadávané výrazy	Předcházení chybám	Přidání omezených možností pro vyplňování některých polí
Prostředí zobrazuje mnoho informací najednou	Minimalistický design	Umožnit skrytí nepoužívaných sloupců
Vyhledávání neopravuje překlepy.	Rozpoznávání a opravování chyb	Doplnění vyhledávání o návrhy na opravu možných překlepů
Pokud nejsou na vyhledávací dotaz nalezeny žádné výsledky, systém nenabízí řešení	Rozpoznávání a opravování chyb	Navrhovat uživateli jiné vyhledávací dotazy
Špatně dostupná nápověda	Pomoc a dokumentace	Vytvoření uživatelské dokumentace a odpovídající odkazování na ní.

Tabulka č. 3 Souhrn problémů s použitelností nalezených v Manuscriptoriu

6.4 Medieval Nordic Text Archive

Medieval Nordic Text Archive (Archiv středověkých severských textů) je projektem Sdružení Menota – síť předních severských archivů, knihoven a výzkumných pracovišť pracujícími se středověkými texty. Cílem Menoty je uchovat a publikovat texty v digitální podobě a přizpůsobit a vyvinout standardy kódování nezbytné pro tuto práci. Obsahem archivu založeného v roce 2001 jsou především texty v severských jazycích, ale také latině. V současnosti obsahuje archiv 50 středověkých severských textů, z nichž některé jsou plně lemmatizované – v archivu je tedy jinými slovy 1,8 milionů slov. [108]

6.4.1 Presentace textu

Textový archiv používá pro publikaci systém pro správu a analýzu korpusů Corpuscle vyvíjený v rámci norské infrastruktury pro jazykové zdroje a technologie – CLARINO. Systém umožňuje současné prohlížení textů kódovaných v různých podobách (viz dále) i skenu stránky zároveň. Různé pohledy na text jsou generovány XSL transformacemi z TEI dokumentů. Nicméně ne všechny texty jsou doprovázeny skeny, a ne všechny skeny jsou kódovány ve více podobách. Podle dostupných informací probíhá přepisování textů ručně, technologie OCR je sice v rámci projektu testována, ale není využívána pro produkci textů. [109]

Texty v archivu jsou kódovány pomocí schématu založeného na Zásadách TEI. TEI XML dokumenty všech publikovaných textů jsou volně ke stažení, stejně jako velmi podrobná příručka poskytující komplexní návod pro kodéry a všechna potřebná schémata včetně ODD dokumentů. V rámci textového archivu se pracuje se třemi verzemi kódování, a to s verzí „faksimile“, kdy je text přepsán písmeno po písmeni s výběrem paleografických charakteristik a zachováním zkratk jako v rukopisu (jedná se tedy v zásadě o transliteraci), úrovní „diplomatickou“, kdy je přepis realizován po písmenech, ale jen s malým výběrem paleografických rysů a identifikací zkratk, a úrovní „normalizovanou“, kdy je standardizován pravopis. Podle pravidel archivu není nutné pro zveřejnění kódovat text ve více než jedné verzi, ostatně relativně mnoho textů obsažených v knihovně je zakódováno jen na úrovni diplomatické [109]. Stejně tak hloubka a detailnost kódování se mezi texty výrazně liší. Ve své minimální podobě obsahují texty v archivu pouze řetězec slov, kde je každé slovo označeno tagem `<w>`, a informace o rozložení textu – tedy konce stránek (`<pb>`), konce řádků (`<lb>`) a elementy `<div>` pro rozlišení kapitol. Některé texty ovšem obsahují výrazně větší množství informací a kódují i v textu se vyskytující jména, metriky, ornamenty, informace o gramatice a podobně.

Ze zásad TEI jsou využity následující moduly:

- *analysis*
- *core*
- *corpus*
- *drama*
- *figures*
- *header*
- *linking*
- *msdescription*
- *namesdates*
- *tagdocs*
- *tei*
- *textcrit*
- *textstructure*
- *transcr*
- *verse*

Použité kódovací schéma obsahuje oproti původním Zásadám TEI navíc rozšíření, která podle sdružení Menota vychází ze dvou hlavních rozdílů mezi středověkými severskými texty a většinou ostatních kompatibilních korpusů: Prvním je velká míra ortografických variací, kvůli kterým je lingvistická analýza velmi obtížná, neboť je velmi složité hledat slova na základně lemmat. Druhou rozdílností je velké množství zkratk přítomných v textu. Tyto problémy jsou řešeny jednak oním výše zmíněným rozdělení do třech verzí a také přidáním několika elementů a atributů k původnímu TEI P5 schématu [109]. Dříve projekt pracoval a upravoval starší varianty Zásad TEI, nicméně zde bude popsána pouze současná verze schématu Menota 3.0 vycházející z TEI P5.

Veškeré přidané elementy náleží jmennému prostoru

`xmlns:me="http://www.menota.org/ns/1.0"` a jedná se o:

- `<me:facsimile>` pro kódování dokumentu ve verzi faksimile
- `<me:dipl>` pro kódování dokumentu v diplomatické verzi
- `<me:normalized>` pro kódování dokumentu v normalizované verzi
- `<me:paleo>` pro kódování dokumentu v paleografické verzi (takovou verzi ale žádný z textů v archivu podle dostupných informací neobsahuje)
- `<me:suppressed>` pro kódování částí textu, které jsou odstraněny editorem
- `<me:textSpan/>` pro kódování libovolných nespojitých struktur
- `<me:alliterati>` pro aliteraci při kódování veršů
- `<me:assonance>` pro vnitřní rým při kódování veršů

Přidané atributy jsou:

- `@me:msa` pro morfosyntaktickou analýzu
- `@me:type` pro účely klasifikace ve spojení s elementy `<ex>` (pro rozepsání zkratky) a `<am>` (posloupnost písmen nebo znaků ve zkratce)
- `@me:key` pro propojení s jiným zdrojem ve formě databázového klíče
- `@me:ref` pro propojení s jiným zdrojem pomocí identifikátoru typu URI
- `@me:level` pro identifikaci úrovně, na které byl dokument přepsán, jedná se o atribut elementu `<normalization>` v záhlaví dokumentu
- `@me:lemmatized` pro identifikaci textů, které byly lemmatizovány, jedná se o atribut elementu `<interpretation>` v záhlaví dokumentu
- `@me:morphAnalyzed` pro identifikaci textů, které byly morfologicky analyzovány, jedná se o atribut elementu `<interpretation>` v záhlaví dokumentu
- `@me:category` pro identifikaci rozpětí textu, jedná se o atribut elementu `<me:textSpan/>` sloužícího k práci s hierarchicky se překrývajícími strukturami.

6. 4. 2 Interoperabilita

Dokumenty nedisponují žádným perzistentním univerzálním identifikátorem. Identifikaci v rámci databáze zaručuje identifikátor nazvaný „Menota“, který je v podstatě přírůstkovým číslem. V TEI XML dokumentu je zapsán například takto: `<idno type="Menota">Ms. 10</idno>`.

Jediná metadata dostupná k textům jsou ta uvedená v záhlaví TEI dokumentu. Zahrnují všechny povinné elementy: `<fileDesc>` pro popis elektronického souboru, prohlášení o názvu (`<titleStm>`), prohlášení o publikaci (`<publicationStm>`) a popis zdrojového dokumentu (`<sourceDesc>`). Vedle toho jsou ale vždy v záhlaví uvedeny i další, pro minimální strukturu TEI nepovinné, údaje. Konkrétně se jedná o popis kódování (`<encodingDesc>`), kde je popsáno mimo jiné, jaké verze kódování byly vytvořeny, popis klasifikačních údajů v elementu (`<profileDesc>`) a často je také přítomen popis provedených revizí elektronického textu (`<revisionDesc>`).

Informace o tom, zda Medieval Nordic Text Archive agreguje dokumenty jiných digitálních knihoven nebo sám dokumenty vystavuje a zpřístupňuje k agregaci nejsou dostupné, pravděpodobně se tak neděje.

6. 4. 3 Webová použitelnost

Úvodní obrazovka po vstupu do textového archivu nabízí přehled všech dostupných textů v podobě tabulky, která shrnuje základní charakteristiky textu, dostupné verze a další údaje (Obrázek č. 20). Kliknutím na tlačítko „Download“ v posledním sloupci tabulky se lze rovnou dostat k TEI XML dokumentu, klikem na název dokumentu se uživatel dostane do prostředí pro prohlížení textu. Levé postranní menu nabízí alternativní způsoby přístupu k obsahu archivu – obecné vyhledávání, vyhledávání podle pojmů (konkordance) a vyhledávání v kolekcích. Prohlížení textů potom disponuje jednoduchým přepínáním stránek daného dokumentu a zaškrtačacími políčky, kterými si uživatel volí, jaké verze přepisu chce mít zobrazené (Obrázek č. 21)

Public catalogue

This catalogue lists all publicly available Menota texts. Texts are displayed on one or more levels: facsimile (very close transcription), diplomatic (less close and with some interpretation), normalised (regularised orthography). Some texts are morphologically annotated and thus allow for more detailed linguistic searches. A search can be performed on a single text or on any combination of texts.

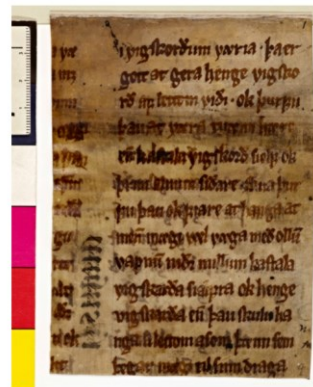
Signature	MS or Work Title	Main Editor(s)	Annotation	H	Version	Date	Lang.	Words	Lemma	Form	Facs	Dipl	Norm	Photo	Status	License	Download
AM 35 fol	Heimskringla 1	Heimskringlafróskjetat ved Universitetet i Oslo, Jon Gunnar Jørgensen		H	1.0.5	c. 1700	isl	68573	none	none	yes				high	CC-BY-SA 4.0	Download
AM 36 fol	Heimskringla 2	Heimskringlafróskjetat ved Universitetet i Oslo, Kjartan Ottosson		H	0.9	c. 1700	isl	105532	none	none	yes				high	CC-BY-SA 4.0	Download
AM 56 4to	Landslög Magnúss Hákonarsonar I Magnús Lagabotes landslöv	Anna C. Horn		H	0.7	c. 1300	nor	49256	none	none	yes	yes			high	CC-BY-SA 4.0	Download
AM 60 4to	Kristinnættis, Bjarflog Magnúss Hákonarsonar ok rettarbotir Kristinnættis, Magnús Lagabotes landslöv, Magnús Lagabotes bylov og rettarbotir	Anna C. Horn		H	0.9	c. 1320	nor	87387	none	none	yes				high	CC-BY-SA 4.0	Download
AM 63 fol	Heimskringla 3	Heimskringlafróskjetat ved Universitetet i Oslo, Jon Gunnar Jørgensen		H	1.0.5	c. 1700	isl	89772	none	none	yes				high	CC-BY-SA 4.0	Download
AM 78 4to	Kristinnættis, Bjarflog Magnúss Hákonarsonar ok rettarbotir Kristinnættis, Magnús Lagabotes bylov og rettarbotir	Anna C. Horn		H	0.8	c. 1320	nor	45516	none	none	yes	yes			high	CC-BY-SA 4.0	Download
AM 113 b fol	Íslendingabók	Matteo Tarsi		H	1.0.4	c. 1700	isl	4023	completely	completely	yes	yes	yes		high	CC-BY-SA 4.0	Download
AM 162 B o fol	A fragment of Njáls saga	Katarzyna Anna Kapitán		H	1.0.2	c. 1390-1440	isl	942	none	none	yes	yes	yes		high	CC-BY-SA 4.0	Download
AM 162 B o fol	Two fragments of Njáls saga	Beeke Stegmann		H	1.0.3	c. 1300-1350	isl	2279	none	none	yes	yes	yes		high	CC-BY-SA 4.0	Download
AM 162 B o fol	A fragment of Njáls saga	Beeke Stegmann		H	1.0.5	c. 1325-1375	isl	2183	none	none	yes	yes	yes		high	CC-BY-SA 4.0	Download
AM 178 fol (1r1-17r17)	Hátríks saga af Bern (part 1 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	9216	partly	partly	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (47r3-49r9)	Hátríks saga af Bern (part 10 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	1375	completely	completely	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (58r19-58v6)	Hátríks saga af Bern (part 3 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	176	completely	completely	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (60r16-61r15)	Hátríks saga af Bern (part 5 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	672	completely	completely	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (108r3-110r9)	Hátríks saga af Bern (part 12 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	1259	completely	completely	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (123r3-124r22)	Hátríks saga af Bern (part 14 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	825	completely	completely	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (129v8-131r8)	Hátríks saga af Bern (part 16 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	764	partly	partly	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (165v8-166v4)	Hátríks saga af Bern (part 18 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	727	completely	completely	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (174v19-185v18)	Hátríks saga af Bern (part 20 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	6147	partly	partly	yes				high	CC-BY-SA 4.0	Download
AM 178 fol (186v9-194r24)	Hátríks saga af Bern (part 22 of 22)	Spráksamlingane		H	0.9.1	c. 1600-1700	isl	4513	completely	completely	yes				high	CC-BY-SA 4.0	Download
AM 242 fol	Codex Wormianus	Karl Gunnar Johansson		H	0.9	c. 1350	isl	62586	none	none	yes	yes			high	CC-BY-SA 4.0	Download

Obrázek č. 20 Úvodní obrazovka textového archivu³³

AM 1056 IX 4to – A fragment of Konungs skuggsjá v. 1.1

Previous page | Next page | Go to: [1] | Levels: [Facs] [Dipl] [Norm] | Translation: [Clean text] [Display photographic facsimiles] | Size: small - medium - large

[Facs]	[Dipl]	[Norm]
[1v r1]	[1v r1]	[1v r1]
1 i vigsokodum varja, þa er	1 i vigsokodum varja, þa er	1 i vigsokodum varja, þa er
2 gort ar gera henge vigsko	2 gort ar gera henge vigsko	2 gort ar gera henge vigsko
3 rd af lettom vidi, ok þurfu	3 rd af lettom vidi, ok þurfu	3 rd af lettom vidi, ok þurfu
4 þau at vera tveim hære	4 þau at vera tveim hære	4 þau at vera tveim hære
5 efi kaatala vigsokod flof ok	5 enr kaatala vigsokod flof ok	5 enr kaatala vigsokod flof ok
6 þrim alnum idare, Sua þur	6 þrim alnum idare, Sua þur	6 þrim alnum idare, Sua þur
7 þu þau ok þare at hanga ar	7 fu þau ok þare at hanga ar	7 fu þau ok þare at hanga ar
8 mefi magi val vega með ollum	8 menr magi val vega með ollum	8 menr magi val vega með ollum
9 vagnum niðr millum kaatala	9 vagnum niðr millum kaatala	9 vagnum niðr millum kaatala
10 vigskaða siftra ok henge	10 vigskaða siftra ok henge	10 vigskaða siftra ok henge
11 vigskaða efi þau skulu ha	11 vigskaða enr þau skulu ha	11 vigskaða. En þau skulu ha
12 nga alrecom alom þeim sem	12 nga alrecom alom þeim sem	12 ga alrecom alom þeim sem
13 þegar magi til finn draga	13 þegar magi til finn draga	13 þegar magi til finn draga...



Previous page | Next page

Obrázek č. 21 Prostředí pro prohlížení textů v Medieval Nordic Text Archive³⁴

³³ Zdroj obrázku: Public catalogue. In: *Medieval Nordic Text Archive: Menota* [online digitální knihovna]. 2007- [cit. 2021-5-2]. Dostupné z: <https://clarino.uib.no/menota/catalogue>

³⁴ Zdroj obrázku: AM 113 b fol. – Íslendingabók v. 1.0.4. In: *Medieval Nordic Text Archive: Menota* [online digitální knihovna]. 2007- [cit. 2021-5-2]. Dostupné z: <https://clarino.uib.no/menota/catalogue>

Medieval Nordic Text Archive disponuje jen jednoduchým rozhraním, ve kterém není složité se vyznat. V kontrastu s ním je velmi sofistikované vyhledávání, které pracuje s dotazovacím jazykem vytvořeným pro systém Corpuscle. Vyhledávání je zároveň poměrně flexibilní, umožňuje klást složité dotazy pomocí vyhledávacího řádku, ale i pomocí menu, ve kterých může uživatel dotaz sestavit, pokud si není jistý jeho syntaxí. Tato menu ovšem představují problém z hlediska použitelnosti, neboť pracují s velkým množstvím zkratk, jež nejsou nijak ustálené a uživatel, který není zvyklý se systémem pracovat, tak nebude schopný jakýkoliv složitější dotaz sestavit (viz Obrázek č. 22). Zkratky a funkcionality vyhledávání jsou sice popsány v dokumentaci k systému, ta je ovšem velmi rozsáhlá a nepříliš logicky dostupná částečně přímo z rozhraní digitální knihovny a částečně jako PDF dokument ke stažení. Na konkrétní sekce (například tu o vyhledávání) není z relevantních částí rozhraní odkazováno. Možným řešením by mohlo být přepracování a lepší strukturování dokumentace, odkazování na její konkrétní části z těch míst knihovny, kde může být potřeba, a doplnění zkratk ve vyhledávacích formulářích o popis, který by se zobrazil například po najetí kurzorem na konkrétní zkratku.

Query

Advanced search | [switch to Basic search](#)

[]

Run Query | **Reset** | **Build graphical query** | Show query expression | Saved queries ...

Choose **positional constraints**: | Ignore structural positions

target
 norm =
 msa: [select]
 repetit

+

Additional
 dipl
 facs
 lemma
 pos
 feature
 langua
 referer
 struct

Choose a **sub**
 (Entir
 Addition
 document

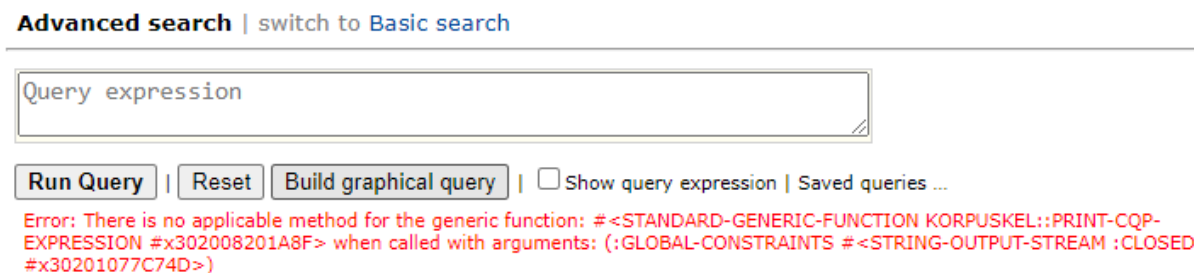
Select values for 'msa'. Only combinations that do occur in the corpus can be selected.

Done | **Add as alternative** | **Delete this alternative** | **Reset** | Page 1 of 2 | **Previous** | **Next**

<input type="checkbox"/> 00000	<input type="checkbox"/> cU	<input type="checkbox"/> gMF	<input type="checkbox"/> mSU	<input type="checkbox"/> rC	<input type="checkbox"/> vR
<input type="checkbox"/> cA	<input type="checkbox"/> cV	<input type="checkbox"/> gMN	<input type="checkbox"/> mSUIIM	<input type="checkbox"/> rP	<input type="checkbox"/> xAJ
<input type="checkbox"/> cAD	<input type="checkbox"/> eN	<input type="checkbox"/> gN	<input type="checkbox"/> mU	<input type="checkbox"/> rS	<input type="checkbox"/> xAP
<input type="checkbox"/> cAG	<input type="checkbox"/> eP	<input type="checkbox"/> gU	<input type="checkbox"/> nD	<input type="checkbox"/> rU	<input type="checkbox"/> xAQ
<input type="checkbox"/> cAN	<input type="checkbox"/> fF	<input type="checkbox"/> iPP	<input type="checkbox"/> nP	<input type="checkbox"/> sD	<input type="checkbox"/> xAT
<input type="checkbox"/> cB	<input type="checkbox"/> fI	<input type="checkbox"/> iRD	<input type="checkbox"/> nS	<input type="checkbox"/> sI	<input type="checkbox"/> xAV
<input type="checkbox"/> cD	<input type="checkbox"/> fP	<input type="checkbox"/> iST	<input type="checkbox"/> nSP	<input type="checkbox"/> sU	<input type="checkbox"/> xCC
<input type="checkbox"/> cF	<input type="checkbox"/> fU	<input type="checkbox"/> iU	<input type="checkbox"/> nU	<input type="checkbox"/> tPS	<input type="checkbox"/> xCS
<input type="checkbox"/> cG	<input type="checkbox"/> gF	<input type="checkbox"/> iWK	<input type="checkbox"/> p1	<input type="checkbox"/> tPT	<input type="checkbox"/> xCU
<input type="checkbox"/> cGD	<input type="checkbox"/> gFN	<input type="checkbox"/> mIN	<input type="checkbox"/> p2	<input type="checkbox"/> tU	<input type="checkbox"/> xDD
<input type="checkbox"/> cN	<input type="checkbox"/> gG	<input type="checkbox"/> mINSU	<input type="checkbox"/> p3	<input type="checkbox"/> vA	<input type="checkbox"/> xDP
<input type="checkbox"/> cO	<input type="checkbox"/> gM	<input type="checkbox"/> mIP	<input type="checkbox"/> pU	<input type="checkbox"/> vP	<input type="checkbox"/> xDQ

Obrázek č. 22 Vyhledávací formulář v Medieval Nordic Text Archive

Vyhledávání také nenabízí našeptavač a nesnaží se uživateli navrhnout alternativní postupy v případě, že na jím zadaný dotaz nejsou nalezeny žádné výsledky. Mnohem problematičtější je ale situace, kdy je do vyhledavače zadán špatně sestavený dotaz. V takovém případě systém zobrazí chybovou zprávu, ze které není pochopitelné, v čem přesně nastala chyba a uživateli není nijak napovídáno, jak má chybu odstranit (Obrázek č. 23).



Obrázek č. 23 Chybová zpráva ve vyhledávání v Medieval Nordic Text Archive

Vyhledávání trpí také problémem s rychlostí. Pokud je zadán moc obecný dotaz, prohledání databáze pravděpodobně trvá poměrně dlouho a po tuto dobu není uživatel nijak informován, zda byla jeho akce zaregistrována, nebo ne. Uživatel vidí buď nezměněnou stránku, ze které dotaz odeslal, nebo stránku čistě bílou. Řešením tohoto problému je implementovat pro tyto případy nějakou formu zpětné vazby, že systém dotaz registruje a pracuje na něm.

Nalezený problém	Odpovídající heuristika	Možné řešení
Při dlouhotrvajícím vyhledávání nedává systém zpětnou vazbu	Viditelnost stavu systému	Implementace zprávy pro uživatele, že systém akci registroval a pracuje na ní
Vyhledávání nemá našeptavač	Předcházení chybám	Přidání našeptavače
Používání zkratk ve vyhledávání	Rozpoznávání před rozpomínáním	Doplnění zkratk o popis
Chybové zprávy ve vyhledávání nejsou konstruktivní	Rozpoznávání a opravování chyb	Změnit znění chybových zpráv
Pokud nejsou na vyhledávací dotaz nalezeny žádné výsledky, systém nenabízí řešení	Rozpoznávání a opravování chyb	Navrhovat uživateli jiné vyhledávací dotazy
Nápověda je nedostatečně strukturovaná a není na ní náležitě odkazováno	Pomoc a dokumentace	Přepracování nápovědy a náležité odkazování na ní

Tabulka č. 4 Souhrn problémů s použitelností nalezených v Medieval Nordic Text Archive

6.5 Oxford Text Archive

Oxford Text Archive (OTA) je digitální repozitář, jehož cílem je shromažďovat, katalogizovat, uchovávat a především distribuovat literární a jazykové sady dat. OTA byl založen na University of Oxford v roce 1976 a i v současnosti je s univerzitou stále spojen – patří do správy Bodleian Libraries, sdružení 28 oxfordských knihoven. Jedná se jeden z nejstarších nepřetržitě provozovaných archivů digitálních zdrojů pro literární a jazykový výzkum na světě. V současnosti obsahuje téměř 35000 textů ve více než 25 různých jazycích, z toho více než 33000 pochází z období před rokem 1800. OTA je také jedním z klíčových center Evropské výzkumné infrastruktury CLARIN (Common Language Resources and Technology Infrastructure). [110, 111]

6.5.1 Presentace dokumentů

Jak předznamenává už název, Oxford Text Archive obsahuje pouze přepisy textů primárních zdrojů. Knihovna je založena na open-source technologii DSpace, která funguje pouze jako repozitář dokumentů, uživatelské rozhraní tedy nedisponuje čtečkou nebo obdobným prostředím pro prohlížení textů. Na webové stránce věnované konkrétnímu titulu je vždy vedle metadat pouze výčet souborů ke stažení (Obrázek č. 26). Vždy je dostupný TEI XML dokument a HTML dokument, velmi často je dostupný také soubor s prostým textem a některé tituly jsou doplněny o formáty ePub a mobi (proprietární formát pro čtečky Kindle), případně další.

Texty dostupné v OTA pochází z různých digitalizačních projektů nebo přímo od uživatelů, kteří mohou své texty skrze rozhraní archivu sami nabídnout. Přijímány jsou nicméně pouze kvalitní texty splňující určené podmínky. Jednou z nich je poskytnutí TEI XML souboru, který odpovídá zásadám TEI a projde validačním procesem [110]. Vzhledem k různému původu textů je možné, že způsoby jejich vzniku pokrývají celou škálu od ručního přepisu až po automatické OCR nepodléhající korekci. Velmi různorodá je také úroveň kódování jednotlivých dokumentů. V některých případech odpovídá 3. nebo i 4. úrovni kódování, tak jak jsou popsány v kapitole 4.1 – v textu jsou tedy označeny odstavce, obrázky, poznámky, použití různých typografických prvků, ale někdy i smysl a funkce textových a strukturálních prvků. Na druhou stranu některé texty disponují pouze záhlavím TEI a jejich přepis je omezen na prostý text, případně je z TEI pouze odkazováno na soubor s prostým textem, který je ve formátu prostého textu.

Oxford Text Archive About OTA Electronic Enlightenment CLARIN [Login](#)

OTA Home / View Item

A proclamation

Please use the following text to cite this item or export to a predefined format: [BIBTEX](#) [CITE5](#)

England and Wales, Sovereign (1685-1688 : James II) ; James II, King of England, 1633-1701 ; and Paterson, William, 1658-1719, 2009, A proclamation, Oxford Text Archive, <http://hdl.handle.net/20.500.12024/A46516>

Share: [f](#) [t](#) [s](#)

Authors: England and Wales, Sovereign (1685-1688 : James II) ; James II, King of England, 1633-1701 ; Paterson, William, 1658-1719

Date of publication: 1687

Type: Text

Language(s): English

OTA identifier: ota:A46516

Images: <https://data.historicalextexts.jsc.ac.uk/view?pubid=eebo-ocm13653056>





Collection(s): Early English Books Online (Phase 1)

Show full item record

This item is **Public Domain** and licensed under: CC0-No Rights Reserved

Files for this item

Download all local files for this item (116.37 KB)

Name	A46516.epub	
Size	14.62 KB	
Format	Unknown	
Description	Version of the work for e-book readers in the EPUB format	
Download file		
Name	A46516.html	
Size	10.37 KB	
Format	HTML	
Description	Version of the work for web browsers	
Download file Preview		
Name	A46516.samuels.tsv	
Size	73.83 KB	
Format	Unknown	
Description	Version of the work with linguistic annotation added, in one-word-per-line format, from the SAMUELS project	
Download file		
Name	A46516.xml	
Size	17.53 KB	
Format	XML	
Description	Version of the work in the original source TEI XML file produced from the Text Creation Partnership version	
Download file		

Local Connections

- Electronic Enlightenment
- Bodleian Libraries
- University of Oxford

The Oxford Text Archive is part of the Bodleian Libraries

[Sitemap](#)





Repository

- Home page
- About OTA
- FAQ
- Contact (email)
- Privacy policy

CLARIN Community Connections

- CLARIN UK
- CLARIN
- LINDAT/CLARIN

Powered by CLARIN DSpace

Oxford Text Archive
Copyright (c) 2019 University of Oxford. All rights reserved.

Obrázek č. 24 Detail titulu v Oxford Tetx Archive³⁵

Pro rozdílnost v hloubce kódování bohužel nelze hodnotit použité moduly TEI, ani to, zda některé texty využívají i vlastní schémata. Přesto lze říct, že všechny zveřejněné texty obsahují alespoň některé základní společné rysy. Všechny bezvýhradně splňují minimální strukturu TEI a obsahují tedy odděleně text dokumentu (`<text>`) a jeho hlavičku (`<teiHeader>`), přičemž v ní jsou přítomny všechny povinné elementy (`<fileDesc>`) pro popis elektronického souboru, prohlášení o názvu (`<titleStmt>`), prohlášení o publikaci (`<publicationStmt>`) a popis zdrojového dokumentu (`<sourceDesc>`). Vedle toho jsou ale vždy v hlavičce uvedeny

³⁵ Zdroj obrázku: A proclamation. In: *Oxford Text Archive* [online digitální knihovna]. University of Oxford, 2019 [cit. 2021-5-2]. Dostupné z: <https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/A46516>

i další, pro minimální strukturu TEI nepovinné, údaje. Konkrétně se jedná o popis klasifikačních údajů v elementu (<profileDesc>) a popis kódování (<encodingDesc>), jehož rozsah se u různých dokumentů výrazně liší – někdy se jedná jen o pár vět prostého textu, jindy o kompletně a strukturovaně vysvětlený proces kódování. Často je také přítomen popis provedených revizí elektronického textu (<revisionDesc>).

6. 5. 2 Interoperabilita

Dokumenty v digitální knihovně mají přidělený perzistentní identifikátor Handle a zároveň mají vlastní unikátní identifikátor „OTA“, který je založený na systému PURL. Samotná digitální knihovna pracuje s metadaty v Dublin Core, která slouží k vyhledávání, ale jsou i volně přístupná a zobrazitelná. Vedle toho je uživatelům nabízena jasná podoba citace daného dokumentu, která je dostupná ke stažení i ve formátech BibTeX³⁶ a CMDI.

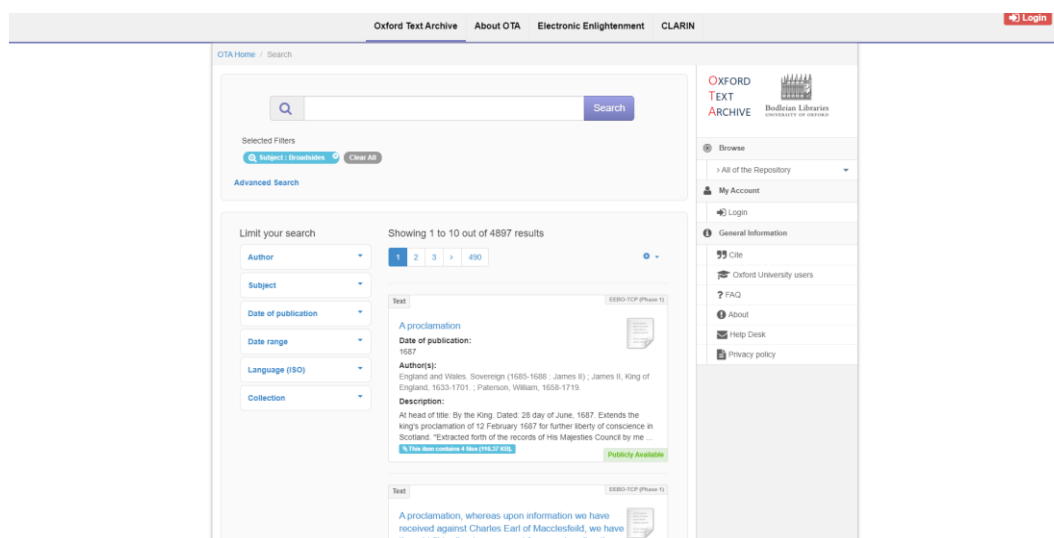
OTA pracuje na bázi nabídek – instituce i jednotliví uživatelé mohou vkládat své vlastní práce a pokud splní určení podmínky, budou jejich texty do systému přidány. Veškerý obsah je dostupný skrze rozhraní OAI-PMH, přes které jsou vystavována data v TEI, DC a OLAC³⁷.

6. 5. 3 Webová použitelnost

Úvodní obrazovka Oxford Text Archive nabízí na první pohled výrazné pole pro vyhledávání mezi dokumenty a hned pod ním možnosti pro procházení obsahu archivu podle předmětu, časového období vzniku primárního dokumentu nebo podle kolekcí. Zároveň jsou na úvodní stránce uživatelům některé dokumenty rovnou doporučovány. Prostředí pro procházení dokumentů nebo výsledků vyhledávání je organizováno do tří sloupců, v levém jsou k dispozici fasety pro zpřesnění vyhledávání, uprostřed jsou řazeny karty jednotlivých nalezených titulů, přičemž kliknutím na název daného titulu se lze dostat na jeho detail (viz Obrázek č. 24 výše). V pravém sloupci je potom dostupné jednoduché menu pro informace o systému a jeho nastavení. Použitá omezení vyhledávání (filtry) jsou zobrazeny pod vyhledávacím řádkem a kliknutím na ně je lze zrušit. (Obrázek č. 25).

³⁶ BibTeX je souborovým formátem a zároveň nástrojem používaným pro práci s odkazy a literaturou v prostředí LaTeX [121]

³⁷ OLAC je formát používaný komunitou Open Language Archives pro výměnu metadat v rámci iniciativy Open Archives (OAO). Metadaty v OLAC vychází z Dublin Core, jsou rozšířena pro specifické použití touto komunitou [122]



Obrázek č. 25 Pohled na obrazovku vyhledávání v OTA³⁸

Nalezené problémy v oblasti použitelnosti se týkají především funkce vyhledávání, které v případě, kdy na vyhledávací dotaz nejsou nalezeny žádné výsledky, nepomáhá uživateli se sestavením jiného dotazu nebo neodkazuje na nápovědu. Vyhledávání se pokouší předcházet možným uživatelským chybám pomocí našeptavače, ale zároveň v pokročilém vyhledávání nijak neomezuje dotazy pro prohledávání konkrétních polí. Například pro pole „Rights“ (práva) jsou k dispozici jen dvě hodnoty, které vrací výsledky. Systém ale i přes to umožňuje zadání jakéhokoli textového řetězce do tohoto pole. Výrazně lepším řešením by byl uzavřený seznam, který by umožnil pouze výběr z některých hodnot. Obdobně nejsou nijak omežována pole vyhledávání v datech publikování nebo rozmezí dat vydání primárních dokumentů. Zároveň je nutné podotknout, že tyto omezující funkce jsou v OTA dostupné, ale ne z pokročilého vyhledávání, nýbrž z položky „All of the Repository“ v pravém postranním menu. Tato implementace se zdá neintuitivní a nekonzistentní.

Vedle zmíněných problémů ve vyhledávání lze spatřovat určité použitelnostní nedostatky v oblasti flexibility systému, neboť nedisponuje žádnou úrovní personalizace, a to ani po přihlášení uživatele. Pro příjemnější a rychlejší používání by bylo vhodné zvážit přidání některých funkcí – například nabízení doporučených dokumentů na základě navštívených titulů nebo umožnění sestavování vlastních kolekcí či označování oblíbených titulů.

³⁸ Zdroj obrázku: Search. In: *Oxford Text Archive* [online digitální knihovna]. University of Oxford, 2019 [cit. 2021-5-2]. Dostupné z: https://ota.bodleian.ox.ac.uk/repository/xmlui/discover?filtertype=subject&filter_relational_operator>equals&filter=Broadsides

Nalezený problém	Odpovídající heuristika	Možné řešení
Vyhledávání neomezuje zadávané hodnoty	Předcházení chybám, Konzistence a standardy	Zavedení omezení pro některé případy, například formou výběru z uzavřeného seznamu
Není dostupná žádná úroveň personalizace	Flexibilita a efektivita	Implementování personalizačních funkcí, například doporučování titulů nebo označování „oblíbených“
Pokud nejsou na vyhledávací dotaz nalezeny žádné výsledky, systém nenabízí řešení	Rozpoznávání a opravování chyb	Navrhovat uživateli jiné vyhledávací dotazy

Tabulka č. 5 Souhrn problémů s použitelností nalezených v Oxford Text Archive

6.6 Projekt Gutenberg

Projekt Gutenberg představuje digitální knihovnu a také iniciativu zaměřenou na digitalizaci knih a jejich distribuci. Lze jej považovat za nejstarší digitální knihovnu, neboť byl založen již v roce 1971. Své poslání definuje Projekt Gutenberg velmi jednoduše: „*Podporovat vytváření a distribuci elektronických knih*“. Spoléhá přitom na práci dobrovolníků z celého světa. V současnosti obsahuje tato digitální knihovna více než 60 000 knih, které jsou zdarma dostupné (podle autorského práva platného v USA). Kvůli závislosti na práci dobrovolníků, neklade digitální knihovna téměř žádné požadavky na to, jaké knihy a jak mají být digitalizovány. Aby byla práce co nejsnazší, nejsou navíc od dobrovolníků navíc vyžadována téměř žádná metadata k textům. [112, 113]

6.6.1 Prezentace textů

Texty samotné nejsou uživatelům v prostředí digitální knihovny nijak prezentovány – knihovna slouží v podstatě jen jako repozitář nabízející stažení souborů obsahující textové přepisy děl. Hlavním formátem, který je dostupný u všech titulů je prostý text. Ten vzniká v drtivé většině případů automatickým OCR, ale prochází následnou korekcí. Nověji zpracované knihy ale bývají stále častěji doplněny o další formáty jako PDF, mobi nebo ePub, které obsahující například i obrázky. Podle informací dostupných na webových stránkách projektu je možné zveřejnit text v prakticky jakémkoliv formátu, v praxi jsou ale upřednostňovány formáty otevřené a upravitelné. [112, 113]

Text v této digitální knihovně tak nemusí být – a často není – nijak kódován. V knihovně se ale i přesto nachází celkem 410 titulů³⁹, které jako jeden ze souborů ke stažení nabízí i TEI XML dokument. Nicméně rozsah kódování se u jednotlivých děl výrazně liší. Vzhledem k tomu že zveřejňovat TEI XML je dobrovolné, nedochází na straně Projektu Gutenberg ani k žádné kontrole struktury tohoto formátu a je tak možné, že u některých titulů není dodržena minimální povinná struktura nebo že dokument není dobře vytvořen (viz kapitola 3.1).

6.6.2 Interoperabilita

Jediným identifikátorem společným všem dokumentům je „EBook-No.“ - pravděpodobně přírůstkové číslo specifické pro digitální knihovnu. Nejspíš jej nelze považovat za perzistentní identifikátor v prvním smyslu slova. Záznamy v digitální knihovně jsou doplněny o stručná popisná metadata prezentovaná spolu s titulem. Metadata obsahují základní informace o

³⁹ Údaj k 2. 5. 2021 pocházející přímo z prostředí digitální knihovny

autorství, názvu titulu, jazyku, klasifikaci v rámci systému Library of Congress a několik dalších údajů. Tyto informace jsou v databázi udržovány ve formě MARC záznamů, jejich individuální stažení ale umožněno není.

Podle dostupných informací nedisponuje Projekt Gutenberg OAI-PMH rozhraním. Data z něj je možné stáhnout a uložit či použít v jiném projektu několika jinými způsoby: Všechny metadatové záznamy mohou být hromadně (v jednom souboru) staženy ve formátu MARC nebo jako prostý text. Je také možné odebírat RSS kanál pro monitorování nově přidaných titulů v digitální knihovně. Zároveň nabízí Project Gutenberg také možnost stažení veškerých souborů, a to prostřednictvím nástroje wget. [114]

6. 6. 3 Webová použitelnost

Na úvodní stránce digitální knihovny jsou návštěvníkovi nabízeny nejnověji přidané tituly, informace o digitální knihovně, nápověda k jejímu používání a podmínky, kterými se zpřístupňování elektronických textů řídí. V záhlaví stránky je dostupné vyhledávací pole a další možnosti pro procházení obsahu digitální knihovny. Jednotlivé dokumenty jsou potom prezentovány na samostatných stránkách obsahujících fotku obálky nebo titulní stranu knihy, vedle níž se nachází sekce, kde jsou vypsány formáty souborů, ve kterých je text dostupný ke stažení. Pod seznamem formátů se nachází nabídka podobných titulů a stručný bibliografický záznam (Obrázek č. 26).

Právě na obrazovce detailu titulu se lze setkat s prvním problémem v oblasti webové použitelnosti – popisná metadata týkající se daného díla se nachází na stránce až úplně dole, pod nabídkou textů ke stažení a doporučenými souvisejícími tituly. Toto řazení neodpovídá běžnému provedení v jiných digitálních knihovnách a může uživatele zmást. Problematická je i samotná sekce nabízející stažení souborů, neboť zde nejsou prezentovány všechny dostupné soubory, na některé se lze dostat až po kliknutí na tlačítko „More files“. Tato akce ovšem přenesení uživatele do prostředí obsahujícího pouhý výčet souborů určité složky na serveru digitální knihovny, není zde žádné formátování ani grafické elementy spojené s digitální knihovnou (viz Obrázek č. 27). S tímto prostředím je spojen další problém použitelnosti, a to že se v něm nenachází odkaz zpět na detail dokumentu, respektive je zde dostupná šipka označená textem „Parent Directory“, která ovšem neslouží k návratu zpět, ale k zobrazení nadřazené složky na úložišti digitální knihovny. Tato složka není, z pochopitelných důvodů, veřejně dostupná a uživatel se tak octne na prázdné obrazovce bez možnosti cokoli provést.

Project Gutenberg > 65 147 free ebooks > 38 by Jean de La Fontaine

Fables de La Fontaine by Jean de La Fontaine



Download This eBook

Format	Size			
Read this book online HTML	1.1 MB			
EPUB (with images)	94.7 MB			
EPUB (no images)	880 kB			
Kindle (with images)	223.0 MB			
Kindle (no images)	2.3 MB			
Plain Text UTF-8	566 kB			
More Files...				

Similar Books

Readers also downloaded...

- [In FR Philosophie, Religion et Morale](#)
- [In FR Contes](#)
- [In FR Poésie](#)
- [In FR Illustrateurs](#)






Bibliographic Record

Author	La Fontaine, Jean de, 1621-1695
Illustrator	Grandville, J.-J., 1803-1847
Title	Fables de La Fontaine
Language	Francois21na
LoC Class	PQ Language and Literatures Romance literatures French Italian Spanish Portuguese
Subject	Aesop's fables -- Adaptations
Subject	Fables, Greek -- Adaptations
Subject	Fables, French
Category	Text
EBook-No.	56327
Release Date	7. 1. 2018
Copyright Status	Public domain in the USA.
Downloads	503 downloads in the last 30 days.
Price	\$0.00

Obrázek č. 26 Detail titulu v Projektu Gutenberg⁴⁰

⁴⁰ Zdroj obrázku: Fables de La Fontaine by Jean de La Fontaine. In: *Project Gutenberg* [online digitální knihovna]. [cit. 2021-5-2]. Dostupné z: <https://www.gutenberg.org/ebooks/56327>

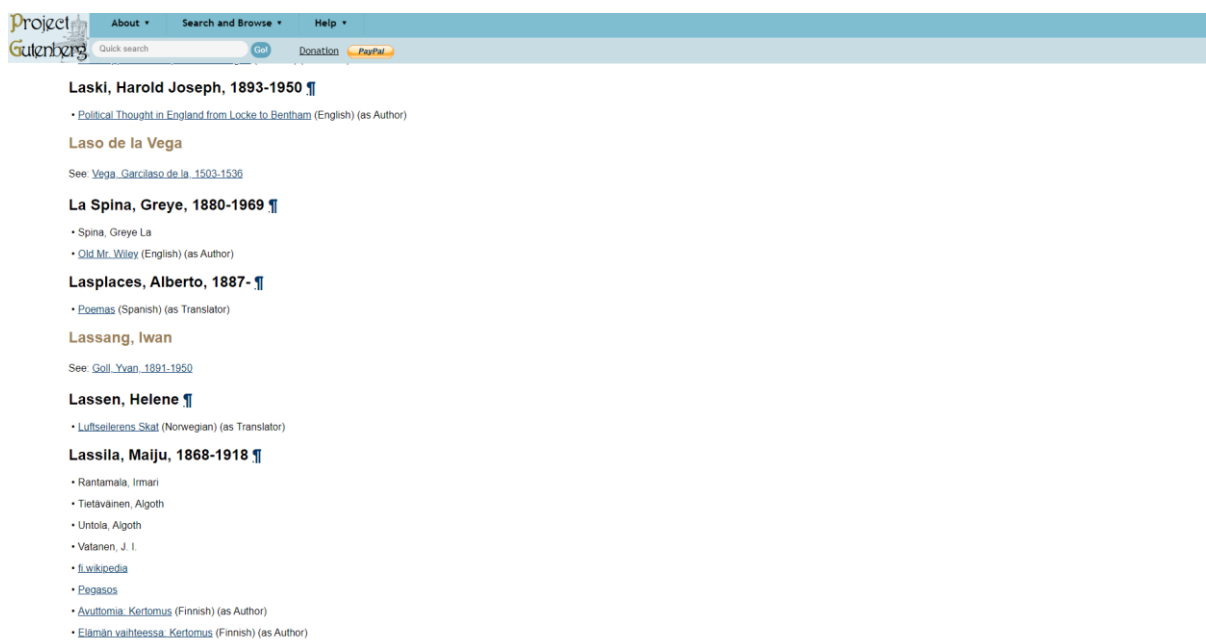
Index of /files/56327

Name	Last modified	Size	Description
 Parent Directory			-
 56327-0.txt	2018-01-07 10:02	566K	
 56327-0.zip	2018-01-07 10:03	210K	
 56327-h.zip	2018-01-07 10:03	93M	
 56327-h/	2018-01-07 12:20		-

Apache Server at www.gutenberg.org Port 80

Obrázek č. 27 Obrazovka pro stažení všech dostupných souborů, Projekt Gutenberg

Prostředí sloužící pro procházení obsahu digitální knihovny není konzistentní, podle způsobu filtrování obsahu jsou tituly někdy organizovány do tabulek, jindy do dlouhých seznamů. V některých případech jsou opatřeny náhledy obálek, ty ale někdy jsou reálné a jindy jen ilustrační. V případech procházení obsahu podle titulů daného jazyka nebo jmen autorů, jsou výsledky prohledávání organizovány do velmi dlouhých seznamů, ve kterých je složité se vyznat. Tomu nenapomáhá ani fakt, že tyto seznamy jsou mezi sebou propojeny takovým způsobem, který není uživateli nijak skrze uživatelské prostředí komunikován. Například při procházení obsahu knihovny podle jazyka jsou díla řazena dle jména autora. Některá jména autora slouží zároveň jako odkaz, skrze který se dostane uživatel na seznam jmen autorů a jejich děl, jiná jména ale vedou rovnou na titul daného autora. Prokliky z těchto seznamů navíc trvají poměrně dlouho a ve chvíli načítání není dávána uživateli dostatečná zpětná vazba. Vhodným řešením těchto problémů by bylo nejspíš kompletní přepracování uživatelského prostředí pro procházení obsahu.



Obrázek č. 28 Procházení obsahu digitální knihovny podle jména autora, Projekt Gutenberg⁴¹

Jako problematická se jeví také nápověda pro uživatele Projektu Gutenberg. Ta je zpracována velmi dobře a je na ní odkazováno jak z úvodní stránky digitální knihovny, tak jejích jednotlivých částí. Problém představuje způsob odkazování, který není jednotný – v některých případech se jedná o žluté tlačítko s nápisem „Help“, jindy o tlačítko modrého otazníku. Jako ještě zásadnější problém se jeví fakt, že ne všechna tlačítka jsou funkční a v některých případech se tak může uživatel místo do nápovědy dostat jen na stránku oznamující nenalezení požadovaného obsahu. Všechny nalezené problémy shrnuje tabulka č. 6.

⁴¹ Zdroj obrázku Browse By Author: L. In: Project Gutenberg [online digitální knihovna]. [cit. 2021-5-2]. Dostupné z: <https://www.gutenberg.org/browse/authors/l>

Nalezený problém	Odpovídající heuristika	Možné řešení
Nepřehledné a vizuálně nekonzistentní prostředí pro procházení obsahu	Viditelnost stavu systému, Konzistence a standardy	Přepracování prostředí do jednotné podoby
Neposkytování zpětné vazby v případě provádění déle trvajících akcí	Viditelnost stavu systému	Implementace zprávy pro uživatele, že systém akci registroval a pracuje na ní
Nabídka umožňující stažení souborů z digitální knihovny má dvě velmi rozdílné podoby	Konzistence a standardy	Nabízení možnosti stáhnout soubory pouze jedním způsobem
Řazení oddílů „stažení souborů“ a „bibliografického záznamu“ neodpovídá konvencím	Konzistence a standardy	Výměna pozic obou sekcí
Chybějící nebo špatně odkazující tlačítko „zpět“	Předcházení chybám	Doplnění funkčního tlačítka „zpět“ na všechny obrazovky digitální knihovny
Některé nabídky nápovědy vedou na neexistující stránky	Konzistence a standardy, Pomoc a dokumentace	Oprava odkazů nápovědy, případně doplnění potřebných stránek

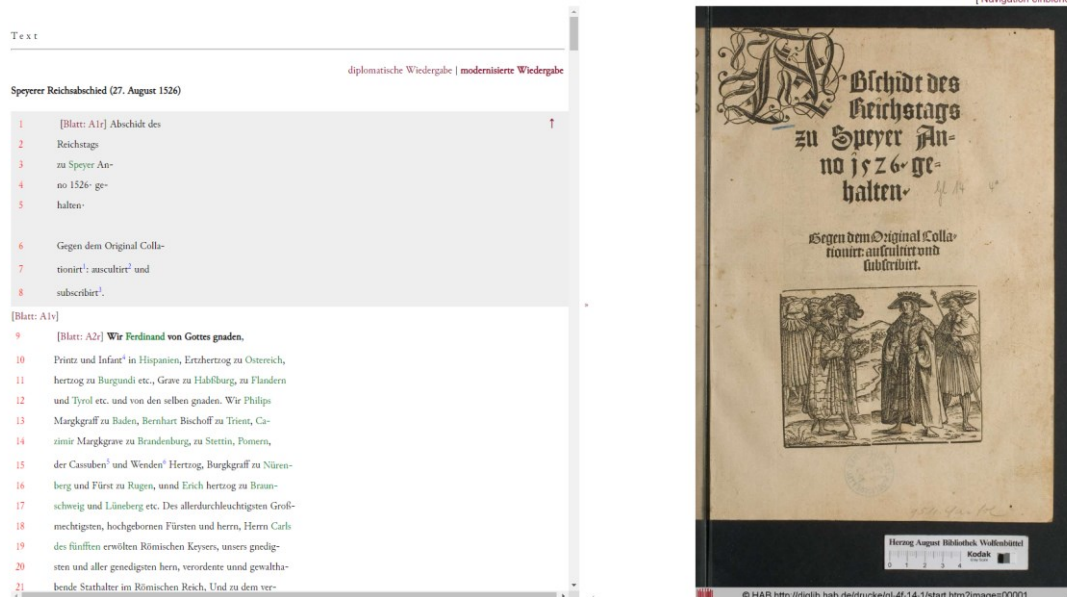
Tabulka č. 6 Souhrn problémů s použitelností nalezených v Projektu Gutenberg

6. 7 Religiöse Friedenswahrung und Friedensstiftung in Europa

Digitální knihovna Religiöse Friedenswahrung und Friedensstiftung in Europa 1500-1800 (v překladu Náboženské uchovávání a budování míru) představuje digitální vydání náboženských mírových dohod pocházející z období mezi lety 1500 a 1800. Knihovna představuje unikátní zdroj pro srovnávací výzkum v oblasti budování náboženského míru v období raného novověku. Obsah digitální knihovny tvoří přibližně 40 textů náboženských mírových smluv, které jsou doplněny podrobnými úvody a komentáři. Digitální knihovna tak jako celek umožňuje komplexně analyzovat politické, právní a teologické podmínky spojené s vyjednáváním náboženského míru, použité argumentační modely, a dosažená řešení, a to při zasazení do kontextu a vzájemných souvislostí. [115]

6. 7. 1 Presentace textů

Digitální knihovna prezentuje čtenářům texty prostřednictvím XML databázového systému eXist. Uživatelské prostředí je velmi jednoduché a nabízí čtenáři pouze menu v levém postranním panelu, které slouží k výběru dokumentu. Texty jsou zde rozděleny do 4 kategorií podle časových období. Po rozkliknutí konkrétního časového období se zobrazí seznam názvů jednotlivých textů. Kromě toho je z levého menu přístupný ještě úvod k digitální edici a vyhledávání ve jménech a místech zmíněných v textech. Při prohlížení samotných dokumentů je uživateli k dispozici v levé straně obrazovky prepis textu, který má dvě podoby – normalizovanou a původní (více dále). V pravé části obrazovky je potom u většiny textů jednoduchá prohlížečka skenů digitální faksimile (Obrázek č. 29).



Obrázek č. 29 Prohlížení dokumentu v digitální knihovně *Religiöse Friedenswahrung*...⁴²

Samotné texty i jejich úvody jsou kódovány v TEI. Jak TEI XML dokumenty, tak XSL transformace sloužící k jejich zobrazení jsou u každého textu dostupné ke stažení. Z veřejně přístupných informací není zřejmé, jak přepisy textů vznikly, v úvahu přichází ruční přepis nebo automatické OCR doplněné ovšem o velké množství lidské práce. Kódování textů odpovídá přibližně čtvrté úrovni, tak jak je uvedena v kapitole 4.1. Označováno je každé slovo (značka `<w>`), vybraná slova jsou navíc rozepsána dvěma způsoby umožňujícími prohlížení dvou variant textu – originální podobu a normalizovanou, snadněji čitelnou podobu. Vytvoření těchto dvou verzí je realizováno použitím elementu `<choice>` uvnitř elementu `<w>` tak, že originální podoba slova je zapisována pomocí značky `<orig>` a normalizovaná pomocí `<reg>`. Texty jsou doplněny o poznámky (`<note>`) a zvýraznění (`<hi>`). Jednotlivé elementy jsou hojně doprovázeny atributem pro popis jejich vykreslení (`@rend`). Hlavní předností textů v této digitální knihovně je, že zakódovány jsou veškeré osoby a místa, která se v textech vyskytují. Přímou v přepisu textu se pro jejich označení používá element `<rs>` (reference string) s atributy `@type` nabývající hodnoty "place" pro místo a "person" pro osobu, a `@ref` pro odkaz na definici daného místa nebo osoby. Definice jsou obsaženy v záhlaví TEI dokumentu v části `<profileDesc>`. Jednotlivá místa jsou zde definována v elementu `<settingDesc>`. Každá

⁴² Zdroj obrázku Speyerer Reichsabschied (27. August 1526). In: *Religiöse Friedenswahrung und Friedensstiftung in Europa 1500–1800: Digitale Quellenedition frühneuzeitlicher Religionsfrieden* [online digitální knihovna]. [cit. 2021-5-2]. Dostupné z: http://exist.ulb.tu-darmstadt.de:8080/exist/apps/edoc/view.html?id=e000001_speyerer_reichsabschied

definice obsahuje název, ID a URL odkaz do databáze Německé národní knihovny nebo portálu Geonames.org na záznam o daném místě. Osoby jsou definovány v rámci elementu `<praticDesc>`, každá osoba má ID, a jméno rozdělené na křestní jméno a příjmení (pokud je to možné), případně je přítomna poznámka a je-li jméno spojeno s názvem nějakého místa, pak je pomocí `@xml:id` propojeno toto jméno s definicí místa.

Protože schéma použité pro kódování textů není dostupné, nelze přesně určit, z jaké moduly byly pro kódování použity. Je ovšem možné určit, že schéma je postavené na Zásadách TEI-P5 a že bylo rozšířeno, a to pravděpodobně v rámci potřeby kódovat jména osob a míst. Například výše zmíněný element `<praticDesc>` není součástí Zásad TEI. Stejně tak je patrné že minimálně definice některých atributů byly upraveny.

6. 7. 2 Interoperabilita

Dokumenty v digitální knihovně nejsou identifikovány žádným perzistentním identifikátorem. Jedinou identifikaci tvoří kombinace blíže neurčitého kódu a názvu dokumentu (například `e000001_schlesischer_majestaetsbrief_1609_zdt`) uvedená jako `@xml:id` elementu `<TEI>` v TEI XML dokumentu. Záhlaví tohoto dokumentu také slouží jako jediný zdroj metadat. Zde jsou obsaženy informace odpovídající popisu elektronického souboru (`<fileDesc>`), prohlášení o názvu (`<titleStmt>`), prohlášení o publikaci (`<publicationStmt>`) a popis zdroje (`<sourceDesc>`). Tyto údaje jsou také nutné pro splnění minimální struktury Zásad TEI. Kromě toho jsou dostupné ještě informace o kódování (`<encodingDesc>`), popis klasifikačních údajů v elementu `<profileDesc>` včetně definic osob a míst a také je přítomen popis provedených revizí elektronického textu (`<revisionDesc>`).

Není dostupná žádná informace o tom, že by byly dokumenty přístupné v této digitální knihovně agregovány (ať už pomocí OAI-PMH nebo jinak), pravděpodobně se tak neděje. Stejně tak knihovna neagreguje záznamy z jiných knihoven

6. 7. 3 Webová použitelnost

Prostředí digitální knihovny (popsané v podkapitole 5.7.1) je velmi jednoduché a zásadě nehrozí, že by uživatel nevěděl, v jaké jeho části se nachází. Problém ale může představovat navigování mezi jednotlivými částmi: Například z hlavní stránky se nelze dostat na vyhledávání v plných textech dokumentů, ale jen na hledání mezi osobami a místy. Odkaz na prohledávání plných textů je umístěn pouze na obrazovce prohlížení dokumentu. Z vyhledávání dále není

An error has occurred

An error has been generated by the application.

wdb0000 no file with ID edoc_ed000227_fg_passauer_vertrag in /db/apps/edoc/data [at line 414, column 7, source: /db/apps/edoc/modules/app.xqm]

In function:

```
wdb:getFilePath(xs:string) [220:15:/db/apps/edoc/modules/app.xqm]
wdb:populateModel(xs:string, xs:string, map(*)) [208:3:/db/apps/edoc/modules/app.xqm]
wdb:getEE(node(), map(*), xs:string, xs:string) [-1:-1:/db/apps/edoc/modules/app.xqm]
templates:process-output(element(), map(*), item(*), element()) [211:9:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:call-by-introspection(element(), map(*), map(*), function(*)) [189:28:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:call(item(), element(), map(*)) [137:36:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:process(node(*), map(*)) [428:17:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:process-output(element(), map(*), item(*)) [230:9:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:process-output(element(), map(*), item(*), element()) [211:9:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:call-by-introspection(element(), map(*), map(*), function(*)) [189:28:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:call(item(), element(), map(*)) [137:36:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:process(node(*), map(*)) [133:51:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:process(node(*), map(*)) [90:9:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
templates:apply(node()+, function(*), map(*)?, map(*)?) [51:5:/srv/existdb/data/expathrepo/shared-0.8.4/content/templates.xql]
```

Obrázek č. 31 Chybová zpráva ve vyhledávání v digitální knihovně Religiöse Friedenswahrung...

V záhlaví kterékoliv z obrazovek digitální knihovny není přítomné logo nebo název knihovny, které by zároveň mohly sloužit jako způsob pro návrat na hlavní stránku. Jedná se přitom v podstatě o standard v oblasti webového designu [116]. Veškeré zjištěné problémy a návrhy možných řešení jsou shrnuty v tabulce č. 7.

Nalezený problém	Odpovídající heuristika	Možné řešení
Při vyhledávání není poskytována zpětná vazba	Viditelnost stavu systému	Implementace zprávy pro uživatele, že systém akci registroval a pracuje na ní
Chybějící tlačítko „zpět“	Uživatel má mít kontrolu a svobodu	Doplnění tlačítka „zpět“
Výsledky vyhledávání neobsahují název titulu, ve kterém byla nalezena shoda	Uživatel má mít kontrolu a svobodu	Umožnit podrobnější popis výsledků vyhledávání
Chybějící logo v levém horním rohu pro návrat na úvodní stranu	Konzistence a standardy	Doplnění loga nebo názvu s odpovídajícím odkazem
Vyhledavač umožňuje vyhledávání i v bázích nenáležících k digitální knihovně	Konzistence a standardy	Omezení možností vyhledávání
Vyhledavač nedisponuje našeptavačem	Předcházení chybám	Doplnění našeptavače
Chybové zprávy zobrazované systémem nejsou srozumitelné a konstruktivní	Rozpoznávání a opravování chyb	Přepracování chybových zpráv tak, aby naváděly uživatele k řešení problému
Neexistuje uživatelská nápověda	Pomoc a dokumentace	Vytvoření uživatelské nápovědy a náležité odkazování na ni

Tabulka č. 7 Souhrn problémů s použitelností nalezených v digitální knihovně Religiöse Friedenswahrung...

7 Shrnutí a vyhodnocení analýzy

Digitální knihovny umožňují jednoduchý přístup k digitálním objektům a vyhledávání v nich. Představují proto důležitý způsob pro rozšiřování přístupu ke kulturnímu dědictví jak mezi vědci a profesionály, tak i širokou veřejností. Texty přitom mohou být veřejnosti prezentovány různými způsoby: jako obrazy naskenovaných stránek primárního dokumentu, jako samostatné textové přepisy, kombinací obou předchozích způsobů nebo vizualizací kódovaných dokumentů, například pomocí XSL transformací. Poslední jmenovaný způsob je zároveň velmi výhodný, neboť může sloužit i k prezentaci komplexních textů obsahujících například několik variant textu (variantu se striktně přepsanými znaky, variantu se standardizovaným pravopisem apod.), kritický aparát nebo poznámky včleněné přímo do textu.

Analyzované digitální knihovny pokrývají všechny zmíněné možnosti prezentace textů, od publikace samotných skenů stránek až po pouhé zveřejnění přepisů, které ve dvou případech nejsou ani k nahlédnutí v prostředí digitální knihovny, ale pouze nabízené ke stažení. V těch knihovnách, které obsahují textové přepisy, vznikly tyto přepisy buď ručně nebo procesem OCR, který byl ale vždy doplněn o korekci.

V současnosti je široce přijímaným způsobem pro kódování textů používání značkovacích jazyků, které umožňují přidat dodatečné informace přímo do digitálně reprezentovaného textu. V zásadě jediným uznávaným standardem na tomto poli jsou zásady definované konsorciem Text Encoding Initiative. Tento fakt potvrzuje i provedená analýza digitálních knihoven, v rámci které bylo zjištěno, že pouze jedna zkoumaná digitální knihovna (Projekt Gutenberg) se Zásadami TEI nijak neparčuje. Knihovny, které nedisponují textovými přepisy, využívají TEI alespoň k vytvoření metadatových záznamů – pracují tedy zejména se záhlavím TEI. Kromě podrobného popisu dokumentu otevírá tento přístup dveře i budoucímu rozšíření záznamů právě o textový přepis, který může být do současného TEI XML dokumentu poměrně jednoduše doplněn.

Samotné konsorcium TEI neoznačuje své zásady za standard, ale hovoří spíše o pravidlech a doporučeních. Je tomu tak i kvůli jejich značné robustnosti – aby byly relativně univerzální a mohly sloužit ke kódování jakéhokoli žánru textu z různých časových období v různých jazycích, obsahují více než 500 elementů. Není přitom předpokládáno, že v praxi budou v rámci kódování jednoho textu nebo jedné kolekce využity všechny tyto elementy. Kódovací schéma TEI je složeno z jednotlivých modulů (některých povinných, jiných volitelných), které je

možné mezi sebou kombinovat a případně i doplňovat tak, aby mohlo být vytvořeno specifické kódovací schéma odpovídající potřebám konkrétního digitalizačního projektu, které bude ale zároveň stále kompatibilní s TEI.

Ve zkoumaných digitálních knihovnách, které disponují prepisem textů a bylo možné zároveň analyzovat i použité kódovací schéma, byly ze Zásad TEI vždy použity všechny povinné moduly a některé volitelné. Z volitelných modulů se nejčastěji opakovaly moduly *figure*, *linking*, *msdescription* a *transr*. Ve všech těchto případech bylo zároveň schéma TEI upraveno pro specifické potřeby projektu. V případech knihoven Deutsches Textarchiv a Manuscriptorium bylo schéma upraveno jeho omezením, tedy výběrem nejen modulů, ale i odstraněním nebo upravením některých elementů, atributů a hodnot, kterých mohou nabývat. V případech knihoven Medieval Nordic Text Archive a Religiöse Friedenswahrung und Friedensstiftung in Europa bylo schéma na jednu stranu omezeno, ale zároveň také doplněno o několik málo nových elementů. Ve většině případů byla tato upravená kódovací schémata volně dostupná ke stažení a doprovázena i širokou dokumentací pro jejich použití.

Úroveň kódování nebylo ve většině případů možné hodnotit, nicméně lze konstatovat, že v digitálních knihovnách obsahujících méně a o to pečlivěji vybraných textů, byla úroveň spíše vyšší, zatímco rozsáhlejší digitální knihovny disponují spíše dokumenty s úrovní velice rozdílnou. Tři z analyzovaných digitálních knihoven nabízí několik verzí či pohledů na jeden text, ale způsoby prezentace těchto verzí se ve všech třech případech liší. Deutsches Textarchiv volí cestu různých vizualizací stejného textu pomocí rozdílných XSL transformací, knihovna Religiöse Friedenswahrung und Friedensstiftung in Europa používá pro kódování dvou verzí nativní elementy Zásad TEI a Medieval Nordic Text Archive umožňuje realizaci rozdílných verzí skrze vlastní přidaný modul, který zavádí nový jmenný prostor obsahující elementy pro odlišení prepisů jednotlivých variant.

Přesuneme-li pozornost od kódování textu k problematice digitálních knihoven, můžeme zde sledovat vzrůstající trendy kladoucí stále větší důraz na aspekty otevřenosti a interoperability, jako je používání otevřených standardů, jednoznačných trvalých identifikátorů a nástrojů umožňující sdílení, znovupoužívání a agregaci dat a jejich metadat. Akcentovány jsou v těchto souvislostech například protokol OAI-PMH, aplikační rámec IIF nebo propojená data (linked data).

Z hlediska interoperability lze shrnout, že 4 ze 7 analyzovaných digitálních knihoven opatřují svoje dokumenty perzistentním identifikátorem, zajímavé může být, že se vždy jedná o jiný

(URN:NBN, DOI, URI, Handle). Stejně 4 digitální knihovny také disponují rozhraním OAI-PMH. Digitální knihovny e-codices a Manuscriptorium pracují také s aplikačním rámcem IIF. Perzistentní identifikaci nepoužívají dvě menší digitální knihovny a Project Gutenberg. Ten také jako jediný poskytuje jen velmi malé množství deskriptivních metadat, která jsou navíc dostupná pouze ve formátu MARC. Ostatní digitální knihovny využívají nejčastěji TEI (respektive záhlaví TEI dokumentu) a Dublin Core, případně jejich kombinaci doplněnou o další formáty.

Vedle zaměření na interoperabilitu a otevřenost se objevuje také sílící důraz na orientaci digitálních knihoven na své uživatele a na jejich zapojení. Právě uživatelé by měli být pro digitální knihovny prioritou. Je totiž třeba si uvědomit, že zatímco kódované (nebo jinak zpracované) dokumenty mohou sloužit k dlouhodobému uchování textu, digitální knihovny představují ve světle rychle se rozvíjejících technologií spíše než prostor pro jejich uchování, pouhý prostředek k jejich prezentaci. Tímto tvrzením není snižována hodnota digitálních knihoven, pouze je více do popředí stavěn fakt, že digitální knihovny by neměly být budovány pro dokumenty, ale pro své uživatele. Proto je velmi důležité dbát v kontextu digitálních knihoven na podobu uživatelského rozhraní, které stojí mezi čtenářem a kolekcí dokumentů. Pozornost by měla být ve vztahu k uživatelskému prostředí věnována problematice přístupnosti (tedy tomu, zda může rozhraní sloužit i uživatelům s různými druhy hendikepu) a aspektům použitelnosti. Ty zahrnují koncepty jako jsou snadnost naučení se práce s uživatelským prostředím, efektivita, zapamatovatelnost funkcionalit, předcházení chybám a způsoby jejich řešení a v neposlední řadě spokojenost uživatelů.

Určitý standard v oblasti webové použitelnosti představují pravidla definovaná Jakobem, Nielsenem, na kterých byla také postavena odpovídající část analýzy. Jako nejčastěji porušované zásady webové použitelnosti byly identifikovány „Rozpoznávání a opravování chyb“, „Předcházení chybám“, „Konzistence a standardy“ a „Viditelnost stavu systému“. Závažnější problémy byly zjištěny (s výjimkou Projektu Gutenberg) u menších digitálních knihoven. Jednoznačně nejčastějším problémem projevujícím se téměř u všech analyzovaných knihoven jsou problémy v oblasti vyhledávání, zejména nedostatečná podpora uživatele ve chvíli, kdy na jeho vyhledávací dotaz nejsou nalezeny žádné výsledky. V takovém případě by měl systém ideálně pomoci uživateli s opravou či úpravou jeho dotazu anebo navrhnout alternativní postup. Důležitým aspektem, který se zdá být také často opomínán, je dostatečně podrobně a strukturovaně zpracovaná nápověda, na kterou by mělo být z uživatelského prostředí vhodnými způsoby odkazováno.

Ačkoli analýza pravidel použitelnosti digitálních knihoven byla provedena jednoduchou heuristickou analýzou, jasně se ukazuje, že některé nedostatky v oblasti webové použitelnosti mají všechny zkoumané digitální knihovny. Opět je zřejmé, že menší digitální knihovny trpí často větším počtem závažnějších chyb. Této problematice by určitě bylo záslužné se více věnovat a provést hlubší a podrobnější analýzy jak kvantitativního, tak kvalitativního typu. I bez nich je ale možné upozornit na to, že použitelnosti digitálních knihoven není v současnosti pravděpodobně věnována taková pozornost, jakou by si především jejich uživatelé zasloužili.

Výsledky praktické části práce ukazují na fakt, že kódování a příprava textů a způsoby jejich prezentace v digitálních knihovnách jsou do jisté míry často chápány jako dva ne příliš související problémy. Z provedené analýzy lze vyvodit, že důraz bývá vždy kladen spíše na jeden z těchto aspektů – buď digitální knihovna obsahuje velmi kvalitně kódovaný text, ale zaostává v oblastech prezentace a interoperability, nebo v těchto oblastech žádné výrazné nedostatky nemá, ale přepisy textů jsou kódovány v nižší míře podrobnosti. Toto pozorování lze samozřejmě postavit i do kontextu toho, že menší a specializovanější digitální knihovny mívají precizněji zpracované kódování menšího počtu textů, zatímco velké knihovny obsahují nesrovnatelně větší množství textů, které ale mají velmi rozdílné úrovně kódování, obsahují-li vůbec kódovaný text. I na základě relativně malého vzorku zkoumaných digitálních knihoven je možné upozornit na nedostatečný důraz, který menší a specializovanější projekty kladou na aspekty interoperability a znovupoužitelnosti svých dokumentů. Ačkoliv nejviditelnějším výstupem digitalizace bývá prezentace textů v prostředí digitálních knihoven, jsou to samotné (TEI) dokumenty a procesy spojené s jejich vznikem, jako je analýza textu, vytvoření kódovacího schématu a metodiky pro kódování, které jsou nejdůležitějšími výstupy digitalizačních projektů, neboť umožňují další výzkum ale i dlouhodobé uchování textů i do nich vložených odborných znalostí.

Souhrnný pohled na výsledky provedené analýzy nabízí tabulky č. 8, 9 a 10.

Prezentace textu v digitální knihovně ⁴⁵							
	DTA	e-codices	Manusc.	MNTA	OTA	PG.	RFFE
Množství dokumentů	6500	2500	150000	50	33000	60000	40
Forma prezentace	skeny, textový přepis	skeny	záznamy a skeny	skeny, textový přepis	přepisy textů ke stažení	přepisy textů ke stažení	skeny, textový přepis
Vznik přepisu	OCR, korekce	-	-	ruční přepis	OCR, korekce, ruční přepis	OCR, korekce	ruční přepis
Text kódován v TEI	Ano	záhlaví TEI	záhlaví TEI	Ano	Ano	Ne	Ano
Dostupnost TEI	Ano	Ano	Ano	Ano	Ano	-	Ano
Použité schéma	vlastní, omezení TEI	nelze určit	vlastní, omezení TEI	vlastní, rozšíření TEI	různá schémata	-	vlastní, rozšíření TEI
Úroveň kódování	3. úroveň	-	-	různá úroveň	různá úroveň	-	4. úroveň
Moduly TEI	core, corpus, drama, figures, header, linking, msdescription, namesdates, tei, textcrit, textstructure, transcr	nelze určit	core, figures, header, linking, msdescription, namesdates, tei, textstructure, transcr	analysis, core, corpus, drama, figures, header, linking, msdescription, namesdates, tagdocs, tei, textcrit, textstructure, transcr, verse	nelze určit	-	nelze určit

Tabulka č. 8 Souhrn zjištění v oblasti prezentace textů v digitálních knihovnách

Interoperabilita ⁴⁵							
	DTA	e-codices	Manusc.	MNTA	OTA	PG.	RFFE
Perzistentní identifikátor	URN:NBN	DOI, vlastní	vlastní	Ne	Handle, vlastní	Ne	Ne
Formáty metadat	TEI, DC, CMDI	TEI	TEI	TEI	DC,	MARC	TEI
Nástroje pro sdílení a agregaci	OAI-PMH	OAI-PMH, IIF	OAI-PMH, IIF	Ne	OAI-PMH	RSS, wget	Ne

Tabulka č. 9 Souhrn zjištění v oblasti interoperability

Počet porušení pravidel webové použitelnosti ⁴³							
	DTA	e-codices	Manusc.	MNTA	OTA	PG	RFFE
Viditelnost stavu systému	1x	1x	-	1x	-	2x	1x
Shoda mezi systémem a reálným světem	-	-	-	-	-	-	-
Uživatel má mít kontrolu a svobodu	1x	1x	-	-	-	-	2x
Konzistence a standardy	-	-	-	-	1x	4x	2x
Předcházení chybám	1x	-	1x	1x	1x	1x	1x
Rozpoznávání před rozpomináním	-	1x	-	1x	-	-	1x
Flexibilita a efektivita	1x	-	-	-	1x	-	-
Minimalistický design	1x	-	1x	-	-	-	-
Rozpoznávání a opravování chyb	1x	1x	2x	2x	1x	-	1x
Pomoc a dokumentace		1x	1x	1x	-	1x	1x

Tabulka č. 10 Souhrn zjištění v oblasti pravidel webové použitelnosti

⁴³ Použité zkratky:

DTA: Deutsches Textarchiv

e-codices : e-codices – Virtual Manuscript Library of Switzerland

Manusc.: Manuscriptorium

MNTA: Medieval Nordic Text Archive

OTA: Oxford Text Archive

PG: Projekt Gutenberg

RFFE: Religiöse Friedenswahrung und Friedensstiftung in Europa

8 Závěr

Stěžejním tématem bakalářské práce bylo podat přehled o standardech používaných pro kódování předmoderních textů a představit způsoby jejich prezentace veřejnosti prostřednictvím digitálních knihoven.

V teoretické části práce byly představeny zásady pro kódování textu definované konsorciem Text Encoding Initiative. Tyto zásady jsou v podstatě jediným široce uznávaným standardem na poli kódování textu, a to i proto, že jsou výsledkem velmi rozsáhlé mezinárodní spolupráce zohledňující různá hlediska, zájmy a způsoby práce s textem. Současná verze zásad nesoucí název TEI-P5 je postavena na značkovacím jazyce XML, tento jazyk byl proto krátce představen a následně byla popsána struktura a modularita kódovacího schématu TEI a jeho možnosti pro popis textových i netextových prvků.

Vedle Zásad TEI byl v práci popsán také proces přípravy kódovaného textu, různé způsoby prezentace textů v digitálních knihovnách a současné trendy v této oblasti. Ty zahrnují problematiku interoperability, trvalé identifikace dokumentů, otevřenosti a zapojení uživatelů. Představeny byly také základní principy použitelnosti webu.

V praktické části bakalářské práce byly analyzovány přístupy sedmi digitálních knihoven k problematice kódování textu, interoperabilitě a webové použitelnosti. Jako základ pro analýzu byl z poznatků poskytnutých literárně přehledovou částí sestaven kontrolní seznam, podle něhož byly informace zjišťovány. Přístupy jednotlivých knihoven k prezentaci textu, jeho kódování i případné hloubce kódování se značně lišily, nicméně lze konstatovat, že kromě jedné knihovny všechny pracují se Zásadami TEI. Dále byly také popsány přístupy jednotlivých knihoven k problematice interoperability a byly identifikovány některé nedostatky v oblasti webové použitelnosti.

Ačkoliv proces přípravy textů včetně jejich kódování a způsob jejich následné prezentace v digitálních knihovnách představují dva úzce související aspekty digitalizace, výsledky praktické části ukazují na častou nevyváženost mezi těmito hledisky. Oběma by přitom měla být věnována náležitá pozornost.

9 Seznam použité literatury

1. HORÁK, František, Emma URBÁNKOVÁ a Bedřiška WIŽĎÁLKOVÁ. *Pravidla jmenné katalogizace starých tisků, prvotisků a rukopisů*. Praha: SPN, 1971. Dostupné také z: <https://ndk.cz/uuid/uuid:f5e298e0-7f22-11e2-a455-005056827e52>
2. ČERNÁ, Alena M. a Boris LEHEČKA. *Metodika přípravy a zpracování elektronických edic starších českých textů*. Praha: Oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i., 2015. Dostupné z: <http://invenio.nusl.cz/record/203171>
3. PANDEY, Prabhat a Roli MISRA. Digitization of Library Materials in Academic Libraries: Issues and Challenges. *Journal of Industrial and Intelligent Information* [online]. 2014, 2(2), 136-141 [cit. 2021-03-19]. ISSN 23013745. Dostupné z: doi:10.12720/jiii.2.2.136-141
4. NYHAN, Julianne. Chapter 6: Text encoding and scholarly digital editions. In: WARWICK, Claire, Melissa TERRAS a Julianne NYHAN. *Digital humanities in practice*. London: Facet Publishing in association with UCL Centre for Digital Humanities, 2012, s. 117-138. ISBN 978-1-85604-766-1.
5. BURDICK, Anne, Johanna DRUCKER, Peter LUNENFELD, Todd Samuel PRESNER a Jeffrey T. SCHNAPP. *Digital humanities*. Praha: Academia, 2019. XXI. století. ISBN 978-80-200-2865-5.
6. RENEAR, Allen H. 17: Text Encoding. In: SCHREIBMAN, Susan, Ray SIEMENS, a John UNSWORTH. *A companion to Digital humanities*. Malden: Blackwell, 2004. ISBN 1-4051-0321-3.
7. History. In: *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-06]. Dostupné z: <https://tei-c.org/about/history/>
8. Design Principles for Text Encoding Guidelines: TEI ED P1. *Text Encoding Initiative* [online]. [cit. 2021-03-19]. Dostupné z: <https://tei-c.org/Vault/ED/edp01.htm>
9. P5 Guidelines. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-06]. Dostupné z: <https://tei-c.org/guidelines/p5/>
10. Module 0: Introduction to Text Encoding and the TEI. *TEI by Example* [online]. 2020 [cit. 2021-03-06]. Dostupné z: <https://ibyexample.org/tutorials/TBED00v00.htm>

11. PIERAZZO, Elena. Textual Scholarship and Text Encoding In: SCHREIBMAN Susan, Ray SIEMENS a John UNSWORTH. *A New Companion to Digital Humanities* [online]. Chichester, UK, 2015, 307-321 [cit. 2021-03-19]. ISBN 9781118680605. Dostupné z: doi:10.1002/9781118680605.ch21
12. Goals and Mission. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-06]. Dostupné z: <https://tei-c.org/about/mission/>
13. HAWKINS, Kevin S. Introduction to XML for Text. In: *Kevin S. Hawkins: Assistant Dean for Scholarly Communication* [online]. 2020 [cit. 2021-03-07]. Dostupné z: <http://www.ultraslavonic.info/intro-to-xml/>
14. Principy formátu XML. In: MLÝNKOVÁ, Irena a Jaroslav POKORNÝ. *XML technologie: principy a aplikace v praxi* [online]. První vydání. Praha: Grada, 2008, s. 15-46 [cit. 2021-03-07]. Průvodce (Grada). ISBN 978-80-247-2725-7.
15. BIRNBAUM, David J. What is XML and why should humanists care?: An even gentler introduction to XML. *Obdurodon.org* [online]. [cit. 2021-03-07]. Dostupné z: <http://dh.obdurodon.org/what-is-xml.xhtml>
16. TEI CONSORTIUM. 1 The TEI Infrastructure: The TEI Guidelines. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-08]. Dostupné z: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html>
17. VANHOUTTE, Edward a Ron VAN DEN BRANDEN, ed. Text Encoding Initiative (TEI). In: *Encyclopedia of Library and Information Sciences: Third Edition*. 3rd Edition. Boca Raton, Florida, USA: CRC Press, 2009, 6106 s. ISBN 9780203757635.
18. Module 1: Common Structure, Elements, and Attributes. *TEI by Example* [online]. [cit. 2021-03-09]. Dostupné z: <https://teibyexample.org/tutorials/TBED01v00.htm>
19. The TEI Header: P5 Guidelines for Electronic Text Encoding and Interchange. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-09]. Dostupné z: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>
20. Default Text Structure: P5 Guidelines for Electronic Text Encoding and Interchange. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-10]. Dostupné z: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>
21. Elements Available in All TEI Documents: P5 Guidelines for Electronic Text Encoding and Interchange. *Text Encoding Initiative* [online]. [cit. 2021-03-10]. Dostupné z: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html>

22. Verse: P5 Guidelines for Electronic Text Encoding and Interchange. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-10]. Dostupné z: <https://tei-c.org/release/doc/tei-p5-doc/en/html/VE.html>
23. Appendix C Elements: P5 Guidelines for Electronic Text Encoding and Interchange. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-13]. Dostupné z: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html>
24. Customization. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-13]. Dostupné z: <https://tei-c.org/guidelines/customization/>
25. Introducing the Guidelines. *TEI: Text Encoding Initiative* [online]. [cit. 2021-03-13]. Dostupné z: <https://tei-c.org/support/learn/introducing-the-guidelines/>
26. RAHTZ, Sebastian a Lou BURNARD. Reviewing the TEI ODD system. *Proceedings of the 2013 ACM symposium on Document engineering* [online]. New York, NY, USA: ACM, 2013, 193-196 [cit. 2021-03-13]. ISBN 9781450317894. Dostupné z: doi:10.1145/2494266.2494321
27. Getting Started with P5 ODDs. *Text Encoding Initiative* [online]. [cit. 2021-03-13]. Dostupné z: <https://tei-c.org/guidelines/customization/getting-started-with-p5-odds/>
28. MÜLLER, Ermenegilda. A TEI Customization for Paper and Watermarks Descriptions. *Digital Medievalist* [online]. 2020, **13**(1) [cit. 2021-03-20]. ISSN 1715-0736. Dostupné z: doi:10.16995/dm.91
29. TRYON, Julia R. *The Rosarium Project: A case of merging traditional reference librarian skills with digital humanities technology* [online]. 2017, **24**(2-4), 171-188 [cit. 2021-03-14]. ISSN 1069-1316. Dostupné z: doi:10.1080/10691316.2017.1329043
30. HAWKINS, Kevin, Michelle DALMAU, Elli MYLONAS a Syd BAUMAN. Best Practices for TEI in Libraries: A guide for mass digitization, automated workflows, and promotion of interoperability with XML using the TEI. *Text Encoding Initiative* [online]. [cit. 2021-03-14]. Dostupné z: <https://tei-c.org/extra/teiinlibraries/4.0.0/bptl-driver.html>
31. WISNESKI, Richard a Virginia DRESSLER. Implementing TEI Projects and Accompanying Metadata for Small Libraries: Rationale and Best Practices. *Journal of Library Metadata* [online]. 2009, **9**(3-4), 264-288 [cit. 2021-03-14]. ISSN 1938-6389. Dostupné z: doi:10.1080/19386380903405173
32. Using the TEI: P5 Guidelines for Electronic Text Encoding and Interchange. *Text Encoding Initiative* [online]. [cit. 2021-03-15]. Dostupné z: <https://tei-c.org/release/doc/tei-p5-doc/en/html/USE.html#CF>

33. Editors. In: *TEIWiki* [online]. [cit. 2021-03-15]. Dostupné z: <https://wiki.tei-c.org/index.php/Editors>
34. TURSKA, Magdalena, James CUMMINGS a Sebastian RAHTZ. Challenging the Myth of Presentation in Digital Editions. *Journal of the Text Encoding Initiative* [online]. 2016, no. 9 [cit. 2021-03-15]. ISSN 2162-5603. Dostupné z: doi:10.4000/jtei.1453
35. CELBOVÁ, Ludmila. Digitální knihovna. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna České republiky, 2003.
36. FOX, Edward A. a Rajat GUPTA. *Introduction to Digital Libraries: Definitions* [online]. [cit. 2021-04-16]. Dostupné z: <https://ei.cs.vt.edu/~dlib/rintro.htm>
37. KATUŠČÁK, Dušan. *Informačná výchova: terminologický a výkladový slovník*. 1. vyd. Bratislava: Slovenské pedagogické nakladateľstvo, 1998. ISBN 80-08-02818-1.
38. WITTEN, I. H., David BAINBRIDGE a David M. NICHOLS. *How to build a digital library*. 2nd ed. Burlington: Morgan Kaufmann, 2010. Morgan Kaufmann series in multimedia information and systems. ISBN 978-012-3748-577.
39. KALISDHA, A. a C. SURESH. Digital Libraries: Definitions, Issues and Challenges. *International Journal of Science and Humanities* [online]. Islamiah College Publications, 2017, 3(12), 95-104 [cit. 2021-04-16]. ISSN 2394 9236. Dostupné z: <http://www.islamiahcollege.edu.in/images/downloads/2341NACQLIS%202018%20JOURNAL.pdf#page=109>
40. RANGANATHAN, S. R. *The Five Laws of Library Science*. Madras: Madras Library Association, 1931.
41. KWANYA, Tom, Christine STILWELL a Peter G. UNDERWOOD. Library 2.0 principles and Ranganathan's fifth law. *Mousaion* [online]. 2010, 28(2), 1-16 [cit. 2021-04-16]. ISSN 0027-2639. Dostupné z: <https://hdl.handle.net/10520/EJC79007>
42. NORUZI, Alireza. Application of Ranganathan's Laws to the Web. *Webology* [online]. 1(2) [cit. 2021-04-16]. Dostupné z: <http://hdl.handle.net/10150/105179>
43. Key themes and challenges in digital libraries. *Exploring Digital Libraries* [online]. Facet, 2018, s. 59-84 [cit. 2021-04-18]. ISBN 9781783300297. Dostupné z: doi:10.29085/9781783300297.005

44. RAEMY, Julien Antoine. *Enabling better aggregation and discovery of cultural heritage content for Europeana and its partner institutions* [online]. Ženeva, 2020 [cit. 2021-04-18]. Dostupné z: https://julsraemy.github.io/assets/doc/Mastersthesis_europeana_raemyjulien_FV.pdf. Diplomová práce. Haute école de gestion de Genève, Information Science Department. Vedoucí práce Arnaud Gaudinat.
45. Open Archives Initiative Protocol for Metadata Harvesting: Interoperability through Metadata Exchange. *Open Archives Initiative* [online]. [cit. 2021-04-18]. Dostupné z: <https://www.openarchives.org/pmh/>
46. CORRADO, Edward M. *Discovery Products and the Open Archives Initiative Protocol for Metadata Harvesting* [online]. 2018, **50**(1), 47-53 [cit. 2021-04-18]. ISSN 1057-2317. Dostupné z: doi:10.1080/10572317.2017.1422905
47. projektu. *Česká Digitální Knihovna* [online]. [cit. 2021-04-18]. Dostupné z: <https://www.czechdigitallibrary.cz/archiv/>
48. Vytěžování Manuscriptoria přes OAI-PMH. *Manuscriptorium: Digital Library of Written Cultural Heritage* [online]. Národní knihovna České republiky. [cit. 2021-04-18]. Dostupné z: <http://www.manuscriptorium.com/cs/oai-pmh-tezeni>
49. IIF Frequently Asked Questions (FAQs). *International Image Interoperability Framework* [online]. [cit. 2021-04-18]. Dostupné z: <https://iif.io/community/faq/>
50. JULIEN ANTOINE, Raemy a Schneider RENÉ. *Suggested measures for deploying IIF in Swiss cultural heritage institutions* [online]. Geneva, Switzerland: HES-SO University of Applied Sciences and Arts, Haute école de gestion de Genève. [cit. 2021-04-18]. Dostupné z: doi:10.5281/zenodo.2640416
51. *Report on Preparing Europeana for IIF involvement Task Force: Appendix B: Frequently Asked Questions* [online]. 2017. Dostupné také z: https://pro.europeana.eu/files/Europeana_Professional/Europeana_Network/Europeana_Network_Task_Forces/Final_reports/Preparing%20Europeana%20for%20IIF%20involvement%20TF%20final%20report/report-on-iif-task-force-2017-appendixb.pdf
52. RAZA, Zahid, Khalid MAHMOOD a Nosheen Fatima WARRAICH. Application of linked data technologies in digital libraries: a review of literature. *Library Hi Tech News* [online]. 2019, **36**(3), 9-12 [cit. 2021-04-18]. ISSN 0741-9058. Dostupné z: doi:10.1108/LHTN-10-2018-0067

53. NEISH, Peter. Linked data: what is it and why should you care? *The Australian Library Journal* [online]. 2015, **64**(1), 3-10 [cit. 2021-04-18]. ISSN 0004-9670. Dostupné z: doi:10.1080/00049670.2014.974004
54. MEINDERTSMA, Joep. A brief introduction to linked data. *Ontola.io: Linked Data Software Development* [online]. [cit. 2021-04-18]. Dostupné z: <https://ontola.io/what-is-linked-data/>
55. ABRAMS, Stephen, Patricia CRUSE a John KUNZE. Preservation Is Not a Place. *International Journal of Digital Curation* [online]. 2009, **4**(1), 8-21 [cit. 2021-04-18]. ISSN 1746-8256. Dostupné z: doi:10.2218/ijdc.v4i1.72
56. VAŠEK, Zdeněk, Martin ŘEHÁNEK a Ladislav CUBR. Metodika pro přidělování a správu životního cyklu unikátních perzistentních identifikátorů digitálních dokumentů podle standardu URN:NBN: verze 2.0 (2018): Český systém pro IDentifikaci a LOkalizaci digitálních dokumentů českého kulturního dědictví. Verze 2.0. 2019. Dostupné také z: <http://www.nusl.cz/ntk/nusl-408011>
57. BRATKOVÁ, Eva. *Sít' trvalých identifikátorů informačních entit*. Verze 1.0. Praha: Ústav informačních studií a knihovnictví FF UK v Praze, 2007. Dostupné také z: https://sites.ff.cuni.cz/uisk/wp-content/uploads/sites/62/2016/01/S%3%ad%5%a5-trval%3%bdch-identifik%3%a1tor%5%af-informa%4%8dn%3%adch-entit_Bratkov%3%a1.pdf
58. BERNERS-LEE, T., R. FIELDING a L. MASINTER. RFC: 3986. *Uniform Resource Identifier (URI): Generic Syntax* [online]. 2005, 61 s. Dostupné také z: <https://tools.ietf.org/html/rfc3986>
59. MOATS, R. RFC: 2141. *URN Syntax* [online]. 1997, 7 s. Dostupné také z: <https://tools.ietf.org/html/rfc2141>
60. ELLIS, M., B. KELLY a R. GARDLER. What Does Openness Mean To The Museum Community?. In: *Museums and the Web 2008. Proceedings* [online]. Montreal, Quebec, Canada: Archives & Museum Informatics [cit. 2021-04-18]. Dostupné z: https://purehost.bath.ac.uk/ws/files/383395/kelly_opus.pdf
61. KRISHNAMURTHY, M. Open access, open source and digital libraries. *Program: electronic library and information systems* [online]. 2008, **42**(1), 48-55 [cit. 2021-04-18]. ISSN 0033-0337. Dostupné z: doi:10.1108/00330330810851582

62. BROMHEAD, Ben. 10 advantages of open source for the enterprise. *Opensource.com* [online]. Red Hat, Inc. [cit. 2021-04-18]. Dostupné z: <https://opensource.com/article/17/8/enterprise-open-source-advantages>
63. *Greenstone digital library software* [online]. New Zealand Digital Library Project, University of Waikato, New Zealand, 2005-2015 [cit. 2021-04-18]. Dostupné z: <https://www.greenstone.org/>
64. *Omeka* [online]. [cit. 2021-04-18]. Dostupné z: <https://omeka.org>
65. *Existdb* [online]. 2018 [cit. 2021-04-18]. Dostupné z: <http://exist-db.org/exist/apps/homepage/index.html>
66. *Kramerius* [online]. Knihovna AV ČR, v. v. i., 2021 [cit. 2021-04-18]. Dostupné z: <https://system-kramerius.cz>
67. KLEINMAN, Molly. *The beauty of "Some Rights Reserved": Introducing Creative Commons to librarians, faculty, and students* [online]. 2008, **69**(10), 594-597 [cit. 2021-04-18]. ISSN 2150-6698. Dostupné z: doi:10.5860/crln.69.10.8077
68. Licence CC. *Creative Commons Česká republika* [online]. 2021 [cit. 2021-04-18]. Dostupné z: <https://www.creativecommons.cz/licence-cc/>
69. HANNESSCHLÄGER, Vanessa. Common Creativity International: CC-licensing and Other Options for TEI-based Digital Editions in an International Context. *Journal of the Text Encoding Initiative* [online]. 2019, no. 11 [cit. 2021-04-18]. ISSN 2162-5603. Dostupné z: doi:10.4000/jtei.2610
70. Social Annotations in Digital Library Collections. *D-Lib Magazine* [online]. 2008, **14**(11/12) [cit. 2021-04-18]. ISSN 1082-9873. Dostupné z: doi:10.1045/november2008-gazan
71. BERTI, Monica, Bridget ALMAS, David DUBIN, Greta FRANZINI, Simona STOYANOVA a Gregory R. CRANE. The Linked Fragment: TEI and the Encoding of Text Reuses of Lost Authors. *Journal of the Text Encoding Initiative* [online]. 2014, no. 8 [cit. 2021-04-18]. ISSN 2162-5603. Dostupné z: doi:10.4000/jtei.1218
72. MATUSIAK, Krystyna K. *Towards user-centered indexing in digital image collections* [online]. 2006, **22**(4), 283-298 [cit. 2021-04-18]. ISSN 1065-075X. Dostupné z: doi:10.1108/10650750610706998

73. NEUHOLD, Erich, Claudia NIEDERÉE a Avaré STEWART. Personalization in Digital Libraries: An Extended View. In: *Digital Libraries: Technology and Management of Indigenous Knowledge for Global Access: 6th International Conference on Asian Digital Libraries, ICADL 2003 Kuala Lumpur, Malaysia, December 8-12, 2003 Proceedings*. Berlin: Springer, 2003, s. 1-17. ISBN 3-540-20608-6. ISSN 0302-9743.
74. Optical Character Recognition (OCR): How it works. *Nicomsoft* [online]. [cit. 2021-04-18]. Dostupné z: <https://www.nicomsoft.com/optical-character-recognition-ocr-how-it-works/>
75. ZAVŘELOVÁ, Alžběta. Projekt PERO: OCR pro historické texty. *Duha: Informace o knihách a knihovnách z Moravy* [online]. 2020, **34**(4) [cit. 2021-04-18]. ISSN 1804-4255. Dostupné z: <https://duha.mzk.cz/clanky/projekt-pero-ocr-pro-historicke-texty>
76. MATUŠÍK, Zdeněk. Faksimile. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna České republiky [cit. 2021-04-18]. Dostupné z: https://aleph.nkp.cz/F/?func=direct&doc_number=000000975&local_base=KTD
77. About the World Digital Library. *World Digital Library* [online]. [cit. 2021-04-18]. Dostupné z: <https://www.wdl.org/en/about/>
78. LEE, Haimin. 15 years of Google Books. *The Keyword* [online]. [cit. 2021-04-18]. Dostupné z: <https://www.blog.google/products/search/15-years-google-books/>
79. YOUNG, Michael J. *XML krok za krokem*. 2. vyd. Brno: Computer Press, 2006. ISBN 80-251-1070-2.
80. HOLUBOVÁ, Irena a Jaroslav POKORNÝ. *XML technologie: principy a aplikace v praxi*. Praha: Grada, 2008. Průvodce (Grada). ISBN 978-80-247-2725-7.
81. TEIC / Stylesheets. In: *GitHub* [online repozitář]. [cit. 2021-04-18]. Dostupné z: <https://github.com/TEIC/Stylesheets>
82. WALSH, John, Grant SIMPSON a Saeed MOADDELI. About TEI Boilerplate. *TEI Boilerplate* [online]. [cit. 2021-04-18]. Dostupné z: <https://dcl.ils.indiana.edu/teibp/>
83. SIEGEL, Erik a Adam RETTER. *EXist: a NoSQL document database and application platform*. Sebastopol: O'Reilly, 2014. ISBN 978-144-9337-100.

84. TURCO, Roberto Rosselli Del, Giancarlo BUOMPRISCO, Chiara Di PIETRO, Julia KENNY, Raffaele MASOTTI a Jacopo PUGLIESE. Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions. *Journal of the Text Encoding Initiative* [online]. 2014, no. 8 [cit. 2021-04-18]. ISSN 2162-5603. Dostupné z: doi:10.4000/jtei.1077
85. ABRAHÃO, Silvia, Cristina CACHERO a Matera MARISTELLA. Web usability and accessibility. *Journal of Web Engineering* [online]. 2008, 7(4) [cit. 2021-04-17]. Dostupné z: https://www.researchgate.net/publication/262234352_Web_usability_and_accessibility
86. NIELSEN, Jakob a Hoa LORANGER. *Prioritizing web usability*. Berkeley: New Riders, 2006. ISBN 0-321-35031-6.
87. NIELSEN, Jakob. Usability 101: Introduction to Usability. *Nielsen Norman Group: World Leaders in Research-Based User Experience* [online]. 2012 [cit. 2021-04-17]. Dostupné z: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>
88. ISO 9241-11:2018 Ergonomics of human-system interaction: Part 11: Usability: Definitions and concepts. 2018.
89. NIELSEN, Jakob. 10 Usability Heuristics for User Interface Design. *Nielsen Norman Group: World Leaders in Research-Based User Experience* [online]. 1998-2021 [cit. 2021-04-17]. Dostupné z: <https://www.nngroup.com/articles/ten-usability-heuristics/>
90. JOHNSON, Jeff. *Designing with the mind in mind: simple guide to understanding user interface design rules*. Burlington: Morgan Kaufmann, 2010. ISBN 978-0-12-375030-3.
91. NORMAN, Donald A. *The Design of Everyday Things: Revised and Expanded Edition*. New York: Basic Books, 2013. ISBN 978-0-465-05065-9.
92. HENRY, Shawn. Introduction to Web Accessibility. *W3C* [online]. [cit. 2021-04-17]. Dostupné z: <https://www.w3.org/WAI/fundamentals/accessibility-intro/>
93. FERNANDEZ, A., S. ABRAHÃO a E. INSFRAN. *A systematic review on the effectiveness of Web usability evaluation methods* [online]. IET, 2012, 52-56 [cit. 2021-04-18]. ISBN 978-1-84919-541-6. Dostupné z: doi:10.1049/ic.2012.0007
94. TAN, Wei-siong, Dahai LIU a Ram BISHU. Web evaluation: Heuristic evaluation vs. user testing. *International Journal of Industrial Ergonomics* [online]. 2009, 39(4), 621-627 [cit. 2021-04-18]. ISSN 01698141. Dostupné z: doi:10.1016/j.ergon.2008.02.012

95. ALSHAMARI, Majed a Pam MAYHEW. Technical Review: Current Issues of Usability Testing. *IETE Technical Review* [online]. 2009, **26**(6) [cit. 2021-04-18]. ISSN 0256-4602. Dostupné z: doi:10.4103/0256-4602.57825
96. MVUNGI, Joel a Titus TOSSY. Usability Evaluation Methods and Principles for the Web.) *International Journal of Computer Science and Information Security* [online]. 2015, **13**(7), 86-92 [cit. 2021-04-18]. ISSN 1947-5500. Dostupné z: <https://sites.google.com/site/ijcsis/vol-13-no-7-jul-2015>
97. Projektüberblick. *Deutsches Textarchiv* [online]. Berlin: Brandenburgische Akademie der Wissenschaften, 2007 [cit. 2021-05-02]. Dostupné z: <https://deustextarchiv.de/doku/ueberblick>
98. GEYKEN, Alexander, Susanne HAAF, Bryan JURISH, Matthias BOENIG, Christian THOMAS a Frank WIEGAND. *ODD file for the DTA 'Base Format' for printed texts (DTABf)* [online]. Version 1.1.0. Berlin: Deutsches Textarchiv, 2017 [cit. 2021-5-2]. Dostupné z: <https://www.deustextarchiv.de/basisformat.odd>
99. GEYKEN, Alexander, Susanne HAAF, Bryan JURISH, Matthias BOENIG, Christian THOMAS a Frank WIEGAND. *ODD file for the DTA 'Base Format' for manuscripts (DTABf-M)* [online]. Version 1.1.0. Berlin: Deutsches Textarchiv, 2020. Dostupné také z: https://www.deustextarchiv.de/basisformat_ms.odd
100. Deutsches Textarchiv. *EADH: The European Association for Digital Humanities* [online]. [cit. 2021-05-02]. Dostupné z: <https://eadh.org/projects/deustextarchiv>
101. Brief History. *E-codices: Virtual Manuscript Library of Switzerland* [online]. swissuniversities [cit. 2021-05-02]. Dostupné z: <https://www.e-codices.unifr.ch/en/about/history>
102. PADLINA, Roberta a Christoph FLÜELER. E-codices Newsletter: Issue N° 12. *E-codices: Virtual Manuscript Library of Switzerland* [online]. [cit. 2021-05-02]. Dostupné z: <https://www.e-codices.unifr.ch/newsletter/archive/issue-12.htm>
103. List of the Institutions whose Data are in Manuscriptorium. *Manuscriptorium: Digital Library of Written Cultural Heritage* [online]. Národní knihovna České republiky [cit. 2021-05-02]. Dostupné z: <http://www.manuscriptorium.com/en/aggregated-resources>
104. Manuscriptoriu. *Manuscriptorium: Digital Library of Written Cultural Heritage* [online]. Národní knihovna České republiky [cit. 2021-05-02]. Dostupné z: <http://www.manuscriptorium.com/cs/o-manuscriptoriu>

105. Agregace dokumentů v Manuscriptoriu: základní informace. *Manuscriptorium: Digital Library of Written Cultural Heritage* [online]. Národní knihovna České republiky [cit. 2021-05-02]. Dostupné z: <http://www.manuscriptorium.com/cs/agregace-dokumentu-v-manuscriptoriu-zakladni-informace>
106. TEI P5 ENRIH Schéma. *Manuscriptorium: Digital Library of Written Cultural Heritage* [online]. Národní knihovna České republiky [cit. 2021-05-03]. Dostupné z: <http://www.manuscriptorium.com/cs/tei-p5-enrich-schema-cs>
107. Manuscriptorium OAI-PMH repository. *Manuscriptorium: Digital Library of Written Cultural Heritage* [online]. Národní knihovna České republiky [cit. 2021-05-03]. Dostupné z: <http://www.manuscriptorium.com/en/oai-pmh-harvesting>
108. *Medieval Nordic Text Archive: Menota* [online]. 2007 [cit. 2021-05-03]. Dostupné z: https://www.menota.org/EN_forside.xhtml
109. Menota Handbook 3.0: Guidelines for the electronic encoding of Medieval Nordic primary sources. *Medieval Nordic Text Archive: Menota* [online]. [cit. 2021-05-03]. Dostupné z: https://www.menota.org/HB3_index.xml
110. About the Oxford Text Archive. *Oxford Text Archive* [online]. [cit. 2021-05-03]. Dostupné z: <https://ota.bodleian.ox.ac.uk/repository/xmlui/page/about>
111. Oxford Text Archive. *TEI: Text Encoding Initiative* [online]. [cit. 2021-05-03]. Dostupné z: <https://tei-c.org/activities/projects/oxford-text-archive/>
112. HART, Michael S. The Project Gutenberg Mission Statement, by Michael Hart. *Project Gutenberg* [online]. [cit. 2021-04-18]. Dostupné z: https://www.gutenberg.org/about/background/mission_statement.html
113. File Formats Utilized by Project Gutenberg. *Project Gutenberg* [online]. [cit. 2021-04-18]. Dostupné z: https://www.gutenberg.org/help/file_formats.html
114. Offline Catalogs and Feeds. *Project Gutenberg* [online]. [cit. 2021-05-03]. Dostupné z: https://www.gutenberg.org/ebooks/offline_catalogs.html#the-project-gutenberg-catalog-metadata-in-machine-readable-format
115. Einleitung. Religiöse Friedenswahrung und Friedensstiftung in Europa 1500–1800: Digitale Quellenedition frühneuzeitlicher Religionsfrieden [online]. [cit. 2021-05-03]. Dostupné z: http://www.religionsfrieden.de/index.php_article_id=18.html
116. WHITENTON, Kathryn. Centered Logos Hurt Website Navigation. *Nielsen Norman Group: World Leaders in Research-Based User Experience* [online]. [cit. 2021-05-03]. Dostupné z: <https://www.nngroup.com/articles/centered-logos/>

117. RSS 2.0 Specification. *RSS Advisory Board* [online]. [cit. 2021-04-18]. Dostupné z: <https://www.rssboard.org/rss-specification>. Verze 2.0.11.
118. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *O'Reilly* [online]. [cit. 2021-04-18]. Dostupné z: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1>
119. Component Metadata. *CLARIN: European Research Infrastructure for Language Resources and Technology* [online]. CLARIN [cit. 2021-05-02]. Dostupné z: <https://www.clarin.eu/content/component-metadata>
120. Co je Drobečková navigace?. *IT-Slovník.cz* [online]. [cit. 2021-05-02]. Dostupné z: <https://it-slovník.cz/pojem/drobeckova-navigace/>
121. FEDER, Alexander. *BibTex* [online]. 2006 [cit. 2021-05-03]. Dostupné z: <http://www.bibtex.org/>
122. SIMONS, Gary a Steven BIRD. OLAC Metadata. *OLAC: Open Language Archives Community* [online]. [cit. 2021-05-03]. Dostupné z: <http://www.language-archives.org/OLAC/metadata.html>

10 Seznam obrázků

Obrázek č. 1: Kódování zvýraznění.....	26
Obrázek č. 2: Kódování dat	27
Obrázek č. 3 Příklad textu prezentovaného v Projektu Gutenberg.....	47
Obrázek č. 4: Prezentace rukopisu v DigiVatLib	48
Obrázek č. 5: Pohled na prezentaci dokumentu ve World Digital Library.....	50
Obrázek č. 6: Pohled na vyhledávání v plném textu dokumentu v prostředí Google Books ...	50
Obrázek č. 7: Prezentace dokumentu v digitální knihovně HathiTrust	51
Obrázek č. 8: Prezentace dokumentu v Národní digitální knihovně	51
Obrázek č. 9: XSL transformace, [80], upraveno	53
Obrázek č. 10: TEI dokument vizualizovaný pomocí XSL transformace	53
Obrázek č. 11: Příklad prezentace dokumentu za použití nástroje eXist-db	54
Obrázek č. 12 Prostředí pro prohlížení textu, Deutsches.....	65
Obrázek č. 13 Detail titulu v Deutsches Textarchiv.....	68
Obrázek č. 14 Prostředí pro procházení dokumentů, Deutsches Textarchiv.....	69
Obrázek č. 15 Prostředí pro prohlížení dokumentu, e-codices	72
Obrázek č. 16 Souhrnná stránka titulu, e-codices.....	73
Obrázek č. 17 Prostředí pro prohlížení skenů, e-codices, upraveno.....	74
Obrázek č. 18 Pohled na uživatelské prostředí digitální knihovny Manuscriptorium.....	78
Obrázek č. 19 Prohlížení skenu v digitální knihovně Manuscriptorium	79
Obrázek č. 20 Úvodní obrazovka textového archivu.....	85
Obrázek č. 21 Prostředí pro prohlížení textů v Medieval Nordic Text Archive.....	85
Obrázek č. 22 Vyhledávací formulář v Medieval Nordic Text Archive	86
Obrázek č. 23 Chybová zpráva ve vyhledávání v Medieval Nordic Text Archive	87
Obrázek č. 24 Detail titulu v Oxford Tetx Archive	90
Obrázek č. 25 Pohled na obrazovku vyhledávání v OTA.....	92
Obrázek č. 26 Detail titulu v Projektu Gutenberg	96
Obrázek č. 27 Obrazovka pro stažení všech dostupných souborů, Projekt Gutenberg	97
Obrázek č. 28 Procházení obsahu digitální knihovny podle jména autora, Projekt Gutenberg	98
Obrázek č. 29 Prohlížení dokumentu v digitální knihovně Religiöse Friedenswahrung.....	101
Obrázek č. 30 Výsledky vyhledávání v digitální knihovně Religiöse Friedenswahrung... ..	103
Obrázek č. 31 Chybová zpráva ve vyhledávání v digitální knihovně Religiöse Friedenswahrung.....	104

11 Seznam tabulek

Tabulka č. 1 Souhrn problémů s použitelností nalezených v Deutches Textarchiv	70
Tabulka č. 2 Souhrn problémů s použitelností nalezených v e-codices	75
Tabulka č. 3 Souhrn problémů s použitelností nalezených v Manuscriptoriu.....	80
Tabulka č. 4 Souhrn problémů s použitelností nalezených v Medieval Nordic Text Archive.	88
Tabulka č. 5 Souhrn problémů s použitelností nalezených v Oxford Text Archive.....	93
Tabulka č. 6 Souhrn problémů s použitelností nalezených v Projektu Gutenberg	99
Tabulka č. 7 Souhrn problémů s použitelností nalezených v digitální knihovně Religiöse Friedenswahrung.....	105
Tabulka č. 8 Souhrn zjištění v oblasti prezentace textů v digitálních knihovnách.....	110
Tabulka č. 9 Souhrn zjištění v oblasti interoperability	111
Tabulka č. 10 Souhrn zjištění v oblasti pravidel webové použitelnosti.....	111

12 Seznam příkladů

Příklad č. 1: XML dokument	16
Příklad č. 2: Minimální struktura TEI dokumentu.....	20
Příklad č. 3: Jednoduchý TEI dokument.....	22
Příklad č. 4: Popis struktury s použitím číslovaných oddílů.....	24
Příklad č. 5: Popis struktury s použitím nečíslovaných oddílů.....	24
Příklad č. 6: Kódování zvýraznění.....	26
Příklad č. 7: Kódování dat	27
Příklad č. 8: Formátování pomocí CSS přímo v zakódovaném textu.....	29
Příklad č. 9: kód XSL transformace, jejíž výsledek zobrazuje Obrázek č. 10.....	53

13 Seznam příloh

Příloha č. 1: Kontrolní seznam	127
--------------------------------------	-----

Příloha č. 1 – Kontrolní seznam

Kontrolní seznam pro analýzu vybraných digitálních knihoven

Část I. - Prezentace textu v digitální knihovně

- **Forma prezentace textu**
 - Jakou formou jsou texty čtenářům prezentovány?
(Jedná se o samotné skeny, přepis nebo kombinaci?)
 - Pokud je dostupný přepis, jak vznikl?
(ručně, automatickým OCR, prošel korekcí apod.)
- **Kódování**
 - Je text kódován v TEI?
 - Pokud ano, je TEI dokument dostupný?
 - Pokud ano, bylo použito schéma předpřipravené konsorciem TEI, nebo bylo nějak upraveno na míru projektu? V čem případná úprava spočívá?
 - Jaké úrovni kódování TEI odpovídá?
 - Jaké moduly TEI jsou použity?
 - Pokud není kódování provedeno v TEI, jakým způsobem je realizováno?
 - Je dostupných více variant textu?

Část II. - Interoperabilita

- **Perzistentní identifikace**

Mají dokumenty prezentované v digitální knihovně perzistentní identifikátor?

 - Pokud ano, jaký?
- **Zobrazení metadat**

Umožňuje aplikace zobrazení metadat dokumentu?

 - Pokud ano, v jakých formátech jsou metadata dostupná?
- **Agregace dokumentů**

Je digitální knihovna agregována případně je sama agregátorem?

 - Pokud ano, využívá pro agregaci OAI-PMH, nebo jiný způsob?

Část III. - Heuristická pravidla webové použitelnosti

- Viditelnost stavu systému
- Shoda mezi systémem a reálným světem
- Uživatel má mít kontrolu a svobodu
- Konzistence a standardy
- Předcházení chybám
- Rozpoznávání před rozpomínáním
- Flexibilita a efektivita
- Minimalistický design
- Rozpoznávání a opravování chyb
- Pomoc a dokumentace