

**Diplomantka:** Bc. Pavla K u p c o v á

**Název diplomové práce:** Srovnávací analýza WebArchivu Národní knihovny ČR se zahraničními projekty

**Oponentka diplomové práce:** PhDr. Eva Bratková, Ph.D.

**Návrh klasifikace:** Dobře

V Praze, dne 10. 9. 2012

### Hodnocení

K obhajobě předkládaná diplomová práce Bc. Pavly Kupcové je věnována aktuálnímu a stále diskutovanému tématu univerzálně profilovaných archivů webu, které provozují velké národní knihovny, popř. jiné velké knihovny nebo instituce. Práce vznikla za příznivých okolností, autorka totiž pracuje v NK ČR v oddělení archivace webu (WebArchivu), jak uvádí v předmluvě. Diplomantka si vytkla za cíl komparaci tří vybraných reprezentantů těchto systémů (cíl je uveden i v zadání práce) s předpokladem jejich předchozí analýzy. Výsledný text práce (bez preliminárií a postliminárií) na s. 8-86 splňuje rozsahové požadavky kladené na diplomové práce (jeho součástí je ale také poměrně značný počet grafických informací, cca 22 normostran.).

Práce vychází jednak z publikované literatury a dalších informačních zdrojů, jednak z vlastní původní analýzy fungování daných systémů, převážně z pohledu uživatele. Diplomantka prezentuje bibliografickou přípravu na s. 87-95 (cca 75 jednotek), seznam referencí však zahrnuje v podstatě jenom domácí publikovanou literaturu k WebArchivu, nikoliv publikovanou literaturu zahraniční (k britskému a australskému systému), která také existuje (weby zahraničních systémů jsou zařazeny korektně). Poněkud překvapuje informace autorky na s. 8, že excerpovala jenom časopisy *Ikaros* a *Národní knihovna*. Otázkou je, proč neprovedla rešerši přímo v databázi KKL, kde jsou zahrnuty i další zdroje (jiné časopisy, sborníky, diplomové práce aj.). Není též jasné, které databáze centra EBSCO použila při rešerši, databáze LISA (ProQuest) není zmíněna vůbec. Řada důležité literatury unikla. K prezentaci bibliografických referencí mám velké námitky, vykytuje se v nich velká řada chyb (předpokládejme, že autorka užila standard ISO 690:2010). Chyby jsou vyznačeny v papírovém exempláři, který jí bude vrácen po obhajobě. Způsob citování pramenů v celé práci (asi harvardský systém?) je v zásadě dodržen, vyskytují se jenom drobné chyby.

Hlavní text diplomové práce je uspořádán do 6 základních kapitol, přičemž poslední je věnována závěru. Z hlediska proporcí je ale práce značně nevyvážená. Analytické části zahrnují celkem 67 s. (tedy téměř celý text diplomové práce), zatímco podstatná komparace je ani ne na 2 s. textu. Analytické části jsou také ne ve stejném poměru (český systém na 36 s., australský systém jenom na 13 s. a britský na 18 s.)

Kapitola první (4 s. textu) je úvodem k celé práci. Domnívám se, že její dělení na druhé úrovni je nadbytečné, neboť existuje jenom část 1.1, další ne. Text je vytvořen na základě literatury známých autorů, autorka jejich myšlenky a fakta řádně cituje. Možná často se uchyluje k citátům, které nejsou doprovázeny žádným vlastním komentářem nebo doplňkovým kontextem. Připomínku lze ale mít stran citace [CELBOVÁ, 2008a], která se vyskytuje poprvé na s. 11. Čtenář se ale do poslední chvíle nedozví, o jaké dílo vlastně jde, v referencích tento titul zcela chybí. Jde ale asi o důležitý zdroj, neboť citace se ještě vyskytuje na s. 14, 20, 25, 26 a jinde. U obhajoby je nutné referenci doplnit.

Nejrozsáhlejší kapitolou diplomové práce je, jak bylo výše vyčísleno, kapitola druhá, věnovaná prezentaci výsledků analýzy českého systému WebArchiv. Diplomantka na velmi

detailní úrovni tento systém popisuje, v prvních podkapitolách je kompletně využito také již publikovaných pramenů. Domnívám se, že text mohl být i jistým způsobem redukován ve prospěch posílení dalších kapitol. Přece jen jde v řadě případů o velmi známá, častokrát zveřejněná fakta. I v diplomové práci samotné se dvakrát opakují informace o typech a strategiích sklizní. Vítám ale naopak pasáže původní analýzy v částech věnovaných systémům – softwarům, vyhledávání informací a statistikám. Od s. 19 začíná autorka postupně zařazovat do textu grafické útvary (obrázky, tabulky, grafy aj.). Musím autorce ale připomenout, že v celé práci, až na malé výjimky, není na obrázky v textu vytvořen korektní odkaz (typu viz Obr. 1; nestačí uvést například frázi „jež jsou vidět na následující tabulce“, jak je uvedeno na s. 23). Tento nedostatek obecně pak čtenáři textu činí značné potíže v dalších kapitolách textu, kde je zařazování obrázků aj. velmi intenzivní. Dochází i k jistým dezorientacím. Větší dezorientace je na například na s. 39, kde autorka sděluje „V levém menu dole je možné ...“, čtenář ale neví, kam se má podívat. Nevhodné je také, aby v jedné podkapitole (2.9.3) byla obsahem jenom tabulka s čísly bez jediné komentující věty (Tabulka 5 má navíc centrované číslice v buňkách, jinde v tabulkách jsou zarovnané doprava či doleva).

Na s. 22 je odkaz na Přílohu 1, ta však není číslem 1 označena (drobná chyba). Faktická připomínka se týká sdělení, že každý zdroj navržený k archivaci je nejdříve vložený do databáze WA Admin (na s. 32) – skutečně se celé texty apod. někde takto ukládají? Anebo se autorka nepřesně vyjádřila? A mohla by autorka při obhajobě doplnit sdělení týkající se připojování adres archivačního uložení zdroje do katalogizačního záznamu ve formátu MARC 21? Jak je možné, že to nejde zjistit? Otázku mám k termínu „semínko ke zdroji“ (uvedeno na s. 35), co to znamená a proč není osvětlující poznámka pod čarou? K části 2.8.2 (Vyhledávání) mám jisté námitky vůči užívaným odborným termínům. Co znamená „trawler“ (na s. 41)? Na téže straně je sdělení „vyhledávání konkrétních názvů a klíčových slov...“ – to není korektní, přece uživatel nevyhledává název nebo klíčové slovo, nýbrž záznamy a dokumenty či zdroje podle názvu, klíčového slova apod. Tyto terminologické nedostatky se vyskytují i na jiných stranách a v dalších kapitolách. Konstatuji, že na s. 47 se autorce povedlo něco unikátního, zachytila (navždy) webovou stránku rozhraní pro vyhledávání nového WebArchivu s pětinasobnou gramatickou chybou (nekorektně zapsané slovo „tématický“).

Popis stejných parametrů australského systému Pandora v kapitole 3. je oproti českému systému hodně zestručněn, přičemž velkou část tvoří grafika (snímané obrazovky systému nebo schémata). Na s. 55 je zařazené ryze odborné schéma převzaté od cizího autora, ovšem diplomantka k němu nepodává příliš osvětlení. Čtenář tak stojí před těžkou situací, často nerozumí, co schéma v detailech znamená. Na téže straně je možná i chyba, autorka sděluje, že levá komponenta (možná pravá?) je Pandas Services Center (ve schématu je ale Pandas Services Server). Lze to při obhajobě osvětlit? Na s. 58 by obrazovka s uvedenými informacemi chtěla více textového objasnění.

Britský systém je popsán v kapitole 4. Text vykazuje taktéž podobné nedostatky, zdůrazním opět, selhává odkazovací systém na doprovodné grafické projevy a autorka užívá nekorektně některé odborné termíny.

Kapitola 5 (komparace systémů) měla být nosnou částí práce, avšak nestalo se tak. Autorka na necelých dvou stranách uvádí tabulku vybraných parametrů ke srovnání a pak krátký textový komentář. Otázkou je, proč tam nejsou parametry další (autorskoprávní, katalogizace aj.). Rozhodně si ale mohla dovolit poněkud delší textový komentář. Celá práce je ukončena sumarizačním závěrem v kapitole 6.

Diplomová práce má jisté potíže z hlediska formálního (grafika, odkazy, chybí seznam zkratk, ponechávání osamocených znaků na konci řádků aj.), stylistického i gramatického. Gramatických chyb se vyskytuje poměrně hodně, připojuji proto jejich seznam na konci hodnocení.

Závěrem konstatuji, že práce Bc. Pavly Kupcové splnila zadaný diplomový úkol. Práci doporučuji k obhajobě jako celek, ale s ohledem na řadu obsahových, formálních a jiných nedostatků ji navrhuji klasifikovat jenom jako **Dobrou**.

*Eva Bratková, v. r.*

*Gramatické chyby a překlepy v textu:*

- s. 8 odst. předposlední, druhá věta, chybí čárka v souvětí za slovem „sklizně“ (podřadicí spojka, vztah následkový)
- s. 10 odst. 2, první věta, chybí čárka v souvětí za slovem „počítače“
- s. 10 odst. poslední, jméno „Internet Archive“ je zapsáno dvakrát, ale v různých tvarech
- s. 13 odst. 2, třetí věta, chybí čárka v souvětí za slovem „debatou“ (podřadicí spojka, vztah následkový)
- s. 15 odst. 1, třetí věta, odborný termín „knihovnických sbírek“ je na s. 17 zapsán jinak („knihovních sbírek“); mělo by to být jednotné (viz TDKIV)
- s. 15 odst. 1, poslední věta, za slovem „i2010“ nebude čárka (není důvod)
- s. 15 odst. 2, záhlaví citace „KALAL“ je v rozporu se záhlavím reference (KÁLAL), co je správně?
- s. 17 odst. 1, první věta, za slovem „Brně“ nebude čárka (není důvod)
- s. 19 odst. 1, třetí věta, za slovem „archivů“ nebude čárka (není důvod)
- s. 19 odst. 1, třetí věta, jméno „Nizozemí“ není zapsáno korektně (správně Nizozemsko)
- s. 19 odst. 1, čtvrtá věta, chybí čárka v souvětí za slovem „multimédia“ (podřadicí spojka, vztah následkový)
- s. 20 odst. 3, poslední věta, chybí čárka v souvětí za slovem „internetu“ (podřadicí spojka, vztah následkový)
- s. 21 odst. první, první věta, neshoda podmětu s přísudkem (většina ... jsou).
- s. 25 odst. 3, předposlední věta, zápis procent ve smyslu podstatného jména (74 % ) je nutné podle pravidel psát s mezerou (v editoru WORD pak s pevnou mezerou, CTRL+Shift+Mezerník), bez mezery se procenta se zapisují pouze ve významu přídavného jména (např. dvacetiprocentní účast – 20% účast)
- s. 26 odst. předposlední, třetí věta, slovo „katalogizace“ asi není ve správném pádě (zaručuje ... katalogizaci?)
- s. 30 odst. 2, první věta, nesprávně zapsané slovo „emailu“ (správně e-mailu podle Slovníku cizích slov, Academia)
- s. 33 odst. 3, poslední věta, chybí čárka v souvětí za slovem „zjistit“
- s. 41 odst. 3, druhá věta, slovo „hledaná“ asi není ve shodě se zkratkou „URL“ (ženský, mužský rod nebo střední rod?)

- s. 41 odst. 4, výčtová položka poslední, chybí čárka v souvětí za slovem „ohledu“
- s. 47 Obrázek 15, pětinásobná gramatická chyba ve slově „Tématický“ (správně Tematický)
- s. 50 odst. 2, čtvrtá věta, chybí čárka v souvětí za slovem „konkrétně“ (podřadicí spojka, vztah následkový)
- s. 71 odst. 1, první věta, z slovem „WebArchiv“ by měla být čárka
- s. 71 odst. 4, první věta, sousloví „které je umístěno“ není ve správném tvaru (není ve shodě s předmětem věty hlavní – „předmětu“)
- s. 80 odst. 3, první věta, čárka v souvětí za slovem „stránek“ má být umístěna až za slovo „tak“ (před vedlejší větou)
- s. 83 Legenda Tabulky 9, slovo „archívů“ by mělo být s krátkým „i“ (sjednocení v celém textu)
- s. 83 odst. 1, třetí věta, chybí čárka v souvětí za slovem „generaci“