

ONLINE VYHLEDÁVÁNÍ DISERTACÍ

PODLE TÉMAT

Eva Bratková

brt@cuni.cz

Univerzita Karlova, Ústav informačních studií a knihovnictví, Česká republika

This paper is licensed under the Creative Commons licence: CC-BY-SA-4.0 (<http://creativecommons.org/licenses/by-sa/4.0/>).

Abstract

Příspěvek vyhodnocuje vyhledávání doktorských disertací podle témat v různých online systémech. Představena je situace v ČR, diskutován je problém úspěšnosti tematického vyhledávání disertací – lze najít vše relevantní nebo postačí najít jenom něco? Porovnána je situace u významných systémů v zahraničí, zejména pak v USA, kde byla již nastoupena také nová cesta zpřístupňování disertací formou propojených otevřených dat, ve které jsou řízené slovníky termínů, vyjadřujících téma, důležitou podmínkou. Jak se vyrovnají vybrané evropské systémy, jejichž disertace jsou již dnes reprezentovány v databázi katalogu WorldCat, s výzvou: „... Časem tyto odkazy [na entity témat] budou nahrazeny trvalými URI na ... zdroje propojených dat“?

Keywords

disertace, bibliografické záznamy, věcné pořádání informací, abstrakta, klíčová slova, řízené slovníky, vyhledávání informací, digitální repozitáře, propojená otevřená data

Introduction

Doktorské disertace patří k **nejvýznamnějším informačním výstupům** univerzit. Nesou cenné obsahy s výsledky vědeckých výzkumů. Disertace jsou pramenem, který by měl být, stejně jako vědecké knihy, kvalitně bibliograficky zpracován včetně témat. Jak často sdělují

i američtí knihovníci, stojí za to připravit co nejširší a nejlepší přístup k disertacím, a proto připravují jejich úplné katalogizační záznamy včetně věcných termínů s využitím řízených slovníků (předmětových heslářů, klasifikačních schémat aj.) (Middleton, 2015, s. 235; McCutcheon, 2008, s. 51). Kvalitně zpracované záznamy disertací lze pak dnes efektivně propojovat v rámci technologií sémantického webu ve formě otevřených propojených dat. Propojování záznamů disertací se **záznamy témat** je pro uživatele dnes podstatné.

Bibliografická registrace disertací na národní nebo mezinárodní úrovni má v řadě zemí dlouholetou tradici. Pro nové tisíciletí je navíc příznačné i jejich efektivní ukládání do digitálních repozitářů za účelem jejich pohodlného online vyhledávání a zejména zpřístupňování. Metadata, která disertace doprovázejí v lokálních digitálních systémech, hrají poté významnou roli také ve službách národní a mezinárodní úrovně, do kterých jsou bezproblémově sklížena. Online vyhledávání disertací podle formálních údajů (autora, názvu práce, jazyka, školy, přiděleného titulu aj.) bývá většinou bezproblémové a zpravidla i úspěšné. Problémem ale je **vyhledávání disertací podle témat**. Je-li vyhledávání závislé na indexu složeném pouze z termínů z názvů disertací, z autorských klíčových slov a abstraktu, popřípadě i z textů samotných, bývá **výtěžnost či úplnost** vyhledávání nízká (**přesnost** může být i vysoká), jak uváděly výsledky dřívějších známých výzkumů¹. Fondy disertací a záznamů o nich stále narůstají a vyhledávání podle témat se proto stává složitějším problémem, určitě pak z pohledu profesionálních řešitelů. Koncový uživatel zpravidla vždy "něco" najde a "něco" si z nabízeného seznamu dále vybere, aniž tuší, že vysoce relevantních disertací je v repozitáři mnohem více.

Tento příspěvek je zaměřen z výše uvedených důvodů výlučně na **disertace**, připravované rámci doktorských programů. Jiné typy závěrečných prací (bakalářské, magisterské aj.) nejsou předmětem rozboru a hodnocení, jakkoliv bývají, zejména v produkčně malých zemích, registrovány v lokálních i národních systémech.

Vyhledávání doktorských disertací podle témat v České republice

V ČR není dnes možné vyhledat záznamy všech obhájených disertací z jednoho místa, a to ani těch, jež byly obhájeny v posledních 10 letech (2007–2016), tedy období, kdy jsou na základě legislativy zpřístupňovány již všude veřejně. Na bázi existujících informačních služeb nelze ani zjistit, kolik bylo přesně obhájeno disertací v jednotlivých letech. Národní služby Theses.cz (<http://theses.cz/>) a NUŠL (Národní úložiště šedé literatury, <http://nysl.cz>) vykazují přibližně stejné roční přírůstky – okolo 1 000 obhájených disertací. Počet ale bude asi vyšší, neboť na národní úrovni nejsou registrovány všechny disertace, zejména z velkých univerzit. Obhájených disertací může být i přes 2 000 ročně.

¹ BLAIR, David C., MARON, M. E. An evaluation of retrieval effectiveness for a full-text document-retrieval system. In: Communications of the ACM. 1985, 28(3), 289-299. ISSN 0001-0782 (Print). ISSN 1557-7317 (Online). Dostupné komerčně z: <http://dx.doi.org/10.1145/3166.3197>. Ve formě zprávy dostupné volně z: <http://hdl.handle.net/2027.42/35415>.

Dlouholeté snahy² o vytvoření úplné národní registrace prozatím nedošly svého cíle. Největší počet záznamů disertací má ve své databázi služba NUŠL: v deklarované retrospektivě 1922–2016 měla v září 2016 celkem **18 304 záznamů**, za léta 2007–2016 jich má **7 496**. O něco menší počet záznamů disertací za období 2007–2016 má ve své databázi služba Theses.cz (**7 410** v září 2016). Celkově má tato služba **7 716 záznamů**. Poznamenat ale lze, že obě služby neregistrují totéž (viz též tabulka 1), z části jde o shodné a z části o rozdílné disertace. Řada disertací chybí, lze je vyhledat jenom v databázích univerzit. Je na jednotlivých univerzitách, do které služby záznamy zasílají nebo je nechají automaticky sklízet z repositářů. V ČR v současnosti neexistuje žádný legislativní předpis pro povinnou národní registraci. A dobrovolné nahlašování není pro ČR příznačné.

Vyhledávání disertací podle témat v systému Theses.cz

Národní registr závěrečných prací Theses.cz, spravovaný Masarykovou univerzitou, slouží, kromě jiného, také k vyhledávání záznamů těchto prací i prací samotných. Uživatelské rozhraní se po celá léta nijak zvlášť nezměnilo, je uživatelsky málo přívětivé jak na straně formulování dotazu, tak na straně vyhodnocování výsledných seznamů záznamů a prezentace úplných záznamů. Nabízí vyhledávání podle několika formálních údajů (roku obhajoby, jména školy aj.). Kupodivu chybí vyhledávání podle typu práce, náhradou je hledisko přidělených titulů (výsledky tak nemusejí být ale přesné).



The screenshot shows the Theses.cz search results page. At the top, it says 'Práce na stejné téma' and 'Theses'. Below that, it says 'Klíčová slova abecedně' and 'řadit abecedně | dle četnosti'. The main content is a large, dense cloud of keywords in various colors and sizes, including: accounting, adaptace, adolescence, advertising, aktivita, banka, bezpecnost, budget, business plan, cena, ceska republika, cestovni, ruch, communication, company, comparison, competition, competitiveness, costs, culture, customer, czech republic, dan, design, deti, development, diagnostika, distribuce, dite, domaci nasili, doprava, dotace, dotaznik, dotaznikove setreni, education, edukace, employee, environment, etika, european union, evaluation, evropska unie, family, financial analysis, financni analiza, financni plan, financovani, gender, historie, history, hodnoceni, hra, hudba, identita, informace, informacni system, inovace, integrace, integration, internet, investice, java, klient, komparace, kompetence, komunikace, komunikacni mix, konkurence, konkurenceschopnost, kriminalita, krize, kultura, kvalita, kvalita zivota, kvalitativni vyzkum, labour market, legislativa, lidske zdroje, likvidita, liquidity, logistika, management, marketing, marketing communication, marketing mix, marketing research, marketing strategy, marketingova komunikace, marketingova strategie, marketingovy mix, marketingovy vyzkum, materska skola, media, mentalni, postizeni, mentalni retardace, migrace, mladez, motivace, motivation, municipality, nabozenstvi, naklady, nezamestnanost, neziskova organizace, obec, obet, odmenovani, optimalizace, organizace, osetrovatelska pece, osobnost, pest analiza, php, planovani, podnik, podnikani, podnikatelsky plan, podpora, prodej, pohybova aktivita, pomerove ukazatele, predskolni vek, preklad, prevence, prevention, proces, profitability, project, projekt, propagace, public relations, quality, questionnaire, ramcovy vzdelavaci program, region, regionalni, rozvoj, reklama, rentabilita, risk, rizeni, rizeni lidskych zdrojů, riziko, rodina, rozpocet, rozvaha, rozvoj, senior, sikana, skola, sluzby, smrt, socializace, socialni prace, socialni pracovník, socialni site, socialni sluzby, spokojenost, spolecnost, spoluprace, sport, spotrebitel, stari, starnuti, strategic analysis, strategicka analiza, strategie, strategy, stres, strukturalni fondy, swot analysis, swot analiza, syndrom, vyhoreni, tourism, trh, trh prace, ucetnictvi, ucitel, unemployment, uver, verejna sprava, volnocasove aktivity, volny cas, vychova, vyvoj, vyziva, vyzkum, vzdelani, vzdelavani, zadluzenost, zak, zakaznik, zakladni skola, zamestnanci, zamestnanec, zamestnanost, zamestnavatel, zavislost, zdravi, zivotni prostredi, zivotni styl, znacka.

Výpis je omezen na štítky vložené u více než 500 prací.

Obrázek 1: "Cloud" abecedně řazených frekventovaných klíčových slov v systému Theses.cz

Vyhledávání disertací podle témat však přináší problémy. Systém nabízí v podstatě jedinou možnost: vyhledávání podle **klíčových slov** (jednoduchých termínů anebo frází). Uživatel musí dotaz zapsat do okna jednoduchého rozhraní. Rejstřík abecedně uspořádaných klíčových slov není k dispozici. Částečnou nabídkou je možnost prohlížení malé množiny

² Jde o snahy vyvíjené po roce 1990, které měly navázat na dřívější úplnou bibliografickou registraci disertací v ČR a zároveň na v SR: *Bibliografický katalog ČSSR. České knihy. Zvláštní sešit. Československé disertace ... 1964-1978*. Praha: Státní knihovna ČSSR, 1965-1981. ISSN 0323-1763. -- *Bibliografický katalog ČSSR. České knihy. Zvláštní sešit. České disertace ... 1979-1988*. Praha: Státní knihovna ČSSR, 1983-1990. ISSN 0232-041X.

vysoce frekventovaných klíčových slov vybraných z celého fondu (zahrnuty ale jsou všechny typy závěrečných prací). Množina je předložena formou efektního "cloudu" – viz obrázek 1, nelze ji však omezit jenom na disertace. Kliknutím na vybraný termín lze získat rozsáhlý seznam výsledných zkrácených záznamů všech typů závěrečných prací, ve kterém je dále složité vysoce relevantní disertace identifikovat a vybrat.

Běžní uživatelé si možná po letech již zvykli na situaci, kdy zapíší jednoduché klíčové slovo, na základě kterého systém předloží seznam vybraných záznamů, určitě však ne všechny obsahově relevantní. Ilustrovat lze příkladem tématu "digitální knihovny". Zápis dotazu pomocí různých synonymních nebo variantních termínů (něco takového koncový uživatel ale pravděpodobně běžně nedělá) vedl v září 2016 pokaždé k jiným výsledkům: „digitální knihovny“ (466 záznamů), „digitální knihovna“ (až 532), "elektronická knihovna" (184), "elektronické knihovny" (145), "digitální repozitář" (105), "digitální repozitáře" (20), "digitální archiv" (252), "digitální archivy" (27), "institucionální repozitář" (26) aj. Poskládat vše dohromady je pro uživatele složité, ruční zpracování dlouhých seznamů záznamů je obtížné, stejně jako vyhodnocování míry obsahové relevance u každého záznamu. Pokus vyhledat práce na téma samotného systému Theses.cz ("theses.cz") vedl v září 2016 k vyhledání až 993 záznamů, ovšem většina záznamů byla irelevantní.

Jako bonus nabízí systém Theses.cz vyhledávání prací na "příbuzné téma" ve formě odkazu ve vyhledaných zkrácených záznamech. Jde o počítačový kalkul založený na přítomnosti společných klíčových slov v záznamech prací. Získanou množinu záznamů příbuzných prací je ovšem také složité vyhodnocovat ručně co do obsahové relevance, uživatel se v seznamu spíše ztrácí, kliká-li vícekrát, ztratí se docela. Objevit práci vysoce relevantní je problém.

Systém Theses.cz, který je založen co do témat jenom na autorských klíčových slovech a abstraktech přicházejících z univerzitních repozitářů nebo studijních systémů, má prozatím problém s kvalitním a účinným vyhledáváním disertací podle témat. Propojování záznamů na předmětové termíny v nějakých řízených slovnících zatím není možné.

Vyhledávání disertací podle témat v systému NUŠL

Systém NUŠL je spravován Národní technickou knihovnou (NTK). Záznamy závěrečných prací všech typů z českých univerzit tvoří v systému NUŠL až 75 % (Charvátová, 2016, s. 91). Centrální uživatelské rozhraní (software FAST), vystavěné nad vlastním repozitářem NUŠL (software Invenio), je moderní a uživatelsky vřidné, má textové filtry pro pohodlnou navigaci. Pokud jde o disertace, nabízí i tato služba vyhledávání podle několika formálních údajů (autora, jména školy, jazyka, typu práce aj.). Limitování vyhledávání pomocí roku obhajoby je nastavitelné pomocí efektní časové osy. Kromě prohlížení seznamů údajů a jejich kombinování lze v systému dotaz také přímo zapisovat.

Vyhledávání disertací podle tématu je i v tomto případě založeno pouze na autorských klíčových slovech, která jsou součástí záznamů sklízených z univerzitních repozitářů nebo databází studijních systémů vybraných univerzit. Efektivnost vyhledávání i v případě této služby není optimální. Termíny, reprezentující téma, nejsou řízeny. Malý kontrolní test vyhledání disertací na téma "digitální knihovny" pomocí zápisu různých synonymních nebo variantních termínů vedl i v tomto případě pokaždé k jiným výsledkům (v září 2016): „digitální knihovny“ (346 záznamů), "elektronická knihovna" (350), "digitální repozitář" (61), "digitální repozitáře" (22), "digitální archiv" (47) a "institucionální repozitář".

Snahy o vylepšení stavu věcí v oblasti věcného popisu a online vyhledávání prací podle témat vedly pracoviště NUŠL k realizaci malého průzkumu stavu autorských klíčových slov s cílem zjistit, bylo-li by například možné jejich mapování na hesla Polytematického strukturovaného hesláře (PSH)³, který je v NTK jako řízený slovník spravován a užíván pro věcný popis. Navíc je již připraven ve formě propojených otevřených dat. Výsledky ukázaly, že prozatím by mapování bylo značně neefektivní (Charvátová, 2016, s. 91-95).

Pokud se nezmění či nezlepší popis disertací ve zdrojových digitálních systémech univerzit, ze kterých jsou metadata pro NUŠL sklízena, zásadně se nezmění ani efektivita vyhledávání podle témat. Vkrádá se otázka, jestli by bibliografická metadata nešlo sklízet raději z katalogizačních databází univerzit, neboť ty ve většině případů na profesionální úrovni popisu témat věnují lepší pozornost. Co tomu brání?

Popis témat disertací v lokálních systémech českých univerzit

Neuspokojivé výsledky vyhledávání českých disertací podle témat v obou národních službách jsou způsobeny tím, že zdrojové záznamy disertací pocházejí z digitálních repozitářů univerzit nebo z databází jejich studijních systémů s tím, že popis témat zajišťují sami autoři (klíčová slova a abstrakta). Znamená to, že i vyhledávání podle témat v lokálních repozitářích vykazuje stejné parametry, popsané výše u národních služeb. Repozitáře nabízejí uživatelům (v aplikaci DSpace či jiné) pod označením "Prohlížení dle předmětu" či jiným označením někdy i delší seznamy termínů označujících totéž – viz ukázka termínů k tématu "informační systémy" z jinak skvěle vypadajícího repozitáře Vysokého učení technického v Brně (VUT).



Obrázek 2: Rejstřík "předmětových hesel", aneb autorských klíčových slov k tématu "informační systémy" v repozitáři VUT (anglické ekvivalenty jsou v jiné části abecedy)

³ Polytematický strukturovaný heslář (PSH) [online]. Praha: Národní technická knihovna, 2000- [cit. 2016-09-26]. Dostupné z (URI): <https://psh.techlib.cz/skos/>.

Pro potřeby tohoto příspěvku byl proveden malý průzkum **popisu témat disertací** obhájených na vybraných českých univerzitách⁴, a to i mimo digitální repozitáře. Cílem bylo zjistit, kde jinde je popis témat realizován lépe, kde je vyhledatelnost disertací podle témat vyšší a kde je šance se v blízké budoucnosti dočkat i propojování záznamů o nich v rámci sémantického webu. Jak je patrné z tabulky 1, popis témat disertací je v **institucionálních repozitářích (IR)** zajišťován výlučně pomocí autorských abstraktů (AB) a klíčových slov (KS). Oba prvky nebo jeden z nich přitom nebývá v řadě záznamů přítomen vůbec (politika škol v tomto směru není jednotná). U dvou škol repozitáře buď neexistují, anebo nezahrnují disertace (Česká zemědělská univerzita a Univerzita Palackého), nebo jsou záznamy k dispozici jenom v databázích studijních systémů.

Zajímavý, byť různorodý přístup k autorským abstraktům a klíčovým slovům ale zaujímají **centralizované katalogy univerzitních knihoven (KK)**⁵. 6 ze 14 katalogů přebírá z repozitářů abstrakt, klíčová slova přebírá až 9 katalogů. Autorská klíčová slova jsou ale lokalizována nejednotně do polí formátu MARC 21, některé katalogy je lokalizují do pole 690, většina do pole 653 (neřízený indexový termín)⁶. Katalog Masarykovy univerzity využívá i lokálního pole M530. Klíčová slova bývají zobrazována v záznamech v rámci OPACů, nejsou však, z pochopitelných důvodů, zařazována do rejstříků předmětových termínů.

Co bylo potvrzeno průzkumem jako potěšitelné a nadějně, je skutečnost, že většina škol věnuje popisu témat v rámci katalogu velkou pozornost. Někdy je zpracování i velmi detailní. 8 ze 14, resp. 11 škol, využívá některý ze známých domácích **řízených předmětových heslářů či tezaurů** (viz tabulka 1). Jde především o Soubor věcných autorit Národní knihovny ČR (CZENAS), heslář (PSH) a také heslář CZMESH (pro biomedicínu). Využívají se i jiné systémy (Český teologický tezaurus, CTT) a také lokální předmětové hesláře (CZ-BrMU, PHFFUK, MFF aj.). V případě PSH se naskýtá možnost okamžitého propojování katalogizačních záznamů na heslář PSH, který už je vystavený ve strukturované formě (hesla mají identifikátory URI), u hesláře CZENAS se pravděpodobně objeví možnost jeho vystavení ve formě propojených otevřených dat v blízké budoucnosti, jak lze předpokládat i ze závěrů výzkumného projektu INTERPI⁷. Až 9 katalogů využívá k popisu témat disertací **klasifikační schéma MDT** (Mezinárodní desetinné třídění). V tomto případě existuje jistá šance otevřeného propojování záznamů na záznamy všech třídění MDT. Prozatím je k dispozici soubor vybraných hlavních třídění ve struktuře propojených otevřených dat (<http://udcdata.info/>). Například třída lékařských věd (Medical science) má URI: <http://udcdata.info/037318>. Přístup k celému souboru MDT (<http://cz.udc-hub.com/cs/login.php>) je v českém jazyce dostupný na základě registrace. V budoucnu, kdy objemy záznamů rychle porostou, bude uspořádanost disertací podle tohoto univerzálního třídění rozhodně ku prospěchu.

⁴ Jde o 14 univerzit, které vykazují nejvíce obhájených disertací v posledních letech. Počty disertací byly vzaty z obou národních služeb a také z databází lokálních systémů včetně centrálních katalogů.

⁵ Ze 14 škol 3 nekatalogizují disertace vůbec, ojedinělé záznamy se ale občas objeví (viz tabulka 1).

⁶ Zjištěna byla v jednom případě i zvláštnost, autorská klíčová slova v katalogizačních záznamech Vysoké školy ekonomické jsou lokalizována v poli 690, v databázi NUŠL, která záznamy přebírá, jsou však posléze nabízena pro potřeby exportu v poli 653.

⁷ Sídlo projektu INTERPI je na URL: <http://authority.nkp.cz/interpi>.

	IR		KATALOG KNIHOVEN					THESES		NUŠL		BASE			OpenAIRE		DART-EU	
	AB	KS	AB	KS	PH	MDT	konspekt	AB	KS	AB	KS	AB	KS	DDT	AB	KS	AB	KS
ČZU			—	—	CZENAS	MDT	konspekt											
ČVUT	AB	KS	—	—	PSH	MDT	konspekt			AB	KS							
JU			—	KS	CZENAS	MDT	konspekt	AB	KS	AB	KS				AB	KS		
MU Brno	AB	KS	AB	KS	CZENAS CZ-BrMU CZMESH AJ.	MDT	konspekt	AB	KS									
MENDELU	AB	KS	AB	KS	—	—	—	AB	KS	AB	KS				AB	KS		
TU Liberec	AB	KS	AB	KS	CZENAS	—	konspekt	AB	KS	AB	KS							
UK Praha	AB	KS	AB	—	CZENAS PHFFUK CTT, MFF CZMESH AJ.	MDT	konspekt											
UPOL			AB	KS	—	MDT	—	AB	KS	AB	KS							
UPA	AB	KS								AB	KS	AB	KS	DDT				
ÚTB Zlín	AB	KS	—	KS	CZENAS ECZENAS AJ.	MDT	konspekt	AB	KS			AB	KS	—				
VŠB-TUO	AB	KS						AB	KS	AB	KS	AB	KS	DDT				
VŠE Praha	AB	KS	AB	KS	CZENAS	MDT	—	AB	KS	AB	KS	AB	KS	DDT	AB	KS	AB	KS
VUT Brno	AB	KS	—	KS	PSH	MDT	—			AB	KS	AB	KS	DDT	AB	KS		
ZČU	AB	KS						AB	KS	AB	KS	AB	KS	—				

Tabulka 1: Přehled údajů věcného popisu disertací z vybraných univerzit ČR v různých informačních systémech a vybraných národních nebo mezinárodních službách [Legenda: IR = institucionální repozitář, AB = abstrakt, KS = klíčové slovo, PH = předmětový heslář]

U 7 katalogů je uplatňována i známá kategorizace KONSPEKT.

Je potěšující, že české disertace mají možnost být co do témat takto lépe nalezitelné, byť jenom prostřednictvím OPACů. Bohužel, zatím jenom v každém katalogu zvlášť. Záznamy českých disertací v současnosti nepostupují, jak tomu je v jiných vyspělých zemích, do **národního souborného katalogu (CASLIN)**. A to je velká škoda, mohli bychom mít přístup k celému fondu záznamů disertací z jednoho místa. Jiným námětem může být myšlenka, že by se záznamy v repozitářích univerzit obohacovaly o určité údaje z katalogů. Myšlenka je logická, ale praktické řešení není možné v tuto chvíli očekávat.

Disertace a databáze České národní bibliografie

V ČR již byla dříve oficiálně vedena celostátní registrace všech obhájených disertací. Tradiční systém zanikl po zrušení příslušné legislativy (po roce 1990). Je škoda, že tato registrace nemohla pokračovat, podobné systémy v jiných zemích úspěšně fungují.

V rámci průzkumu zpracování témat disertací v katalogích byla kontrolována i databáze České národní bibliografie (ČNB, <http://aleph.nkp.cz/F/>). Je skutečností, že ČNB registruje dnes řadu tezí (autoreferátů) disertací, neboť jsou na některých univerzitách⁸ oficiálně publikovány (s identifikátory ISBN). Systém ČNB je dostává jako povinné výtisky. Podléhají následně kvalitnímu zpracování v databázi ČNB včetně detailního předmětového a systematického popisu⁹. Například záznam tezí disertace autora Petra Maršálka s názvem "Únavové zkoušky ozubených kol" (ISBN 978-80-248-2991-3) získal až 4 třídníky schématu MDT a kromě

⁸ Jde například o Univerzitu Tomáše Bati ve Zlíně, Vysoké učení technické v Brně, VŠB - Technickou univerzitu Ostrava aj.

⁹ Katalogizace publikovaných tezí je realizována i na úrovni univerzitních katalogů.

formálních předmětových hesel také dvě tematická hesla z řízeného slovníku CZENAS¹⁰. Díky spolupráci NK ČR a OCLC je záznam již také v databázi WorldCat (<http://www.worldcat.org/oclc/855464854>) a v současné chvíli je dostupný i na webu ve struktuře propojených otevřených dat (ve standardech schema.org a RDFa). Záznam tezí citované disertace jako díla je k dispozici volně, má identifikátor URI: <http://worldcat.org/entity/work/id/1374607131>. Záznam díla bude propojen na záznamy českých řízených předmětových hesel včetně anglických ekvivalentů (viz blok "about" na obr. 3), jakmile je NK ČR vystaví jako propojená data s příslušnými identifikátory URI. Prozatím se při pokusu na ně klikat objeví anglická zpráva: „This is a placeholder reference for a Topic entity, related to a WorldCat Entity. Over time, these references will be replaced with persistent URIs to... and other Linked Data resources“.



The screenshot shows a WorldCat record for a thesis. The title is "Únavové zkoušky ozubených kol = Gear fatigue tests : autoreferát disertační práce". The URI is <http://worldcat.org/entity/work/id/1374607131>. The record includes a "type" section with the URI <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> and a specific type <http://schema.org/CreativeWork>. The "about" section lists several topic-related URIs: <http://schema.org/about>, http://experiment.worldcat.org/entity/work/data/1374607131#Topic/zkousky_na_unavu_materialu, http://experiment.worldcat.org/entity/work/data/1374607131#Topic/fatigue_tests, http://experiment.worldcat.org/entity/work/data/1374607131#Topic/ozubena_kola, and http://experiment.worldcat.org/entity/work/data/1374607131#Topic/cog_wheels.

Obrázek 3: Část záznamu tezí české disertace jako díla z experimentálního serveru propojených otevřených dat OCLC

Lze v tuto chvíli jen litovat, že české disertace nejsou systémem národní bibliografie zpracovávány. Jejich kvalitní záznamy by tak mohly být součástí významných světových systémů včetně systému Google Books, který přebírá záznamy z katalogu WorldCat. Znásobila by se jejich vyhledatelnost, též z mobilů koncových uživatelů.

Vyhledávání doktorských disertací podle témat v zahraničí

Ve srovnání se situací v ČR v oblasti zpracování a vyhledávání disertací podle témat je situace v zahraničí podstatně lepší. V řadě zemí jsou samozřejmostí provozy různorodých systémů národní nebo i nadnárodní úrovně pro jejich zpracování a online vyhledávání z jednoho místa včetně zpřístupňování jejich textů. Zpracování témat podle řízených slovníků typu

¹⁰ Kompletní záznam tezí dané disertace je v databázi ČNB dostupný na URL: http://aleph.nkp.cz/F/?func=direct&doc_number=002466242&local_base=CNB.

předmětových heslářů či klasifikačních schémat, jež jsou v řadě případů již vystaveny zdarma ve formě propojených otevřených dat, je samozřejmostí. Jejich vyhledatelnost je nepochybně veliká, též v rámci vyhledávače Google. Dále v textu bude krátce charakterizována jako příklad situace v USA a Německu.

USA

USA mají dnes dva systémy, jejichž databáze mají podobný počet záznamů disertací (Procious, 2014, s. 144). Registrace doktorských disertací je stoprocentní.

První je databáze **PQDT (ProQuest Dissertation & Theses)** společnosti ProQuest. Nachází se v ní přibližně 3 500 000 záznamů disertací a též vybraných magisterských prací. Registrace doktorských disertací z USA je úplná (ročně je obhájeno okolo 40 000 doktorských disertací). Zapojeno je asi 700 univerzit. Bibliografické záznamy databáze PQDT jsou velmi podrobné co do formálního popisu, přidělován je také identifikátor ISBN. Online vyhledávání disertací zajišťuje společnost ProQuest v rešeršním systému FAST (nový je připravován). Co do témat, záznam je opatřen autorským abstraktem, profesionálové ProQuest disertace zařídí do vlastního univerzálního kategorizačního schématu (ProQuest Subject Categories). Slovní ekvivalenty kategorizace jsou navíc lokalizovány i v poli předmětových hesel. Záznam obsahuje zpravidla také autorská klíčová slova. Díky spolupráci s univerzitami lze záznamy z databáze PQDT převzít (přebírány jsou vybrané užitečné údaje) do univerzitních katalogů, kde podlehnou kvalitnímu dopracování, zejména pokud jde o téma disertace. Následně se záznamy předávají do databáze WorldCat (OCLC). Vyhledatelnost disertací podle tématu se dále zvyšuje v některých oborových databázích (PsycINFO, ERIC aj.), do kterých společnost ProQuest předává dle dohody jejich záznamy. Záznamy jsou v nich dále doplňovány o jemnější předmětové termíny z tezurů a třídíky z oborových klasifikací.

Druhou databází, která obsahuje velký počet záznamů disertací z USA, je databáze **WorldCat**. Záznamy pocházejí z univerzitních katalogů, jsou opatřeny hesly řízeného předmětového hesláře LCSH a automaticky také hesly nového hesláře FAST (OCLC) a také třídíky klasifikace LCC nebo DDC. Díky technologické vyspělosti OCLC jsou záznamy těchto disertací dnes vystaveny ve formě propojených otevřených dat s tím, že propojení na předmětové hesláře je fakticky živé. Například disertace práce autora N. R. Kala s názvem "A case study on robustness of Dynamic Time Warping for activity recognition using wearable computers" (URI: <http://www.worldcat.org/oclc/846506856>), obhájená na Texaské univerzitě v Dallasu v roce 2012, má vedle abstraktu až 8 předmětových hesel propojených ze záznamu díla (URI dané disertace jako díla je <http://worldcat.org/entity/work/id/1356254544>) na záznamy relevantních heslářů dostupných na webu: LCSH (například heslo "Dynamic programming", URI: <http://id.loc.gov/authorities/subjects/sh85040313>) a FAST (například heslo "Wireless sensor nodes", URI: <http://id.worldcat.org/fast/1750044>).

Vyhledatelnost disertací z USA podle témat je tak významně zajištěna, a to i v podmínkách rozrůstajícího se sémantického webu.

SRN

Podobně příznivá je vyhledatelnost podle témat u všech disertací obhájených na německých univerzitách. Díky zákonu o povinném odevzdávání a uchovávání disertací se daří realizovat jejich úplnou registraci v rámci systému Německé národní bibliografie. Záznamy jsou dostupné

z katalogu Německé národní knihovny, tedy také z jednoho místa. Zveřejňování záznamů vyniká precizitou, veškeré fyzické záznamy podstatných "věcí" (objektů) již mají jedinečné identifikátory URI včetně záznamů předmětových hesel a jmenných autoritních dat. Téma disertace je reprezentováno především pomocí řízeného předmětového hesláře SWD (Schlagwortnormdatei), který je dnes součástí autoritního souboru GND (Gemeinsame Normdatei), vystaveného již také ve struktuře propojených otevřených dat. Příkladem může být disertace autora Floriana Kleina s názvem Metadaten-Verwaltung in einem verteilten RAM-basierten Speicherdienst, obhájená na Düsseldorfské univerzitě v roce 2015 (<http://d-nb.info/1079652574>). Disertace má přidělen identifikátor URN (přístup k textu je zajištěn ze dvou míst), záznam je primárně propojen na národní autoritní záznam autora (<http://d-nb.info/gnd/187515727>), záznamy obou oponentů (<http://d-nb.info/gnd/1018114750> a <http://d-nb.info/gnd/122535316>) a celkem 6 záznamů řízených předmětových hesel: Speicher <Informatik> (<http://d-nb.info/gnd/4077653-0>), Metadaten (<http://d-nb.info/gnd/4410512-5>), Verwaltung (<http://d-nb.info/gnd/4063317-2>), Dienst <Informatik> (<http://d-nb.info/gnd/4835035-7>), Speicherverwaltung (<http://d-nb.info/gnd/4182146-4>), Anwendungssoftware (<http://d-nb.info/gnd/4120906>). K dispozici je navíc i třídnic klasifikačního schématu Německé národní knihovny. Viz obrázek 4.

Login →	
→ Über die Deutsche Nationalbibliothek	
Link zu diesem Datensatz	http://d-nb.info/1079652574
Titel	Metadaten-Verwaltung in einem verteilten RAM-basierten Speicherdienst / Florian Klein. Gutachter: Michael Schöttner ; Martin Mauve
Person(en)	Klein, Florian Schöttner, Michael (Gutachter) Mauve, Martin (Gutachter)
Verlag	Düsseldorf : Universitäts- und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf
Zeitliche Einordnung	Erscheinungsdatum: 2015
Umfang/Format	Online-Ressource
Hochschulschrift	Düsseldorf, Heinrich-Heine-Universität, Diss., 2015
Persistent Identifier	URN: urn:nbn:de:hbz:061-20151202-144245-2
URL	http://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=36464 (Verlag) (kostenfrei zugänglich)
Sprache(n)	Deutsch (ger)
Anmerkungen	Langzeitarchivierung gewährleistet
Schlagwörter	Speicher <Informatik>* ; Metadaten* ; Verwaltung* ; Dienst <Informatik>* ; Speicherverwaltung* ; Anwendungssoftware* (*maschinell ermittelt)
Sachgruppe(n)	004 Informatik
Online-Zugriff	Archivobjekt öffnen

- In meine Auswahl übernehmen
- Druckansicht
- Versenden
- MARC21-XML-Repräsentation dieses Datensatzes
- RDF (Turtle)-Repräsentation dieses Datensatzes
- Dokumentation Linked Data
- BIBFRAME-Repräsentation dieses Datensatzes
- Projekt BIBFRAME
- Korrekturanfrage

Teilen Tweet

[i](#) [⚙](#)

Obrázek 4: záznam německé disertace v katalogu Německé národní knihovny s identifikátorem URN a také s hyperlinky na záznamy předmětových hesel řízeného hesláře GND (Gemeinsame Normdatei)

Díky spolupráci s OCLC jsou také záznamy německých disertací dostupné v databázi WorldCat, a jsou tudíž dostupné také ve formě otevřených propojených dat. Klikání na německá národní předmětová hesla prozatím hlásí stále zprávu, jež byla ocitována výše v textu u příkladu popisu tezí jedné české disertace. Zpráva už vlastně ale není zcela pravdivá, německá předmětová hesla již URI mají, v rámci experimentu OCLC budou, jak lze předpokládat, brzy korektně propojena.

Conclusion

Digitální repozitáře českých univerzit i dvě národní služby, které registrují a zpřístupňují disertační práce, se zdárně rozvíjejí a rozrůstají co do objemu dat. Záznamy disertací z některých univerzit již postupují také do mezinárodních služeb (BASE, OpenAIRE a DART-Europe - viz tabulka 1). Něco podstatného jim však chybí, a to lepší a kvalitnější přístup k vyhledávání disertací podle témat. S narůstajícími objemy záznamů a disertací samotných hrozí do budoucna vážné ztráty při jejich kvalitním a efektivním vyhledáváním. Najít jenom "něco" z určitého množství obsahově relevantních disertací možná v daný okamžik uspokojí laika (studenta apod.), neuspokojí už ale profesionálního rešeršéra, jehož zadaným úkolem může být vyhledání všech disertací na určité téma. Malý průzkum v ČR ukázal, že určitá naděje lepšího popisu témat existuje, a to v rámci katalogů univerzitních knihoven. Řada knihoven (nikoliv všechny) přiděluje předmětová hesla a třídíky klasifikace MDT a kategorie systému Konspekt. Existuje šance, že na webu budou záznamy disertací moci být v budoucnu propojeny na záznamy těchto heslářů a klasifikací. Vystavení některých pořadacích systémů ve struktuře propojených otevřených dat je již realitou, jiné se připravují. Jde o mocnou technologii, která by možná pomohla změnit nepříznivou českou situaci v oblasti popisu témat disertací. Lze popřípadě doufat, že by se přece jenom daly nějakým způsobem integrovat popisy témat disertací, uvedené k repozitářích a katalogích zvlášť? Lze doufat, že by do národního souborného katalogu CASLIN mohly postupovat i záznamy českých disertací? To jsou již otázky na další diskusi.

References

- CHARVÁTOVÁ, Michaela, 2016. Sjednocování věcného popisu agregovaných záznamů v repozitáři NUŠL. Praha. 121 s., 2 příl. Diplomová práce (Mgr.). Univerzita Karlova, Filozofická fakulta. Dostupné také z: <https://is.cuni.cz/webapps/zzp/detail/161853/>.
- McCUTCHEON, Sevim aj., 2008. Morphing Metadata: Maximizing Access to Electronic Theses and Dissertations. *Library Hi Tech*. **26**(1), 41-57. Dostupné také komerčně z (DOI): <http://dx.doi.org/10.1108/07378830810857799>.
- MIDDLETON, Cedar C. aj., 2015. A Process for the Original Cataloging of Theses and Dissertations. *Cataloging & Classification Quarterly*. **53**(2), 234-246. Dostupné komerčně také z: <http://dx.doi.org/10.1080/01639374.2014.971997>.
- PROCIOS, Aaron W., 2014. WorldCat, the Other ETD Database: An Exploratory Study. *The Reference Librarian*. **55**(2), 144-150. Dostupný komerčně také z (DOI): <http://dx.doi.org/10.1080/02763877.2014.880276>.