

**Posudek oponenta na disertační práci Ing. Jana Macha
„Správa, vyhledávání a zpřístupňování elektronických vysokoškolských
kvalifikačních prací“**

Předložená práce se zabývá velmi aktuální problematikou uchovávání, vyhledávání a zpřístupňování vysokoškolských kvalifikačních prací v elektronické podobě, a to velmi komplexně, počínaje přesným vymezením kategorie vysokoškolské kvalifikační práce a cílů uchovávání těchto prací, přes souhrnný popis různých výzkumně-vývojových projektů na toto téma, kterých se autor práce zúčastnil, srovnání různých známých softwarových nástrojů, až po vlastní návrh nástroje Validátor VŠE, který by měl (v součinnosti s jinými používanými nástroji) zejména přispět k efektivní antiplagiátorské kontrole předkládaných kvalifikačních prací.

Kapitola 1, „**Úvod**“ (s. 19–26), definuje předmět, cíle, metody a další parametry předkládané disertační práce a seznamuje s předchozími aktivitami autora v oblasti této práce. Kapitola 2, „**Výchozí stav**“ (s. 27–68), podává výše uvedené podrobnější definice základních pojmů, rozebírá základní právní podmínky uchovávání a zpřístupňování jak vlastních prací, tak údajů o nich, a shrnuje systémy a nástroje používané do r. 2014. Kapitola 3, „**Průzkum zpřístupňování vysokoškolských kvalifikačních prací v roce 2014**“ (s. 69–89), popisuje provedený průzkum a jeho výsledky.

Kapitola 4, „**Mapování metadat eVŠKP**“ (s. 90–106), rozebírá různé existující standardy pro popis eVŠKP (vysokoškolských kvalifikačních prací v elektronické formě) a možnosti převodu sledovaných metadat. Kapitola 5, „**Metriky pro měření užití eVŠKP v online prostředí**“ (s. 107–143), se zabývá možnostmi statistického vyhodnocování elektronických repozitářů a jejich vhodností pro eVŠKP. Kapitoly 6, „**Výběr systému centralizovaného vyhledávání**“ (s. 144–154), a 7, „**Vyhledávací rozhraní eVŠKP**“ (s. 155–166), v podstatě shrnují různé další specifické zkušenosti získané autorem při účasti na řešení vybraných dílčích úkolů v rámci tématu této práce.

Kapitola 8, „**Plagiátorství u eVŠKP**“ (s. 167–198), podává přehled dostupných antiplagiátorských systémů a specifikuje a zdůvodňuje způsob, jakým se autor rozhodl experimentálně vyhodnotit tyto systémy. Mimo jiné je tu stanoveno **7 „hypotéz o systémech na odhalování plagiátů“**.¹ Následují tabulky ukazující výsledky hodnocení a shrnutí tohoto hodnocení ve smyslu kategorizace systémů podle toho, jak splňují jednotlivé výše uvedené „hypotézy“ (testovací kritéria). Ze závěrů kapitoly 8 pak vychází projekt popsáný v následující kapitole 9, „**Validátor VŠE**“ (s. 199–217), jehož byl autor práce projektovým manažerem.

¹ Ačkoli vzhledem k formálnímu způsobu provádění hodnocení systémů asi nelze mít zásadní námitky vůči tomu, jak zde autor práce použil termín „hypotéza“, spíše bych doporučil hovořit v této souvislosti o „**dílčích testech**“ nebo o „**testovacích kritériích**“. Jde tu o to, že „nesplnění hypotézy“ u konkrétního systému snad neznamená, že by hypotéza neplatila, ale spíše to, že systém nesplnil určitý test!

Kapitola 10, „**Závěr a přínosy disertační práce**“ (s. 218–220), přehledně shrnuje, čeho bylo v rámci této práce dosaženo a jaké využití by to mohlo najít.

Celkově lze konstatovat, že práce velmi komplexně a přehledně řeší prakticky významnou úlohu, respektuje pravidla vědecké práce a přináší určité konkrétní praktické výsledky. Proto práci doporučuji k obhajobě.

V rámci obhajoby by se nicméně měl autor vyjádřit k následující **připomínce**:

Nejsm si jistý, *nakolik* v dnešních podmínkách *duplicitní výskyt jednotlivé věty* ve vyhodnocovaném a jiném odborném dokumentu svědčí skutečně o plagiátorství — dokonce i kdyby většina vět vyhodnocovaného dokumentu byla podobně duplicitní (ale s různými jinými dokumenty). Vzhledem k tomu, že se u kvalifikační práce požaduje, aby obsahovala smysluplná, pravdivá a přiměřeně srozumitelná tvrzení na nějaké předem dané téma, je počet skutečně přijatelných vět poměrně značně omezený ve srovnání s celkovými kombinatorickými možnostmi jazyka. A protože takové práce předkládá (v celosvětovém měřítku) mnoho miliónů studentů každý rok, obávám se, že požadovat něco takového, jako aby většina jednotlivých vět v těchto pracích byla skutečně originálních, tj. *dosud nikde nikým nenapsaných*, prostě přestává být reálné. (Většina studentů nejsou géniové, a tudíž neobjevují skutečně nové věci.) Obávám se, že pokud v blízké době budeme mít technickou možnost porovnat text předkládané kvalifikační práce se *všemi texty z podobných oborů, které už byly napsány*, nutně zjistíme, že většinu vět, ze kterých se skládá, už *někdo někde* napsal — aniž by tím pádem muselo jít o plagiátorství. **Zná autor zde předkládané práce nějaká statistická data, která by vyvrace-la tuto mou obavu?**

Pokud by výše naznačená obava byla oprávněná, znamenalo by to, že vysoká úhrnná podobnost posuzovaného textu s *velkým množstvím* jiných textů vlastně sama o sobě nic závažného neznamena. **Měl by autor pro takový případ nějaké doporučení, jak se vyhnout zbytečně mnoha planým poplachům a přiblížit se optimální automatizované detekci skutečného plagiátorství?**

V Praze 5.5.2015

doc. RNDr. Petr Strossa, CSc.