

Posudek vedoucího bakalářské práce

Filip Kersch, Standardy používané pro kódování předmoderních textů a způsoby jejich prezentace v digitálních knihovnách, bakalářská práce předložená k obhajobě na Ústavu informačních studií a knihovnictví FF UK, Praha 2021.

V současné době se v digitálních knihovnách setkáváme nejen s obrazy stran digitalizovaných (či born-digital) knih nebo časopisů, ale je zde obvykle v nějaké podobě přístupný také plný text dokumentu. Vedle toho je dostupných mnoho digitálních knihoven, které poskytují pouze korpusy textů (a nástroje související s jejich využitím). V oblasti digitální reprezentace textů se jako všudypřítomný standard prosadilo schéma TEI (Text Encoding Initiative), které je užíváno i v digitálních knihovnách. Předkládaná práce se soustředí na způsoby prezentace textů v digitálních knihovnách a související aspekty interoperability a dodržování pravidel použitelnosti.

V úvodní kapitole se student věnuje nezbytným definicím pojmů (předmoderními texty rozumí texty publikované do roku 1800) a také obecnému přehledu o standardech používaných pro kódování textů, přičemž klade důraz především na obecné souvislosti této problematiky, což je vzhledem ke zpracovávanému tématu v pořádku. V následující kapitole přechází přímo ke standardu TEI, který je v současnosti postaven na základě (meta)jazyka XML. Věnuje se stručně i jeho historii, která odráží obecné trendy v oblasti reprezentace (kódování) textů v posledních třiceti letech. Dále se věnuje přímo Zásadám TEI (TEI Guidelines), které představují nejen samotné schéma, ale na obecnější rovině také pravidla a doporučení, kterými je vhodné se řídit při práci s markupem. Představuje zde nejprve jazyk XML, následně celkovou strukturu schématu TEI, které se v současné době skládá z více než dvacítky modulů. Důležitá je zde poslední podkapitola, která pojednává o možnostech přizpůsobení schématu pro specifické potřeby, tedy přístupu, který je v digitálních knihovnách často využíván.

Důležitá je také kapitola nazvaná „Proces přípravy textů“. Student se zde zabývá vytvářením kódování pro dokumenty; na základě literatury rozlišuje pět úrovní kódování od základní (spojované s vytvořením prostého textu pomocí OCR) až po tu nejpodrobnější.

V navazující kapitole věnované prezentaci textů v digitálních knihovnách jsou vedle nezbytných definic pojmů také postiženy trendy v oblasti digitálních knihoven. Jsou zde také představeny způsoby prezentace textů v digitálních knihovnách: prostý text, skeny stránek a kombinace obojího. Prezentovány jsou i zásady použitelnosti webu.

Šestá kapitola, která obsahuje analýzu vybraných knihoven, představuje spolu se sedmou kapitolou (vyhodnocení analýz) vlastní jádro práce. Podrobná pozornost je zde věnována celkem sedmi digitálním knihovnám, z nichž pět můžeme považovat za elektronické textové korpusy a dvě za digitální knihovny zpřístupňující zejména středověké rukopisy. Zatímco textové korpusy jsou si po obsahové (nikoliv formální) stránce vzájemně podobné, digitální knihovny zpřístupňující rukopisy se od nich liší tím, že obsahují buď jen plné texty katalogů, a to včetně přepisů starších (švýcarský portál e-codices), nebo vedle katalogů také plné texty vytvořené ručním přepisem či pomocí OCR starší tištěné edice (Manuscriptorium). Tím se výzkumný terén částečně zneřehledňuje, ale oba zdroje jsou pro výzkum předmoderních textů natolik významné a v mnohém i charakteristické, že bylo určité přínosné zahrnout je do formální analýzy.

Kapitola obsahuje oddíly věnované jednotlivým digitálním knihovnám. V každém z nich následuje po krátkém historizujícím úvodu poměrně obsáhlá charakteristika prezentace textu v dané digitální knihovně, včetně posouzení, o kterou úroveň kódování textu se jedná. Následují

části věnované interoperabilitě a webové použitelnosti včetně tabulky, která shrnuje studentem zjištěné problémy, pokud jde o webovou použitelnost. Všechny závěry jsou podloženy a doplněny ilustrativními screenshoty.

Shrnutí a vyhodnocení analýzy v sedmé kapitole kromě porovnání zkoumaných aspektů nabízí obecnější syntetizující závěry týkající se povahy těchto digitálních knihoven, resp. způsobu prezentace textů v nich. V závěru je velmi instruktivní srovnávací tabulka, jež na rozdíl od tabulek v jednotlivých oddílech nabízí jen údaje, které je možné považovat za profilující.

Po stránce stylistické je text zcela vyhovující a dobře čitelný, nemám připomínek ani ke gramatické stránce textu.

Závěrem konstatuji, že se jedná o velmi zdařilou bakalářskou práci, která se kromě pečlivého zpracování vyznačuje i originálním přístupem studenta k prezentované problematice. Vedle originality oceňuji zejména vzácnou vyváženost mezi analytickým přístupem a syntetizující prezentací zjištěných poznatků, promyšlené formulace a tvůrčí aplikaci poznatků z (většinou zahraniční) literatury na zkoumané téma. Jedná se kromě jiného asi o první srovnávací analýzu webové použitelnosti těchto textových korpusů a digitálních knihoven. Oceňuji také, že student ve srovnání dokázal velmi vhodně aplikovat poznatky a taxonomie z úvodních teoretických kapitol.

Vzhledem k výše uvedenému bakalářskou práci doporučuji k obhajobě a navrhuji hodnotit ji známkou **v ý b o r n ě**.

V Praze dne 3. června 2021

Mgr. Jindřich Marek, Ph.D.

Ústav informačních studií a knihovnictví FF UK

Aspekt práce	Vysvětlení	Bodové hodnocení
metodologie a věcné zpracování tématu	<i>hodnotí se zvládnutí obsahové (věcné) stránky práce, splnění zadaného úkolu, využití výzkumných metod apod.</i>	38 / 40
přínos	<i>hodnotí se přínos a novost práce pro informační vědu v ČR</i>	17 / 20
citování, korektnost citování, využití inf. zdrojů	<i>hodnotí se zejména využití odkazů a pramenů v textu, korektnost citování; hodnotí se využití cizojazyčných pramenů</i>	20 / 20
slohové zpracování	<i>hodnotí se čtivost textu, skladba vět, odborný styl vyjadřování, rozsah práce</i>	15 / 15
gramatika textu	<i>hodnotí se gramatická úroveň práce, hrubky, čárky ve větách apod.</i>	5 / 5
CELKEM		95 / 100

Hodnocení bc. a dipl. prací

Bodový zisk za práci	Hodnocení
0-50 bodů	Neprospěl, nedoporučeno k obhajobě

51-60 bodů	Dobře (3)
61-80 bodů	Velmi dobře (2)
81-100 bodů	Výborně (1)