

Univerzita Karlova v Praze

Filozofická fakulta

Ústav informačních studií a knihovnictví

Studijní program: informační studia a knihovnictví

Studijní obor: informační studia a knihovnictví

Barbora Ševčíková

Věcné pořádkání šedé literatury
(Subject control of grey literature)

Diplomová práce

Praha 2007

Autor práce: Barbora Ševčíková

Vedoucí práce: Mgr. Josef Schwarz

Oponent práce:

Datum obhajoby:

Hodnocení:

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracovala samostatně a že jsem uvedla všechny použité prameny a literaturu.

V Praze dne 11. 8. 2007

Barbora Ševčíková

Identifikační záznam

ŠEVČÍKOVÁ, Barbora. *Věcné pořádní šedé literatury (Subject control of grey literature)*. Praha, 2007. ii, 143 s., 37 s. příl. + 1 CDROM. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce Mgr. Josef Schwarz.

Abstrakt

Tématem diplomové práce je představení a zhodnocení systémů věcného pořádní v oblasti šedé literatury se zaměřením na Systém pro informace o šedé literatuře v Evropě (System for Information on Grey Literature in Europe, SIGLE) a fond šedé literatury ve Státní technické knihovně, která byla hlavním partnerem pro SIGLE v ČR. V rámci srovnání je dále popsána funkce a věcné pořádní v americkém systému Síťové digitální knihovny diplomových a disertačních prací (Networked Digital Library of Theses and Dissertations, NDLTD) a australském systému Australian Digital Theses (ADT).

V práci je nejprve charakterizována šedá literatura, definice, dosavadní vývoj a trendy. Následně je popsán elektronický formát diplomové a disertační práce, jejího vznik a vývoj. Jádro práce bude tvořit samotný popis a srovnání systémů věcného pořádní informací SIGLE, NDLTD, ADT a Polytematického strukturovaného hesláře (PSH). V části věnované praktickému využití předmětových pořádní systémů je zachycena dosavadní praxe v užívání selekčních jazyků všeobecně a konkrétně v oblasti šedé literatury se specializací na vysokoškolské kvalifikační práce.

V závěru jsou shrnuty výsledky srovnání, možnosti využití pořádní systémů v databázi šedé literatury, která vzniká ve Státní technické knihovně v rámci projektu Digitální knihovna pro šedou literaturu.

.

Klíčová slova

Předmětové heslo, klíčové slovo, věcné pořádní informací, indexace, věcný popis, šedá literatura, diplomová práce, disertace, elektronické diplomové práce a disertace

OBSAH

Předmluva	i
1. Úvod	1
2. Šedá literatura	2
2.1. <i>Charakteristika, dosavadní vývoj a trendy</i>	2
2.2. <i>Co je to šedá literatura?</i>	3
2.2.1. Definice a vlastnosti šedé literatury	3
2.2.2. Krize šedé literatury	4
2.2.3. Propagace	4
2.2.4. Prosazení šedé literatury v digitální době	5
2.2.4.1. Proč nepublikovat	5
2.2.4.2. Proč publikovat	5
2.2.4.3. Producenti	6
2.2.4.4. EAGLE-SIGLE	6
3. Elektronické diplomové a disertační práce	7
3.1. <i>Rozdělení</i>	7
3.2. <i>Zpřístupnění - dostupnost</i>	8
3.2.1. Podpora informační gramotnosti studentů	8
3.2.2. Browsing	9
3.2.3. Systémy pro věcné pořádkání informací	9
3.3. <i>Rozvoj a propagace</i>	10
3.4. <i>Úložiště pro elektronické diplomové a disertační práce</i>	10
3.4.1. Rizika ukládání	12
3.5. <i>Závěr</i>	13
4. Charakteristika vybraných systémů	13
4.1. <i>Systém informací o šedé literatuře v Evropě (SIGLE)</i>	13
4.1.1. Úvod	13
4.1.1.1. Historie	14
4.1.1.2. Evropská asociace pro využívání šedé literatury (EAGLE)	14
4.1.2. Členové	14
4.1.3. Analýza SIGLE	15
4.1.3.1. Slabiny systému SIGLE	16
4.1.3.2. Problematika tvorby jednotné báze šedé literatury	16
4.1.3.3. Východisko z krize	17
4.1.4. SIGLE v ČR	17
4.1.4.1. Úvod	17
4.1.4.2. Nekompatibilita formátů	17
4.1.4.2.1. Možnosti řešení	18
4.1.4.3. Státní technická knihovna koordinátorem pro SIGLE	18
4.1.4.4. Nové aktivity ve Státní technické knihovně	18
4.1.5. Technické údaje o SIGLE	19
4.1.5.1. Typy registrovaných dokumentů	20
4.1.5.2. Geografický záběr	20
4.1.5.3. Obsahový záběr	21
4.1.6. Bibliografický záznam	22
4.1.6.1. Pravidla	22
4.1.6.2. Identifikační popis	23
4.1.6.3. Věcný popis	23
4.1.6.3.1. Systém pro věcné pořádkání informací SIGLE	23
4.1.6.3.2. Struktura věcného pořádkacího systému SIGLE	24

4.1.6.3.3.	Konečné uplatnění věcného pořádacího systému SIGLE.....	25
4.1.6.3.4.	Příklady záznamů ze SIGLE.....	25
4.1.7.	Nárůst záznamů v jednotlivých databázích šedé literatury	28
4.1.8.	Závěr.....	28
4.2.	<i>Síťová knihovna diplomových a disertačních prací (NDLTD)</i>	29
4.2.1.	Úvod	29
4.2.2.	Technické zázemí	30
4.2.3.	Členové a rozvoj fondu elektronických diplomových a disertačních prací.....	30
4.2.4.	Standards popisu	32
4.2.4.1.	Úvod	32
4.2.4.2.	Metadata	32
4.2.4.3.	Jmenné autority	32
4.2.5.	Věcné pořádní	33
4.2.5.1.	Příklady záznamů a popis použitých předmětových hesel	33
4.2.5.2.	Vyhledávání pomocí MARIAN.....	36
4.2.5.3.	Multijazyčné vyhledávání	37
4.2.6.	Budoucnost NDLTD	38
4.2.6.1.	Rozvoj projektů	38
4.2.7.	Závěr.....	39
4.3.	<i>Program australských digitálních disertací (ADT)</i>	40
4.3.1.	Úvod	40
4.3.2.	Členové a rozvoj ADT.....	41
4.3.3.	Analýza systému.....	42
4.3.4.	Technické zázemí	43
4.3.5.	Standards popisu	44
4.3.5.1.	Úvod	44
4.3.5.2.	Metadata	44
4.3.5.3.	Jmenný popis	45
4.3.5.4.	Věcný popis	46
4.3.5.5.	Předmětová brána	47
4.3.6.	Budoucnost ADT.....	48
4.3.6.1.	Rozvoj projektů	49
4.3.7.	Závěr.....	50
4.4.	<i>Selekční jazyky v STK</i>	50
4.4.1.	Historický vývoj	50
4.4.2.	Polytematický strukturovaný heslář (PSH)	51
4.4.3.	Termíny	52
4.4.4.	Struktura	54
4.4.5.	Indexace dokumentů hesly PSH.....	55
4.4.5.1.	Výběr konkrétních hesel.....	55
4.4.5.2.	Určení významnosti hesel.....	56
4.4.5.3.	Počet hesel	56
4.4.5.4.	Prekoordinace a postkoordinace.....	57
4.4.5.5.	Aktualizace PSH.....	57
4.4.6.	Rozvoj.....	57
4.4.7.	Oblasti rozvoje.....	58
4.4.7.1.	Spolupráce s dalšími subjekty	58
4.4.7.2.	Zpřístupnění hesláře	58
4.4.8.	Oblasti útlumu	59
4.4.9.	Závěr.....	59
5.	Věcné třídění šedé literatury	59
5.1.	<i>Definice</i>	59

5.2.	<i>Selekční jazyk</i>	60
5.3.	<i>Popis tvorby pořádacích znaků</i>	61
5.3.1.	Moderní principy	62
5.3.2.	Automatická indexace	62
5.4.	<i>Rozdělení pořádacích systémů</i>	63
5.4.1.	Jazyky deskriptorového typu (tezaury).....	63
5.4.2.	Volně tvořené předmětové termíny	64
5.4.3.	Klíčová slova	65
5.5.	<i>Harmonizace selekčních jazyků</i>	66
5.6.	<i>Zhodnocení kvality selekčního jazyka</i>	67
5.7.	<i>Věcný popis elektronických dokumentů</i>	68
5.7.1.	Problematika indexace.....	68
5.7.2.	Možnosti přístupu k indexaci elektronických dokumentů.....	68
5.7.3.	Věcný popis elektronických diplomových a disertačních prací	69
5.7.3.1.	Metody zpřístupnění bibliografického záznamu	69
5.7.3.2.	Aplikace metod.....	70
5.7.3.3.	Nároky na lidské zdroje vs. užitečnost předmětových hesel	70
5.7.3.3.1.	Výhody přidávání předmětových hesel do záznamů.....	70
5.7.3.3.2.	Nevýhody přidávání předmětových hesel do záznamů	72
5.7.3.3.3.	Alternativní řešení	73
5.7.3.4.	Závěr.....	73
5.7.4.	Standardy	74
5.7.5.	Využití PSH k indexaci elektronických diplomových a disertačních prací	74
5.7.5.1.	Srovnání předmětového pořádání SIGLE vs. heslář PSH.....	74
5.8.	<i>Vyhledávání v elektronických disertacích a diplomových pracích</i>	75
5.8.1.	Úvod	75
5.8.2.	Uživatelé a jejich schopnosti	75
5.8.3.	Dotazování vs. prohlížení	76
5.8.3.1.	Prohlížení ve Státní technické knihovně	77
5.8.4.	Vyhledávání podle autora, názvu, předmětového hesla	78
5.8.5.	Fulltextové vyhledávání	79
5.8.6.	Závěr.....	80
5.9.	<i>Budoucnost pořádacích systémů</i>	81
6.	Závěr	82
	Seznam grafů	84
	Seznam tabulek	84
	Seznam obrázků	84
	Seznam použité literatury	85
	Přílohy	96

Předmluva

Předkládaná diplomová práce je zaměřena na předmětové pořádací systémy, které již byly použity pro pořádání v databázích elektronických i tištěných diplomových a disertačních prací. K výběru tohoto tématu mě instirovala jednak samotná práce na správě Polytematického oborového hesláře, který byl vytvořen a je dále rozvíjen ve Státní technické knihovně, a která přispěla k zohlednění praktických aspektů v této práci, jednak spolupráce s pracovním týmem projektu Digitální knihovna pro šedou literaturu rovněž ve Státní technické knihovně. Nezanedbatelnou roli při výběru tématu hrála skutečnost, že zatím nebylo v české odborné literatuře z oblasti knihovní a informační vědy konkrétněji zachyceno. Bohužel je nutno konstatovat, že ani ve světové literatuře ani na webových sídlech jednotlivých zpracovatelů těchto dokumentů není literatura na téma věcné pořádání šedé literatury snadno k nalezení (k dispozici byly pouze dílčí příspěvky k tématu přednesené na konferencích či části v časopisecky publikovaných článcích). Příkladem můžou být manuály SIGLE, které jsou v elektronické podobě dnes už prakticky nedostupné a jediná možnost je nahlédnout do papírové verze, a to jen v knihovnách, kde se systémem SIGLE pracovali.

Problematika zveřejňování vysokoškolských kvalifikačních prací a snaha jejího řešení pomocí různých iniciativ jako je např. projekt Archivace a zpřístupnění elektronických verzí vysokoškolských kvalifikačních prací či Digitální knihovna šedé literatury je velmi aktuální. V rámci těchto projektů se řeší též metadatový standard pro popis elektronických diplomových a disertačních prací jehož součástí je i předmětové pořádání, které zatím nebylo jasně specifikováno. Prozatím je, např. v rámci návrhu metadatového standardu eVŠKP (elektronické vysokoškolské kvalifikační práce) pole pro předmět či klíčové slovo definováno jako nepovinné. V této práci se mimo jiné snažím dokázat, že definování tohoto pole jako povinného je opodstatněné. Také vzhledem k trendu zpřístupňování databází dokumentů v elektronické podobě na internetu a s tím souvisejícího vzájemného propojování elektronických informačních zdrojů pomocí nejrůznějších technologií pokládám za důležité věnovat se problematice předmětových hesel a jejich přidělování.

Práce je vypracována v souladu se schváleným zadáním. Jejím cílem je představení a zhodnocení systémů věcného pořádání v oblasti šedé literatury se zaměřením na databázi SIGLE a potřeby věcného pořádání a vyhledávání ve fondu šedé literatury ve Státní technické knihovně s přihlédnutím k možnostem Polytematického strukturovaného hesláře. Práce má tedy hodnotící, částečně deskriptivní charakter a v závěru je nastíněno možné konkrétní řešení ve smyslu maximalistického přístupu ve vyhledávání dokumentů. Do hodnocení a závěru jsou

začleněny prezentované poznatky z praxe. Naopak teoretické poznatky jsou více zohledněny v popisu jednotlivých systémů. Téma práce se vztahuje k současnému stavu archivace a zpřístupňování elektronických vysokoškolských kvalifikačních prací, proto mají některé části práce charakter návrhů. Při zpracování byly použity publikované i nepublikované dokumenty. Práce je doplněna seznamem použité literatury, které se vztahují jak k problematice věcné pořádání informací, tak vyhledávání a k jednotlivým systémům, které se zabývají šedou literaturou. K textu, jehož rozsah splňuje podmínky dané *Studijním a zkušebním řádem FF UK*, jsou připojeny 3 přílohy a CD-ROM s elektronickou verzí práce ve formátu PDF. Použité informační zdroje jsou citovány v souladu s ČSN ISO 690 a ČSN ISO 690-2. Citace v textu jsou vždy uzavřeny do hranatých závorek a obsahují příjmení autora (respektive název korporaci či zkrácený název díla) a rok.

V diplomové práci je nejprve v první části charakterizovány šedá literatura a dále v textu jsou analyzovány jednotlivé systémy zabývající se archivací a zpřístupněním šedé literatury v minulosti i dnes. V poslední části je specifikováno věcné pořádání a vyhledávání v databázích šedé literatury. V textu jsou prezentovány dílčí závěry, jejichž syntéza tvoří závěr diplomové práce.

Z poskytnutí cenných rad, konzultací a pomoci v průběhu zpracování práce bych ráda poděkovala vedoucímu své práce J. Schwarzovi, dále L. Skolkové oba z Ústavu informačních studií a knihovnictví Filozofické fakulty Univerzity Karlovy v Praze, Š. Žížkové a dalším kolegům a kolegyním ze Státní technické knihovny.

1. Úvod

Pořádání informací je dílčím procesem vstupního zpracování informací, jenž probíhá v rámci budování informačního systému. Procesem předcházejícím pořádání informací je informační analýza dokumentů, kterou se zjišťují významné identifikační a obsahové charakteristiky dokumentu. *Pořádání informací* je vytváření organizovaných souborů informací na základě určitého systému.

Věcné pořádání informací lze podle základní charakteristiky použitého systému pořádání rozdělit na systematické pořádání informací a předmětové pořádání informací. *Systematické pořádání informací* je proces, při kterém jsou informace vřazovány na dané místo v rámci systematicky uspořádaného souboru (systému) lidského poznání, přičemž jejich postavení se v zásadě řídí rodo-druhovými vztahy a slovní formulace obsahu dokumentu bývá většinou nahrazena znaky umělého jazyka (notacemi). Pro systematické pořádání informací se používá také označení třídění, klasifikace, systematické zpracování, systematická katalogizace ad. Systematickým pořádáním informací se vzhledem k zaměření práce nebudu dále zabývat.

Předmětové pořádání informací je proces, při kterém jsou informace vyjádřeny souborem abecedně uspořádaných hesel. Pro předmětové pořádání informací se používá také označení heslování, předmětové třídění, předmětová klasifikace, předmětové zpracování ad. Je nutno podotknout, že používání termínů označujících předmětové pořádání informací jako třídění, resp. klasifikaci není vhodné, protože způsobují terminologickou konfúzi [SCHWARZ, 1999].

Tvorba tezauru je proces sestavování (*compilation*), aktualizace (*updating*), údržby (*administration*) apod. slovníku deskriptorového selekčního jazyka. *Indexace* (indexování, *indexing*) je „proces vyjádření výsledku analýzy dokumentu prostřednictvím prvků selekčního jazyka nebo přirozeného jazyka, obvykle s cílem umožnit zpětné vyhledávání“ [ČSN ISO 5127-3a/3.2.1-03]. *Automatická indexace* (automatické indexování, *automated indexing*) je „vyjádření obsahu dokumentu pomocí automatického výběru slov nebo termínů z textu nebo pomocí automatického přiřazování termínů selekčního jazyka“ [ČSN ISO 5127-3a/3.3.3-01].

Pořádání a indexace je v této práci zaměřena na *šedou literaturu*. Konkrétně na *elektronické diplomové a disertační práce*. Tyto termíny jsou vysvětleny podrobněji dále v textu.

2. Šedá literatura

2.1. Charakteristika, dosavadní vývoj a trendy

O první zmínku o výzkumných zprávách, předchůdcích šedé literatury, se ve skutečnosti odborníci přou. Někteří tvrdí, že to bylo v roce 1909, kdy vyšla první kolekce zpráv Reports & Memorands společnosti Advisory Committee for Aeronautics in the United Kingdom. Jiní tvrdí, že to bylo dříve, ale dokument neobsahoval slovo „zpráva“. Jednalo se noviny The Professional Paper vydané 1902 společností The US Geological Survey.

Rozvoj zaznamenaly **výzkumné zprávy** až během druhé světové války, spolu s akcelerací vědeckého výzkumu pro válečné účely a předávání informací o nich. Stalo se tak zejména v hlavních bojujících mocnostech, tj. Německu, Velké Británii a Spojených státech amerických.

V 70. a 80. letech 20. století docházelo k nárůstu různých podobných studií publikovaných ve vybraných knihovnických časopisech a na **konferencích věnovaných šedé literatuře**, což vedlo k myšlence nebo spíše nutnosti zlepšit všeobecné povědomí o těchto studiích. Nejvíce známým časopisem byl bezpochyby Seminars of York, který položil základ spolupráce mezi evropskými knihovnami a dokumentačními centry, což posléze vedlo ke vzniku databáze SIGLE.

Mimochodem termín **šedá literatura** se poprvé oficiálně objevil v roce 1970 názvu německého článku v souvislosti se špatně nalezitelnou literaturou. Výraz šedá literatura ale byl již užíván od roku 1920 a označoval širší soubor dokumentů (tj. nejen výzkumné zprávy), které byly nedostupné uživatelům běžnými komerčními vydavatelskými cestami.

První mezinárodní konference o šedé literatuře GL'93 představila velmi ambiciózně nový typ dokumentu. Ve stejném roce byla ustavena **Komise pro vědecké a technické informace** (The Committee on Scientific and Technical Information, COSATI), která měla sloužit k šíření informací v těchto dvou oborech. Zajímavé produkty COSATI již pocházejí ze 60. let 20. století. Byly to zejména Struktura předmětových kategorií a podkategorií (Subject Category and Subcategory Structure) z roku 1964 a Standard pro katalogizaci a vládní vědecké a technické zprávy (Standard for Descriptive Cataloguing and Government Scientific and Technical Reports) z roku 1966. Oba nástroje byly široce využity jako základ pro bibliografický popis v systémech typu Mezinárodního informačního systému pro jadernou energetiku (International Nuclear Information System, INIS), Mezinárodního informačního systému pro zemědělství a technologii (International Information System for the Agricultural Sciences and Technology, AGRIS) či SIGLE.

Mezinárodní konference o šedé literatuře se od roku 1993 konaly pravidelně každé dva roky a pořádala je nizozemská komerční služba GreyNet <http://www.greynet.org/> . Bohužel plánovaná konference v roce 2001 se již nekonala, ale další pokračovaly již od roku 2003. V roce 2007 slaví služba GreyNet 15 let od svého založení. Tematicky byly konference orientovány na typologie šedé literatury, její producenty nebo bibliografický popis atd. Mezi **producenty** patří např. akademické organizace, orgány státní a místní správy, obchodní organizace nebo průmyslové organizace. Na to navazuje **typologie** šedé literatury, která může být např. zpráva, preprint, disertace, normy, dokumenty z konferencí atd. Typologie se uvádí v různém rozsahu a záleží na autorovi zpracování [LUZZI, 2000].

2.2. Co je to šedá literatura?

2.2.1. Definice a vlastnosti šedé literatury

Jedna z nejpochopitelnějších definic šedé literatury byla stanovena na konferenci GL'97 a zní

„Šedá literatura je literatura produkována na všech úrovních státní, resp. místní správy, akademických a obchodní a průmyslových organizací. Může existovat jak v tištěné tak v elektronické formě a není kontrolována komerčními nakladateli.“

Na otázku, co je tedy vlastně šedá literatura, je často odpověď podobně nejasná, jako je nejasný např. význam pojmů „poloveřejná“ nebo „nekonvenční“. Jindy je termín vysvětlován na příkladech, jako jsou výzkumné zprávy, disertace, sborníky, úřední věstníky, preprinty či překlady. Řada těchto druhů dokumentů však nepředstavuje nikterak nekonvenční informační prameny. Jsou konvenční již svým vzhledem a lze si je i předplatit. Snad nejpraktičtější definice za šedou literaturu pokládá „dokumenty, které nejsou dostupné běžnými distribučními kanály“ [WOOD, SMITH, 1991].

Kromě specifických způsobů distribuce se ale šedá literatura vyznačuje mnoha dalšími vlastnostmi – nízkým nákladem, obtížně zpracovatelnými formami, např. u mikrofiší či mikrofilmů, různými vydavatelskými a výrobními normami, nízkou publicitou, nedostatečným bibliografickým zpracováním a špatnou dostupností v knihovnách. Šedá literatura má také jinou důležitou vlastnost – obsahuje informace, které jdou často do mnohem větších detailů než u běžné literatury a které jsou obecně dostupné rok až rok a půl před vydáním časopiseckého článku nebo knihy. V každém oboru lze tedy šedou literaturu bezpochyby pokládat za potenciální zdroj užitečných informací.

2.2.2. Krize šedé literatury

Nejvýznamnějším nedostatkem knihovnicko-informačních služeb působících v rámci Evropské unie byla v roce 1994 právem označována situace v sekundárních informačních fondech, resp. bibliografickém zachycení a v katalogizaci právě těchto primárních dokumentů. Tato situace částečně přetrvává dodnes. To znesnadňuje přístup k informacím, které se jinak než v oficiálně nepublikované formě nevyskytují nebo se v nejlepším případě standardně vydavatelsky publikují v běžných komerčních časopisech a knižních publikacích. Problém prudkého rozvoje relativně obtížně definovatelného okruhu druhů dokumentů, doprovázejících soudobý vývoj informačního zázemí Evropské unie, vznikl pochopitelně již dříve než na přelomu 80. a 90. let 20. století. Už v roce 1963 se touto literaturou zevrubně zabývala slavná **Weinbergova zpráva** „Science, government, and information - the responsibilities of the technical community and the government in the transfer of information“¹ (viz Příloha č. 1), analyzující stav a východiska nové vědní (a také informační) politiky USA v konfrontaci s tehdejší vývojem sovětského kosmického programu. Weinberg tuto literaturu již výstižně identifikoval jako velmi významnou a perspektivní informační dokumentovou kategorii. Pouze ji nepojmenoval [BRATKOVÁ, VLASÁK, 1994].

2.2.3. Propagace

Nechť a nebo neschopnost producentů propagovat vzniklou šedou literaturu způsobuje problémy abstraktním a indexačním službám, které ji chtějí shromažďovat a zpřístupňovat koncovým uživatelům, kteří o ni jeví zájem. V Evropě i jinde se procesy akvizice, bibliografického zpracování a zpřístupňování primárních pramenů šedé literatury ve velkém soustředily do několika specializovaných institucí. Dvě velké instituce – Institut pro vědecké a technické informace (L'Institut de l'information scientifique et technique, INIST) ve Francii a Centrum pro dodávání dokumentů v Britské knihovně (British Library Document Supply Centre, BLDSC) ve Velké Británii jsou příkladem národních středisek, která v širokém záběru shromažďují konvenční i šedou literaturu. Např. BLDSC získává celosvětovou šedou literaturu ze všech oborů již 25 let [WOOD, SMITH, 1991]. Zahnutí šedé literatury do rostoucího rejstříku signálních médií však může vést k nežádoucímu růstu neuspokojených požadavků, pokud signál o existenci dokumentu neobsahuje zároveň informaci o jeho dostupnosti (což je obecně platné pravidlo). Ve srovnání s pokroky USA bylo v této oblasti bibliografické zpracování a poskytování těchto dokumentů v Evropě špatné. Dnes se sice zájemci např. o disertace dostanou až přímo k dokumentu, ale pouze prezenčně a v příslušné

¹ Dostupné i v české verzi ve Státní technické knihovně [BLOCH, 1964].

fakultní nebo univerzitní knihovně. Stále se vedou diskuze o přizpůsobení dostupnosti vysokoškolských prací tak, aby byla v souladu s Autorským zákonem, což vede k dalšímu omezení dostupnosti šedé literatury [ČESKO, 2000].

2.2.4. Prosazení šedé literatury v digitální době

Šedá literatura je ceněna zejména pro originální, hodnotný obsah a aktuálnost. Zejména v dnešním období síťové komunikace informací vykazují velkou aktuálnost preprinty článků do odborných časopisů. Literatura tohoto typu obsahuje v mnoha případech detailnější informace než ty, jež se objevují v publikovaných zdrojích.

Postupem času přišla nevyhnutelně tendence na nějakém místě tuto literaturu sdružovat a posléze zpřístupňovat. To vedlo ke tvorbě společných databází na národní i mezinárodní úrovni. Jedním z nich byl i mezinárodní systém pro šedou literaturu SIGLE.

2.2.4.1. Proč nepublikovat

Mezi autory šedé literatury je mnoho těch, jejichž mateřské či sponzorské instituce mají pocit, že není nutné publikovat dané materiály v časopisech. Často je to třeba jen proto, že nejsou zainteresováni na počtu citací dané práce, protože financování jejich projektů není závislé na tomto způsobu hodnocení. Jejich primárním cílem je dané informace co nejrychleji a současně nejlevněji rozšířit a zpřístupnit je vymezenému okruhu lidí nebo organizací. Přitom se chtějí vyhnout značnému časovému zpoždění souvisejícímu s časopiseckým publikováním. A dnešní informační technologie, např. tzv. desktop publishing², využití blogů³ atd. jim umožňují mít publikační činnost pod vlastní kontrolou [WESSELS, 1994].

2.2.4.2. Proč publikovat

Hlavními typy institucí, které vydávají šedou literaturu, jsou výzkumné ústavy, univerzity (jejich katedry, ústavy apod.), mezinárodní i místní úřady a orgány a průmyslové podniky. Samozřejmě existují platné důvody, proč „publikovat šedě“. Např. některé zprávy mají i komerční hodnotu, a je proto zájem zachovat materiál jako důvěrný. Dalším důvodem k šedému publikování je skutečnost, že obsah zveřejňovaných materiálů je takového charakteru, že lze očekávat zájem pouze velmi úzké skupiny lidí, nebo že je příliš dlouhý (případně krátký) pro běžné komerční publikování. A mohou jistě existovat i další důvody.

² Výraz **desktop publishing** (zkráceně **DTP**) pochází z angličtiny. Jedná se o tvorbu tištěného dokumentu za pomoci počítače. K tomu, aby mohl dokument vzniknout, je zapotřebí kromě počítače a obsluhy, obvykle DTP operátor nebo grafik, také sázecího počítačového programu [WIKIPEDIA, heslo desktop publishing].

³ **Weblog** (stažením anglického **web log** – zhruba: webový záznam[ník]) nebo zkráceně a neformálněji **blog** je webová aplikace obsahující periodické příspěvky na jedné webové stránce [WIKIPEDIA, heslo blog].

Přesto však je většina šedé literatury, s níž pracují knihovny a producenti databází, vydávána a to zejména oficiálními institucemi, organizacemi a orgány [WESSELS, 1994].

2.2.4.3. *Producenti*

Šedá literatura není běžně zpracovávána pro producenty databází z oblasti vědy a techniky. Producenti takových databází jako je BIOSIS nebo MEDLINE se jen obtížně vyrovnávají s množstvím časopisecké literatury, a tak šedou literaturu – i s přihlédnutím k obtížím spojeným s jejím shromažďováním – vyloučili ze záběru svých databází. Přesto však existují producenti, kteří se touto literaturou také zabývají. Jde o specializované mezinárodní databáze z oblasti techniky a technologie, příkladem mohou být báze dat produkované společností Engineering Information Inc. např. Compendex, Inspec, GeoBase, Ei Patents a další. Mnoho uživatelů/vědců těchto databází je současně též autory šedé literatury a proto doceňují její hodnotu [WESSELS, 1994].

2.2.4.4. *EAGLE-SIGLE*

Uživatelé ze zemí dnešní Evropské unie se mohli pro šedou literaturu obrátit do báze **Systému pro informace o šedé literatuře v Evropě** (System for Information on Grey Literature in Europe, SIGLE). Vyhledávat v ní bylo možné prostřednictvím databázových center BLAISE, STN International a SUNIST (Serveur Universitaire National de l'information Scientifique et Technique, France). Uživatelé ze zemí, které nebyli členy EU, mohli do SIGLE vstupovat především prostřednictvím databázového centra STN International. Báze SIGLE byla ovšem dostupná také ve formě CD-ROM, které vydávala dnes již neexistující společnost SilverPlatter a prostřednictvím magnetopáskové služby s měsíční periodicitou [WESSELS, 1994].

Síla SIGLE spočívala ve využívání existujících prostředků. Šedá literatura zahrnutá do SIGLE byla spolupracujícími centry a knihovnami shromážděna a bibliograficky zpracována pro jejich vlastní účely a tak bylo třeba vynaložit úsilí pouze na aplikaci příslušného věcného pořádacího systému a na překlad záznamů do angličtiny v případě, že originální titul byl publikován v jiném jazyce. **Evropská asociace pro využívání šedé literatury** (EAGLE) usilovala o zdokonalení svých služeb pro uživatele ve třech směrech:

- větší rozsah,
- dokonalejší věcný přístup a
- nové produkty [WESSELS, 1994].

Dokonalejší věcný přístup k dokumentům zpracovaným pro SIGLE byl potřebný především vzhledem k nárůstu databáze a také ve snaze udělat ji uživatelsky přívětivější. Dokumenty se věcně zařadily pomocí 246 širších předmětových kategorií. V rámci kategorií bylo možné vyhledávat v angličtině, pomocí slov z anglického názvu publikace, názvovými slovy z titulu v jazyce originálu a klíčovými slovy, která byla doplněna pro objasnění titulu. Doplnující klíčová slova bohužel obsahovalo pouze 15 % záznamů v bázi a EAGLE dále zkoumalo možnosti, jak zajistit jejich širší používání.

SIGLE a EAGLE pracovaly pouze s tištěnou verzí prací a jejími bibliografickými záznamy. Kdyby nezanikly řešily by dnes pravděpodobně problémy spojené zejména s elektronickými diplomovými a disertačními pracemi, které jednoznačně nabývají na významu a stávají se trendem.

3. Elektronické diplomové a disertační práce

Disertace a diplomové práce sdružují informace z rozsáhlých výzkumů a jsou přehledové a hodnotné zdroje informací, které je možno produkovat buď v tištěné nebo elektronické podobě. Z odborných odhadů vyplývá, že v Centrum pro dodávání dokumentů v Britské knihovně je okolo tří milionů těchto dokumentů, ale půjčuje se jich jen malá část. Nízká půjčovatelnost šedé literatury je přímo spojena s problémem identifikace vhodných materiálů. Na vzestupu je momentálně sběr, pořádání, zpřístupňování a archivace elektronických verzí diplomových a disertačních prací, anglicky nazývaných Electronic theses and dissertations (ETD).

3.1. Rozdělení

Existují dva základní typy těchto dokumentů. První typ, který je výrazně preferován od počátku sbírání podkladových materiálů až po jeho tvorbu, je **práce psaná samotným autorem**. Jinak řečeno, jedná se o práce, které píše sám student již během studia **pomocí nějakého softwaru** jako např. Microsoft Word či LaTeX, a zpřístupňuje je elektronicky přímo v databázi na své univerzitě, na webových stránkách, na internetu atd. Běžný formát dokumentu, např. MS Word „DOC“ je lehce konvertovatelný do formátů, PDF nebo XML, které jsou vhodnější pro účely uchovávání a zpřístupňování. Tyto formy dovolují zpřístupnit dokumenty přes snadno dostupné sítě obvykle pomocí metadat (obvyklá struktura obsahuje název, rok, autor, abstrakt a popis). Takto zpřístupněná elektronická diplomová nebo disertační práce může být nalezena téměř odkudkoliv, podle oblasti již zabývá, pomocí

vyhledávání a prohlížení metadat nebo vyhledáváním v celém textu dokumentu, přičemž text může obsahovat klidně i multimediální komponenty (obrázky, video nebo hudbu).

Druhý typ elektronických diplomových a disertačních prací je **elektronický soubor**, který byl **vytvořen naskenováním tištěné verze** diplomové práce nebo disertace. Výsledky této metody jsou mnohem obtížněji zpracovatelné než výše zmíněné e-born dokumenty. Jsou náročnější z hlediska paměti, podpora fulltextového vyhledávání je problematická, manipulace s nimi není lehká (nemohou být např. volitelně zmenšovány kvůli hrozící zhoršené kvalitě pro čtení) a nevedou studenty – autory - ke standardním postupům elektronického publikování. Nicméně takto naskenované a dostupné stránky za méně peněz umožňují následný tisk a studium originálního textu a jsou řešením pro retrospektivní digitalizaci [UNESCO, 2001].

3.2. Zpřístupnění - dostupnost

Jednou z výhod zpřístupnění diplomových a disertačních prací elektronicky je **redukce opakování výzkumů**, které již byly provedeny, protože studenti jsou upozorňováni na již dříve zpracované práce jiných kolegů. Výjimkou jsou nestandardní práce, např. habilitační práce, které jsou zřídka prezentovány v databázích (jen několik jich je v databázích UMI Services, např. Dissertation Abstracts). S pomocí celosvětově dostupné kolekce elektronických vysokoškolských kvalifikačních prací mohou studenti rychle hledat práce podle svého zájmu z jakéhokoliv místa na světě a v mnoha případech zkoušet a učit se z nich bez velkých nákladů. Práce zpřístupněné přes internet nevyžadují **žádné náklady** ve srovnání se stovkami a tisíci nutnými na preferovanou tištěnou verzi a její publikování na papíru nebo jiných médiích. Počítačová síť a software dostupný v mnoha institucích pro všechny studenty, jednoznačně vedou k napsání práce elektronicky, čímž klesají náklady na její vytvoření.

3.2.1. Podpora informační gramotnosti studentů

Na mnoha kampusech je pro studenty připravena **poradna a přípravné kurzy** zaměřené na psaní a další dovednosti, které jsou potřebné při tvorbě elektronického dokumentu a rovněž i pro uživatele digitální knihovny. Pokud si student vybere osobní počítač a požaduje užívat vlastní software k tvorbě elektronických diplomových a disertačních prací, může být tento později užitečný pro další výzkum a rozvoj jejich tvorby. Skrytý záměr podpůrných aktivit je připravit další generaci studentů na jejich zapojení jako znalostních pracovníků do informační společnosti 21. století.

Hlavní **zájmy univerzit** v oblasti vysokoškolských kvalifikačních prací jsou :

- znalosti studentů jak se zapojit a používat digitální knihovny,
- budování univerzitních digitálních knihoven, služeb a infrastruktury,
- rozšíření sdílení výsledků výzkumů prováděných na jednotlivých univerzitách,
- elektronické diplomové a disertační práce ve vysoké kvalitě a zvláště při jednotlivých vyhledáváních studenty

[UNESCO, 2001].

3.2.2. Browsing

Prohlížení (browsing) je klíčem k přístupu k elektronickým diplomovým a disertačním pracím. Prohlížení probíhá v podobném duchu, v jakém se člověk rozhlíží po knihovně. Často vtáhne vědce do učebního procesu spojeného s koncepcí, oblastmi a použitým slovníkem v konkrétním oboru. Vědec se obvykle pohybuje okolo „prostoru konceptu“ a hledá **blízká nebo vzdálenější podobná témata**, příklady či aplikace teorií a metod. Např. v lékařské práci podněcuje prohlížení vědce k přemýšlení o nemoci, její léčbě, umístění v organismu, symptomech a dalších spojitostech. V mnoha oborech obsahuje prohlížení vysvětlení a taxonomický systém, ovládání organizačního schématu nebo organizační struktury, či umí přeměrovat uživatele z termínu na bližší slovo nebo frázi (více v kap. 5.8.3 a 5.8.4) [FISHER, 1995].

3.2.3. Systémy pro věcné pořádání informací

Řada vědeckých disciplín má k pořádání informací a jejich vyhledávání rovněž oficiální pořadací a klasifikační systém. Některé jsou dost široké, v některých případech pokrývají celé obory, např. předmětová hesla vytvořená v Kongresové knihovně (Library of Congress Subject Headings, LCSH) nebo Deweyho desetiné třídění (Dewey Decimal Classification System, DDC). Tedy pro hlubší charakterizování výzkumné práce v konkrétní oblasti je vhodnější použít pořadací systém z této oblasti. V oboru informatiky je populární systém ACM (Association for Computing Machinery Classification System), v medicíně MeSH (Medical Subject Headings) nebo UMLS (Unified Medical Language System), který je široce využíván. Pro fyziku je to pak PACS (Physics and Astronomy Classification Scheme). Nakonec tedy digitální knihovna zpřístupňující elektronické diplomové a disertační práce by měla prosazovat prohlížení a dotazování pomocí širšího schématu, např. DDC stejně jako užší vyhledávání pomocí integrovaných oborových heslářů a tezurů [DAHLBERG, 1974].

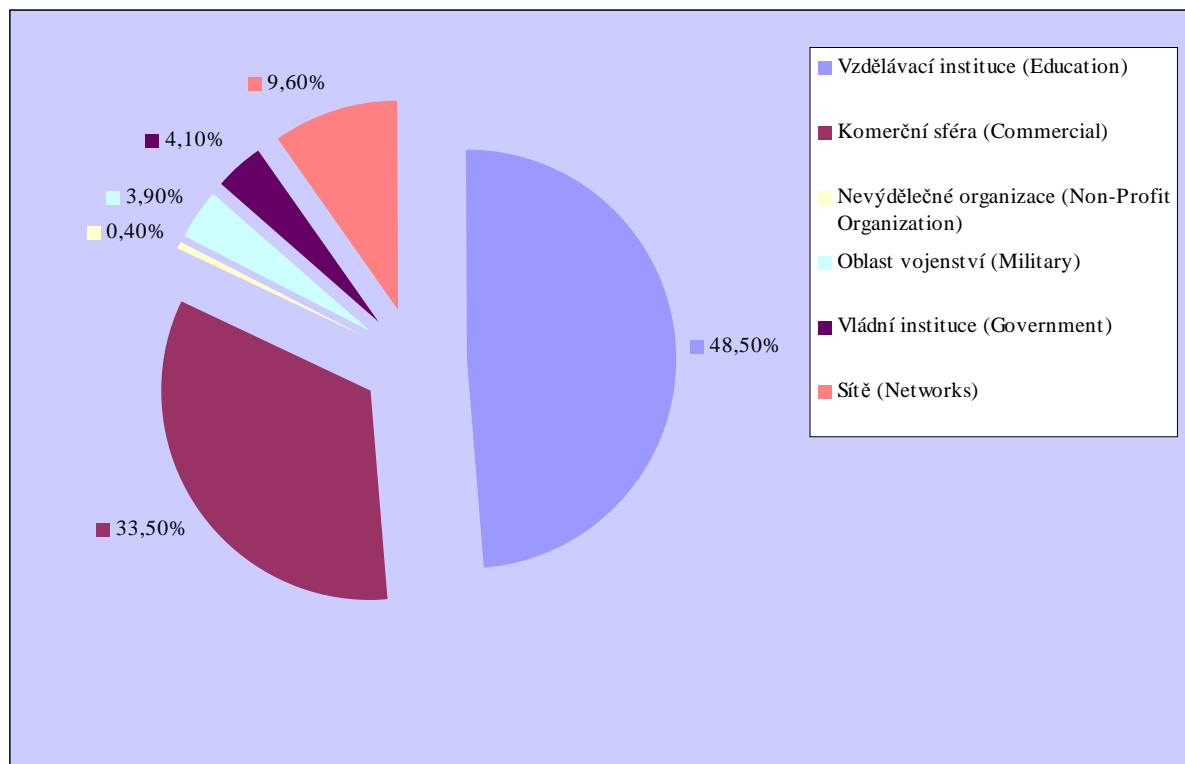
3.3. Rozvoj a propagace

Co se týká rozvoje elektronických diplomových a disertačních prací, za všechny jeden příklad. Studenti z Technické univerzity ve Virginii, kteří jsou současně autory a produkují elektronické diplomové a disertační práce, oznámili v roce 1999 ambiciózní publikační plán. Více než 85 % dotazovaných řeklo, že plánují publikovat na základě své diplomové nebo disertační práce článek, kapitolu, knihu atd. Jen 13,7 % neplánuje publikovat nic. Přes dvě třetiny studentů, kteří končí studium, řeklo, že plánuje publikovat článek nebo stať ve sborníku z konference. Jen malá část (cca 26 ze 445 dotázaných, tj. 5,8 %) plánuje publikovat kapitolu, zatímco 31 (7 %) plánuje publikovat celou knihu. Rovněž výsledky ankety ohledně zpřístupnění elektronických vysokoškolských kvalifikačních prací jsou uspokojivé. Celkem 56,5 % respondentů považují proces jejich zpřístupnění za méně složitý, než čekali, a téměř jedna čtvrtina pak za více složitý, než očekávali. Podobně je tomu s procesem převodu těchto prací do formátu PDF [McMILLAN, 1999].

3.4. Úložiště pro elektronické diplomové a disertační práce

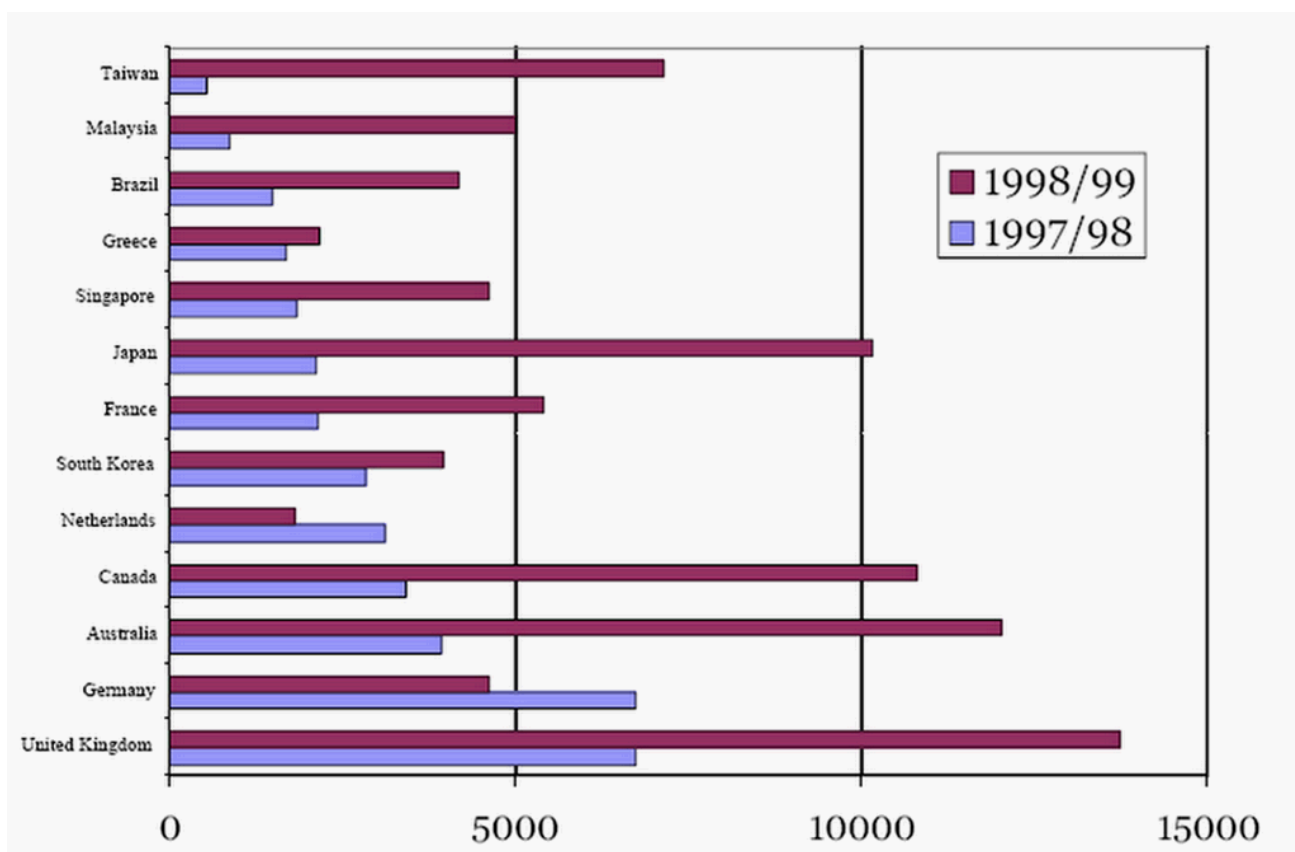
Nástup elektronických diplomových prací a disertací přináší nutnost, aby knihovny vzaly na sebe zodpovědnost za umístění těchto prací včetně jejich zpřístupnění z informačního serveru a jejich archivaci. S nimi začíná pro knihovny éra zlepšení metod a způsobů uchovávání informací, stává se z nich úplná digitální knihovna. Zpřístupňují informace rychleji, lépe tak slouží uživatelům a to s ušetřením personálu i místa na skladování papírových dokumentů. Některé druhy dokumentů jsou pro knihovny nové, ale většina vyžaduje podobné zpracování a vyhledávání jako tradiční dokumenty.

Graf 1: Složení uživatelů ETD v USA [McMILLAN, 1999].



Graf 1 ukazuje, jakému spektru uživatelů a v jakém rozsahu slouží knihovny poskytováním elektronických diplomových a disertačních prací. Zajímavá je vínová část, která naznačuje, že uživatelé z komerční sféry tvoří celou třetinu ze všech uživatelů těchto informací. Je tedy zřejmé, že potenciál využitelnosti elektronických diplomových a disertačních prací je třeba hledat i mimo akademickou sféru.

Graf 2: Složení uživatelů ETD v mezinárodním pojetí [McMILLAN, 1999]



Na grafu 2 je pak vidět, jak s databází NDLTD pracují uživatelé z různých zemí světa. Samozřejmě, že čím vyspělejší stát a ekonomika, tím více je o elektronicky dostupné práce a jejich využití zájem např. při vědeckých výzkumech.

3.4.1. Rizika ukládání

Možná rizika pro elektronické diplomové a disertační práce jsou ve stejném nastavení kritérií jejich zpracování, která byla aplikována již u tištěných prací. V některých institucích je na elektronické diplomové a disertační práce nahlíženo stejně jako na tištěné, jsou definovány stejnými podmínkami a podléhají stejným postupům zpracování a zpřístupnění. To vyvolává různé problémy, které mají dopad i přímo na elektronické publikování. Někteří odborníci se domnívají, že se jedná v podstatě o tištěné práce pouze převedené do elektronické podoby, ale aplikace metod vytvořených pro tištěné verze je problematická. Zvláště v případech, kdy elektronická verze diplomové a disertační práce obsahuje hypertextové odkazy či multimedia. Další nezanedbatelný problém, který se v institucích podporujících tento druh prací objevil, je dostatečné softwarové vybavení počítačů uživatelů (tj. studentů i vyučujících), kteří práce pročítají. Seth Kratz z Bradley University v Illinois přímo řekl, že je výborné pokud student postupuje dle pokynů a napíše práci v elektronickém formátu jako hypermediální diplomovou či disertační práci, ale pokud si to on i kdokoliv další nebude moci na svém počítači přečíst,

vzniká problém [FOX, 2004]. Je tedy třeba dobře promyslet formáty a potřebný software, také vzhledem k jeho zastarávání a možnostem migrací verzí, k tvorbě běžných elektronických i hypermediálních diplomových prací či disertací.

3.5. Závěr

Změny v tradičním vnímání vysokoškolských prací byly vyvolány zejména změnou akademického prostředí. Elektronické publikování jim dává nový rozměr, větší hodnotu, kterou může zajistit citovanost práce, zlepšení propagace na trhu informací. Změny lze shrnout do tří základních okruhů:

- Čím dál více studijních materiálů se zpřístupňuje elektronicky a on-line.
- Ze studijního prostředí se stává obchodní prostředí, které se snaží konkurovat okolnímu světu informací.
- Studijní plány jsou téměř šité na míru nejen zahraničním, ale i běžným prezenčním studentům [FOX, 2004].

4. Charakteristika vybraných systémů

Do této kapitoly jsem vybrala Systém informací o šedé literatuře v Evropě (SIGLE), Síťovou knihovnu diplomových a disertačních prací (NDLTD), Program australských digitálních disertací (ADT) a popis selekčních jazyků používaných v STK.

Důvodem mého výběru samotných systémů byla jejich architektura a jejich obliba uživateli a s tím spojené množství implementací. Systém SIGLE je příkladem jediného celoevropského fungujícího systému pro šedou literaturu. U systémů a selekčních jazyků používaných v STK jsem se doústředila na doposud nepopsaný historický obraz využití předmětových hesel nad fondy v STK.

4.1. Systém informací o šedé literatuře v Evropě (SIGLE)

4.1.1. Úvod

SIGLE je zkratka pro Systém informací o šedé literatuře v Evropě (System for Information on Grey Literature in Europe). Jako základní funkce bylo stanoveno získávání, zpracování a zpřístupňování bibliografických informací o zprávách a jiné šedé literatuře a dále pak informace, kde je dokument dostupný v plném textu. Oficiální informace a dokumentace k SIGLE byly umístěny v nizozemské Národní knihovně v Haagu.

4.1.1.1. Historie

System SIGLE vznikl v roce 1980 pro zpřístupňování dokumentů formou DDS (document delivery service⁴). Jednalo se o šedou literaturu vytvářenou v zemích Evropské unie. Základním nástrojem systému SIGLE je bibliografická báze dat připravovaná decentralizovaně s centrální kontrolou v účastnických institucích a zpřístupňovaná on-line prostřednictvím databázových center nebo na CD-ROM. On-line hlavně za pomoci center STN International a na CD-ROM ho distribuovala firma SilverPlatter, kterou v roce 2001 převzalo databázové centrum OVID Technologies. Hlavním cílem bylo zlepšení registrace a šíření bibliografických informací o těchto dokumentech [LUZZI, 2000].

Za zakladatele lze považovat Evropskou komisi, která odstartovala a finančně zaštitila vznik SIGLE a Evropské asociace pro využívání šedé literatury (European Association for Grey Literature Exploration, EAGLE).

4.1.1.2. Evropská asociace pro využívání šedé literatury (EAGLE)

System SIGLE využíval zdrojů velkých Evropských informačních a dokumentačních center, která byla členy výše zmíněného sdružení EAGLE, jež bylo založeno v roce 1985, aby pokračovalo ve vedení projektu a otevřelo jej pro všechny evropské země. To se úplně podařilo splnit v roce 1992.

Šlo o sdružení velkých evropských informačních a dokumentačních i knihovnických center, které bylo řízeno Valným shromážděním tvořeným zástupci patnácti členských zemí.

Každá účastnická instituce odpovídala za vytváření fondů šedé literatury produkované v dané zemi a za přípravu detailních informací o ní pro bázi dat SIGLE. Instituce rovněž ručily za to, že všechny registrované dokumenty byly dostupné na základě lokace, která byla uvedena v každém záznamu [BRATKOVÁ, VLASÁK, 1994]. Za uvedení záznamu do systému se platilo zhruba 15 LUF⁵ (asi 12 Kč) a účastnický poplatek činil cca 30 000 LUF ročně (asi 22 500 Kč).

4.1.2. Členové

Mezi členy patřily:

- Všeobecná knihovna a knihovna humanitních věd Katolické univerzity (Belgie),
- Akademie věd ČR a Státní technická knihovna (Česká republika),
- INIST (Francie),

⁴ Elektronické dodávání dokumentů.

⁵ Lucemburský frank.

- Centrum pro dodávání dokumentů v Britské knihovně - British Library Document Supply Center, BLDS (pro Irsko, Velkou Británii a Komisi Evropských společenství),
- Ústřední knihovna Národní Rady pro vědu a výzkum (Itálie),
- Lotyšská akademie věd (Lotyšsko),
- Národní knihovna (Lucembursko),
- Ústřední knihovna Budapešťské technické a ekonomické univerzity (Maďarsko),
- FIZ Karlsruhe (Německo),
- Univerzitní knihovna a Technická informační knihovna Hannover (Německo),
- Královská knihovna (Nizozemí),
- Nizozemský institut vědecko-informačních služeb (Nizozemí)
- Knihovna Technické univerzity Delft (Nizozemí),
- Nadace pro vědu a technologii při Ministerstvu pro vědu a technologii a SID – Informační a dokumentační služba Lisabon (Portugalsko),
- Všeruské vědecké technické informační centrum (Ruská federace),
- Centrum vedecko-technických informací (Slovenská republika),
- Centrum vědeckých informací a dokumentace (Španělsko)

[BRATKOVÁ, VLASÁK, 1994].

4.1.3. Analýza SIGLE

Je zajímavé a užitečné podívat se na přednosti a nedostatky systému SIGLE, jak je přednesli D. N. Wood a A. W. Smith z Britské knihovny v roce 1991 na konferenci I'91 [WOOD, SMITH, 1991]. Tato analýza totiž umožňuje posoudit, do jaké míry může být systém vzorem pro mezinárodní spolupráci i v jiných oblastech bibliografické činnosti. Přijmeme-li kritérium množství sebraných záznamů, lze jakýkoliv systém, který během desetileté existence dospěl ke shromáždění a mezinárodní distribuci 30 000 záznamů ročně, pokládat za úspěšný.

Jedním z faktorů, které vedly k tomuto úspěchu, je bezpochyby podpora ze strany Komise Evropských společenství (KES). Její výchozí finanční dotace do doby, než se konsolidovala národní střediska, byla zásadní. Angažování KES navíc zvýšilo prestiž systému SIGLE v očích těch zemí, jež o zapojení do tohoto systému uvažovaly. Její trvalá podpora také zajistila minimum provozních a jazykových problémů, častých při podobné mezinárodní spolupráci, a to poskytnutím jednacích prostorů a překladatelských služeb. Systém měl štěstí i v tom, že získal podporu několika významných institucí ochotných finančně přispět. Úspěch lze přisoudit i včasnému přijetí dobře definovaných „pravidel hry“ a vytvoření funkční řídicí

struktury, v níž Rada určovala politiku systému a Technická skupina zajišťovala vytyčení technických cílů a jejich dosažení [WOOD, SMITH, 1991].

4.1.3.1. Slabiny systému SIGLE

Pokud se týká slabin systému, můžeme uvést následující body: ne všichni potenciální účastníci byli stejnou měrou schopni přizpůsobit své národní zvyklosti. V některých zemích bylo poměrně obtížné nalézt prostředky na podporu systému. Někteří je naopak získali snáze právě tím, že významná část šedé literatury jejich země byla plánována převést do mezinárodních systémů. Tam, kde existují centrální registrace prací, bylo samozřejmě snazší zajistit účinnější spolupráci. To vysvětluje i fakt, že podíl Velké Británie, díky BLDSK jako centrálního registru, byl v součtu vyšší, a že u zemí s relativně decentralizovaným postavením institucí, jako je např. Belgie, byl vstupní objem menší. Zájem projevil i další země, např. Řecko, které však nebylo vůbec schopno spolupráce, zejména z důvodu absence jakéhokoliv funkčního národního systému.

4.1.3.2. Problematika tvorby jednotné báze šedé literatury

Ačkoliv SIGLE lze pokládat za úspěch díky vzájemné spolupráci národních „vstupních“ center, bylo obtížné docílit konzistence v pokrytí celého spektra šedé literatury, protože jednotlivá centra mají odlišné národní role a funkce. Centrum pro dodávání dokumentů v Britské knihovně, britské národní centrum SIGLE, mělo a stále má nejširší odpovědnost za zpřístupňování všech druhů šedé literatury ve všech oblastech. Ukázalo se, že pro sjednocení činností bylo nezbytné zřídit funkci placeného výkonného tajemníka, jehož posláním bylo organizace porad a realizace dalších akcí. Dále začalo být zřejmé, že EAGLE musela být samostatným právním subjektem, protože byl nucen vstupovat do smluvních vztahů. Musel být také pokládán za obchodní subjekt s vlastním financováním, rozpočtem a obchodními cíli. Práce realizované pro EAGLE musely být, jak jen to bylo možné, prováděny za honorář. Spoléhání na dobrovolnou práci totiž mohlo vést k její nižší kvalitě a prioritě, případně by práce byly vůbec zastaveny. Důležité bylo, aby finanční struktura měla podobu, která zajistí kontinuální závazky všech účastníků. Od každého z nich, bez ohledu na objem vstupů, byl očekáván významný podíl na celkových nákladech. Samozřejmě bylo nutné prověřování technických stránek systému a tím zajištění jeho efektivnosti. Bylo tedy důležité disponovat skupinou technických expertů, kteří by společně pracovali na vývoji a aktualizaci v rámci systému, např. pořádacího systému, vstupních formátů, uživatelských příruček apod. Technické centrum v tom hrálo absolutně hlavní roli. Ovšem ať by byly snahy v rámci EAGLE jakkoliv dobré, bez efektivního zpracovatelského centra, muselo skončit na nule [WOOD, SMITH, 1991].

4.1.3.3. *Východisko z krize*

Velká péče se tudíž musela věnovat volbě centra a průběžnému sledování jeho efektivity, jinak by kvalita a efektivita systému mohly upadat. V provozu SIGLE nebylo snahou formulovat rigidní pravidla členství. Určité zásady byly pochopitelně nutné, avšak na druhé straně právě jejich vysoká flexibilita umožnila řadě zemí podle jejich možností větší či menší účast. I když Velká Británie, SRN a Francie přispěli do systému zdaleka nejvíc, v porovnání s ostatními účastníky, nikterak to nesnížilo podíl dalších států na řízení a rozvoji systému. Jejich vstupy postupně rostly a účast se zvětšovala [WOOD, SMITH, 1991]. Tak, jak další země nacházely zdroje a politickou vůli připojit se, systém stále sílil. Hledaly se cesty k úpravě statutu, aby do EAGLE mohly vstoupit i nečlenské země tehdejších Evropských společenství (ES), a počítalo se s tím, že se z EAGLE stane vskutku celoevropské sdružení. Ani tyto snahy bohužel nakonec neudržely systém SIGLE v chodu.

4.1.4. SIGLE v ČR

4.1.4.1. *Úvod*

Česká republika se stala jednou z prvních zemí tzv. východního bloku, které byly přijaty do tohoto systému. Byla reprezentována především Knihovnou AV ČR (KNAV), která byla přijata již v roce 1994 na základě svých dlouholetých zkušeností s registrací šedé literatury ve vlastním systému ASEP (Automatizovaný systém evidence publikovaných prací). Knihovna přizpůsobila vstupní formát ASEP pro ukládání záznamů požadavků pro formát SIGLE [HORÁKOVÁ; SKÁLOVÁ, 1995].

Koordinačním centrem pro předávání záznamů do SIGLE se v roce 1995 stala Státní technická knihovna (STK). Oficiální název centra zněl Kooperační systém šedé literatury (KSŠL) a podle soudobé legislativy do něj spadaly české vysoké školy. Tvorby dat se neúčastnily všechny vysoké školy, a to z nejrůznějších důvodů ať už organizačních, legislativních, technických nebo technologických.

4.1.4.2. *Nekompatibilita formátů*

Velkým problémem byla pravidla a formát SIGLE pro ukládaná data, jež neumožňoval bezproblémovou konverzi dat o disertacích a katalogizačních formátů užívaných v České republice. Získaná data byla v dohodnutých dávkách transferována do výpočetního centra FIZ Karlsruhe přes FTP server Britské knihovny.

Po technologické stránce byly nejbližšími úkoly finalizace a odzkoušení technologie sběru dat. Byl dokončen překlad druhé úrovně tematických skupin SIGLE ve spolupráci s Knihovnou Akademie věd (KNAV) a výchozí soubor autorit korporací. Oba datové soubory

byly k dispozici producentům šedé literatury. Zavedení FIBRE (Friendly Input of Bibliographic Records) pro ukládání dat si vyžádalo určité změny původního datového formátu. K řešení otázky vhodného nástroje věcného popisu, který by v české databázi překlenul nedostatky tematických řad SIGLE, zejména nerovnoměrného pokrytí některých oborů, se STK a KNAV už nedostaly. Po stránce organizační bylo třeba dořešit i smluvní a finanční aspekty spolupráce mezi STK a univerzitami [BAYER, SKÁLOVÁ, 1995].

4.1.4.2.1. Možnosti řešení

H. A. Wesselse, ředitel EAGLE, v roce 1993 odpovídal na otázky redaktorky časopisu Infocus [WESSELS, HANKOVÁ, 1995]. Mimo jiné tam uvedl, že vedení EAGLE reálně uvažovalo o možnosti snižovat poplatky za členství v EAGLE (viz kap. 4.1.1.2.), až do té míry, že od roku 1997 by mohly být členské poplatky zrušeny. Chtěl v rámci nové ekonomické strategie zvažovat možnost oceňovat každý vstupní záznam, který členská organizace do systému vloží, určitou hodnotou a na základě toho pak odvíjet celý ekonomický mechanismus účasti. Hlavním záměrem bylo zjednodušit připojení dalších členů a zajistit co nejlepší ekonomickou samostatnost a existenci SIGLE. Jak dnes už víme, tato strategie úspěšná nebyla.

4.1.4.3. Státní technická knihovna koordinátorem pro SIGLE

Kromě STK a KNAV byla jako partner kontaktována i Národní knihovna ČR pro oblast konferenční, vládní, úřední (governmental) literatury. Bohužel ze strany NK ČR bez odezvy. Po roce 1999 se v souvislosti s platností nového vysokoškolské zákona [ČESKO, 2005] Státní technická knihovna stala, na základě smlouvy s Národní knihovnou ČR, pracovištěm pro celostátní zpracování disertačních prací. Stala se hlavní organizací v ČR spolupracující se systémem SIGLE a na jejím serveru byla vystavována i celková národní databáze SIGLE .

4.1.4.4. Nové aktivity ve Státní technické knihovně

Po ukončení funkce systému SIGLE bylo třeba uvažovat o dalším postupu spojeném s evidencí šedé literatury v ČR. Plánovaný projekt STK s názvem Digitální knihovna pro šedou literaturu navazuje na existující projekty ve vyspělých státech EU a zároveň respektuje českou Koncepti trvalého uchování knihovních sbírek tradičních a elektronických dokumentů do roku 2010 zpracovanou Národní knihovnou ČR [NÁRODNÍ KNIHOVNA ČR, c2007]. V návaznosti na průběžné výstupy projektu budou v průběhu let 2008-2011 připraveny Národní úložiště šedé literatury a Národní registr vysokoškolských kvalifikačních prací.

Specifickými cíli projektu jsou:

- Zajistit aktuální i dlouhodobou integraci, ochranu a uchovávání šedé literatury.
- Vybudování platformy pro spolupráci knihoven včetně Státní technické knihovny, archivů, akademických institucí (vysokoškolská pracoviště, Akademie věd ČR aj.) a dalších subjektů, které publikují a využívají šedou literaturu.
- Příprava metadatového standardu šedé literatury.
- Příprava standardů výměny metadat a plných textů mezi lokálními úložišti a národními registry.
- Příprava metodických pokynů pro sběr a zpřístupnění šedé literatury s ohledem na zákonnou úpravu ochrany autorských práv.
- Doporučení metod, technologií a nástrojů pro optimalizaci a zvýšení výkonnosti digitálního zpracování sbírek šedé literatury.
- Zajistit jednotnou dostupnost vědeckých zdrojů prostřednictvím centralizovaných národních repozitářů.
- Doporučení pro uživatelské rozhraní národních registrů řešící vyhledávání v metadatech a plných textech šedé literatury a vysokoškolských kvalifikačních prací (VŠKP).
- Příprava modelového národního úložiště za účelem testování a verifikace výše uvedených doporučení, technologií a nástrojů. Modelová aplikace bude sloužit pro následnou přípravu plnohodnotných národních úložišť šedé literatury a VŠKP [Státní technická knihovna, 2007]

Díky spolupráci s Odbornou komisí pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací při Asociaci knihoven vysokých škol ČR (<http://www.evskp.cz>), ve které řešitelský tým dlouhodobě spolupracuje na odborných aktivitách Komise souvisejících s projektem, je zaručena několikaletá odborná příprava projektu, jeho proveditelnost a praktická využitelnost výsledků v praxi. S Knihovnou Akademie věd ČR a s Ústavem informačních studií a knihovnictví FF UK je navázána spolupráce odborného charakteru tak, aby bylo zaručeno vyřešení existujících problémů v oblasti řešeného výzkumného směru.

4.1.5. Technické údaje o SIGLE

V listopadu 2003 bylo v SIGLE uloženo 833 000 záznamů a roční přírůstek činil asi 60 000 záznamů za rok. Aktualizace tehdy probíhala v měsíčních intervalech. Bohužel systém zanikl na konci roku 2005 (viz kap. 4.1.3.1 a 4.1.3.2.) [SIGLE, 2005].

4.1.5.1. Typy registrovaných dokumentů

- technické nebo výzkumné zprávy
- disertace
- materiály z konferencí a z jednání
- číselná data
- překlady odborné literatury⁶
- technická doporučení a normy
- popis počítačových programů
- bibliografie

4.1.5.2. Geografický záběr

Záznamy nově vytvořené šedé literatury byli ve formátu SIGLE připravováni spolupracujícími národními institucemi (viz Tab. 1). Zaslaly je do střediska pro zpracování dat, které mělo za úkol sloučit jednotlivé národní vstupy do systému a připravit strojem čitelný výstup na matriční magnetickou pásku SIGLE. Po dlouhá léta byla zpracovatelským centrem italská organizace ITALSIEL⁷, která se zabývala informačními službami. Zpracovávala veškeré vstupy, které pak transferovala směrem k databázovým centřům, jež data zpřístupňovala, a dříve také směrem k firmě SilverPlatter, která připravovala výstup na CD-ROM (v roce 2001 některé služby převzalo a provozovalo centrum OVID Technologies). Protože společnost ITALSIEL spotřebovávala zhruba 50 % veškerých financí systému, rozhodlo se vedení systému v roce 1998 změnit centrálního zpracovatele na FIZ Karlsruhe.

Tab. 1: Záznamy SIGLE z pohledu geografického k roku 2003 [BRATKOVÁ, 2003]

Země	Procentuální zastoupení záznamů z jednotlivých zemí
Velká Británie	50,4 %
Německo	25,8 %
Francie	10,1 %
Nizozemí	5,5 %
Itálie	2,6 %
Rusko	2,5 %
Komise EU	1,3 %

⁶ Překlady přestaly být registrovány v roce 1997, pro jejich vyhledávání je k dispozici báze UNESCO „World Translation Index“

⁷ Patřila do společnosti Grupo IRI Finisiel Řím.

Španělsko	0,9 %
Belgie	0,8 %
Ostatní	0,1 %

4.1.5.3. **Obsahový záběr**

V tabulce 2 je zobrazeno obsahové zaměření záznamů, které byly do systému SIGLE zasílány z celé Evropy. Další tabulka číslo 3 vyčísluje počet záznamů, kterým byla tématicky přiřazena jednotlivá hesla z pořádacího systému informací SIGLE.

Tab. 2: Záznamy SIGLE z pohledu obsahového k roku 2003 [BRATKOVÁ, 2003]

Obor	Počet záznamů v procentech
Sociální /humanitní vědy	40 %
Biologie a medicína	12 %
Fyzika	9 %
Elektronika	6 %
Vědy o zemi	5 %
Strojírenství	5 %
Matematika	4 %
Energetika	4 %
Chemie	4 %
Zemědělství	3 %
Životní prostředí	3 %
Materiály	3 %
Jiné	2 %

Tab. 3: Počet záznamů přiřazených jednotlivým předmětovým kategoriím [SIGLE, 2005]

Kategorie	Počet záznamů
Aeronautika	4 158

Zemědělství, rostlinstvo a zvěrolékařství	17 044
Znečištění životního prostředí, ochrana a kontrola	17 668
Humanitní vědy, psychologie a sociální vědy	6 657
Biologie a lékařství	81 269
Chemie	25 089
Vědy o zemi a atmosféře	30 597
Elektronika a elektrotechnika, výpočetní technika	39 479
Energie a síla	24 407
Materiály	19 016
Matematické vědy	22 775
Strojírenství, stavebnictví, řízení výroby a lodní stavitelství	47 563
Výrobní postupy a zařízení	4 342
Vojenství	507
Raketová technologie	69
Navigace, spojení, detekce a protiopatření	4 418
Arzenály	300
Fyzika	57 823
Pohon a paliva	3 222
Kosmické technologie	1 737

4.1.6. Bibliografický záznam

4.1.6.1. Pravidla

Pravidla pro bibliografický popis dokumentů v SIGLE vycházejí původně z popisu a výměnného formátu Energy Technology Data Exchange (ETDE) Mezinárodního

informačního systému pro jadernou energetiku (International Nuclear Information System, INIS). Původní verze pravidel formátu SIGLE pro šedou literaturu z roku 1991 byla v roce 1997 nahrazena verzí novou (4. vyd.), která znamenala uvolnění dříve velmi striktních pravidel u několika údajů, např. jak zapisovat názvy korporací. Specifikace formátu záznamů je dána sadou manuálů SIGLE 1-5.

4.1.6.2. Identifikační popis

Každý bibliografický záznam obsahoval základní kontrolní identifikační údaje, stručný bibliografický popis, předmětové kategorie a klíčová slova a také anotaci, která ale nebyla povinná. Záznamy bylo možné na přání knihovny vytvářet také přímo ve formátu SIGLE pomocí speciálního vstupního editoru FIBRE (Friendly Input of Bibliographic Records), který byl na smluvním základě dodáván z Mezinárodní agentury pro atomovou energii (International Atomic Energy Agency, IAEA). Jeho distributorem byl FIZ Karlsruhe. Přípravovala se také verze pro Windows.

Systém SIGLE používal svůj vlastní formát, který se bohužel neshoduje ani se standardy UNIMARC ani s MARC 21.

4.1.6.3. Věcný popis

4.1.6.3.1. Systém pro věcné pořádkání informací SIGLE

Báze dat SIGLE byla polytematická prakticky v nejširším smyslu tohoto označení. Pro předmětové pořádkání byl využíván kódový systém, který byl znám z oborového třídění grantových programů, a byl tedy běžně dostupný. Struktura vychází z US COSATI (Federal Council for Science and Technology, Committee on Scientific and Technical Information) Subject Category List, který je souhrnou verzí Tezauru technických a vědeckých termínů (Thesaurus of Engineering and Scientific Term) [ENGINEERS JOINT COUNCIL, 1969]. Pořádkací systém SIGLE má analogickou strukturu. Na rozdíl od systémů používaných v databázích z oblasti vědy a techniky pokrýval nejen oblast vědecké činnosti, ale i vývoj technologií (např. v pořádkacím systému UNESCO je vývoji technologií věnováno pouze 19 % hesel, v ANNEX II⁸ 16 %) a i svým rozsahem (přes 300 000 záznamů) byla v Evropské unii nepoužívanějším systémem pro tuto oblast. Rovněž jednoduchou strukturou členění (250 oborů a podoborů) byl pro daný účel nejvhodnější. Počítalo se i s využitím podpůrných prostředků, např. slovníku klíčových slov. Při indexování se doporučovalo zařazení projektů do dvou až tří oborů zároveň [BAYER, 1995].

⁸ Soubor definovaných pojmů z oblasti ekonomiky a finančnictví.

4.1.6.3.2. **Struktura věcného pořadacího systému SIGLE**

Seznam předmětových kategorií SIGLE (SIGLE Subject Category List) tedy sestával ze 22 hlavních tříd⁹ definovaných na základě oborů (viz tab. 4), které se dále dělily do 250 předmětových podtříd. Každé heslo mělo rozsáhlý počet vysvětlujících poznámek za účelem maximální pomoci indexátorovi k nalezení nejvhodnějšího kódu pro každý dokument [SIGLE Manual. Part 1, 1997]. Volba oborů a rozdělení materiálu jsou značně nevyrovnané a často spíše kuriózní. Veškeré humanitní obory mají jednu kategorii, kam patří jak historie, filozofie, teologie a další, tak i management, ekonomika nebo sport, zatímco vojenství jsou věnovány kategorie čtyři. Nicméně vzhledem k celkovému rozsahu této báze nelze říci, že by to představovalo závažnější problém. Báze dat SIGLE byla tedy zdrojem informací spíše pro obory humanitní než přírodovědecké a technické. Ve srovnání např. chemie a jejích tehdejších přírůstků představovala SIGLE 0,3 % informací zaregistrovaných za stejné časové období v databázi Chemical Abstracts, přičemž řada odkazů se překrývalo. Jednotlivé kódy k heslům vypadaly např. takto:

- Aeronautika 01
- Všeobecný oddíl 010
- Aerodynamika 01A
- Aeronautika 01B

Tab. 4: Předmětová struktura Seznamu předmětových kategorií SIGLE

Aeronautika
Zemědělství, rostlinstvo a zvěrolékařství
Znečištění životního prostředí, ochrana a kontrola
Nepoužito
Humanitní vědy, psychologie a sociální vědy
Biologie a lékařství
Chemie
Vědy o zemi a atmosféře
Elektronika a elektrotechnika, výpočetní technika
Energie a síla
Materiály
Matematické vědy

⁹ Dvě nejsou použity.

Strojírenství, stavebnictví, řízení výroby a lodní stavitelství
Výrobní postupy a zařízení
Vojenství
Raketová technologie
Navigace, spojení, detekce a protiopatření
Nepoužito
Arzenály
Fyzika
Pohon a paliva
Kosmické technologie

4.1.6.3.3. Konečné uplatnění věcného pořadacího systému SIGLE

Do konce roku 1994 se používalo k věcnému pořádání databází šedé literatury u nás dnes již zastaralého a nedostačujícího rubrikátoru MSVTI (Mezinárodní systém vědeckých a technických informací), od roku 1995 jej nahradily tematické skupiny SIGLE (modifikovaný tezaurus COSATI). Na tomto místě je třeba ocenit pomoc jednotlivých ústavů AV ČR při práci na překladu druhé úrovně tohoto pořadacího systému, která je zárukou kvalitního výběru českých ekvivalentů a eliminace sémantických chyb, vznikajících při mechanickém překladu. Bibliografický záznam práce měl dvě části, přičemž v té druhé byla autorovi nabídnuta možnost zveřejnit tuto práci i v systému SIGLE. Rozhodnul-li se autor pro zveřejnění, doplnil anglický překlad názvu a klíčová slova v angličtině. Toto rozhodnutí bylo plně v kompetenci autora, neboť se jednalo o výsledky práce, které mohly být např. předmětem utajení, patentového řízení apod. K elektronickému vstupnímu formuláři byl připojen program umožňující okamžitou kontrolu a případně opravu. Data byla poté konvertována a zpřístupněna v software CDS/ISIS [BAYER; SKÁLOVÁ, 1995].

4.1.6.3.4. Příklady záznamů ze SIGLE

První dva příklady jsou vybrány z databáze SIGLE. Jedná se o záznamy vědeckých zpráv. Z obou je patrné, že v poli deskriptorů (descriptors) jde pouze o dvoustupňové řazení. První heslo v příkladu 1 *Computer* patří do kategorie 09 – Elektronika a elektrotechnika, výpočetní technika a druhé heslo *software* spadá do podkategorie H – software. Obdobně je tomu u příkladu 2 jen podkategorie hesla je zde G – hardware. Hlubší členění pořadací systém neumožňoval. Třetí příklad je záznam diplomové práce vytvořený v STK. Naproti dvěma prvním je okamžitě zřejmé, že heslář PSH poskytuje mnohem detailnější věcný a obsahový

popis dokumentu. To je způsobeno jeho hlubším členěním a větším obsahem hesel, než měl pořádací systém SIGLE.

Tab. 5: Příklad 1

TI - TITLE:	A query-driven approach to executable specifications.
AU - AUTHOR(S):	Tuzhilin,-Alexander; Balasubramanian,-P.
CS - CORPORATE SOURCE:	Ecole Nationale Superieure des Telecommunications, 75 - Paris (FR). Dept. Informatique. (021061S)
DP - DATE OF PUBLICATION:	1994
PN - PAGINATION:	23 p.
SE - SERIES:	ENST. D (ISSN 0751-1345) 94-015
RN - REPORT NUMBER(S):	ENST-94-D-015
AV - AVAILABILITY:	Available at INIST (FR), Document Supply Service, under shelf-number : RP 12468
DT - DOCUMENT TYPE:	R-Report
DE - DESCRIPTORS: Primary:	09H-Computer-software
TA - TITLE AUGMENTATION:	temporal query language; loose coupling
CN - CONTROL NUMBER:	FR9402041
NC - NATIONAL CENTRE:	FR France. See Guide.
PY - PUBLICATION YEAR:	1994

Tab. 6: Příklad 2

TI - TITLE:	Information systems design. Software engineering workshops
OT - ORIGINAL TITLE:	Conception des systemes d'information. Les ateliers de genie logiciel
AU - AUTHOR(S):	Thiery,-Odile
CS - CORPORATE SOURCE:	Centre National de la Recherche Scientifique (CNRS), 54 - Vandoeuvre-les-Nancy (FR). Centre de Recherche en Informatique de Nancy (CRIN); Nancy-1 Univ., 54 - Vandoeuvre-les-Nancy (FR). Centre de Recherche en Informatique de Nancy (CRIN); Nancy-2 Univ., 54 -

	Vandoeuvre-les-Nancy (FR). Centre de Recherche en Informatique de Nancy (CRIN); Institut National Polytechnique de Lorraine (INPL), 54 - Vandoeuvre-les-Nancy (FR). Centre de Recherche en Informatique de Nancy (CRIN). (024333Z; 029942W; 024334A; 024336C)
DP - DATE OF PUBLICATION:	1993
PN - PAGINATION:	90 p.
RN - REPORT NUMBER(S):	CRIN-93-E-022
AV - AVAILABILITY:	Available at INIST (FR), Document Supply Service, under shelf-number : RP 12456
LA - LANGUAGE:	(In French)
DT - DOCUMENT TYPE:	R-Report
DE - DESCRIPTORS: Primary:	09G-Computer-hardware
CN - CONTROL NUMBER:	FR9401992
NC - NATIONAL CENTRE:	FR France. See Guide.
PY - PUBLICATION YEAR:	1993

Tab. 7: Příklad 3

001^CZ0303322
008^20K/I/M/U
100^Karlík, M.
110^České vysoké učení technické v Praze, Praha (CZ)
111^Thesis
230^Struktura a mechanické vlastnosti uspořádaných slitin na bázi aluminidu železa Fe ₃ Al.
401^Praha (CZ)
402^České vysoké učení technické v Praze
403^2002
500^45 s..
600^(CS)
611^Státní technická knihovna Praha
901^PSH: slitiny; feroslitiny; mechanické vlastnosti

4.1.7. Nárůst záznamů v jednotlivých databázích šedé literatury

Časové rozložení dokumentů a růst báze v letech 1980 – 1994 komentoval Jaroslav Šilhánek. [ŠILHÁNEK; NOVOTNÁ, 1996] citováno v [BALÍKOVÁ, 2001]. Největší počet dokumentů má letopočet 1990 a od tohoto roku pak počty v bázi zaznamenaných citací trvale klesaly. Pokud docházelo k větším zpožděním v zápisu dokumentů do systému, byla vážně ohrožena informační hodnota databáze. Provozovatelé systému tehdy neřešili bohužel otázku, u jakého typu dokumentů se pokles nejrazantněji projevil. Lze to např. posoudit podle disertací, jež představovaly prakticky třetinu celého objemu dat v bázi SIGLE. Sekundární zpracování disertací je možné považovat téměř za stoprocentní v Británii. Jedná se o produkci v národním jazyce, takže po nalezení vhodného dokumentu to znamenalo další starosti s jeho překladem. Tento fakt nebyl pochopitelně příliš motivující. Značně překvapivé bylo zjištění, že počet záznamů o disertacích z německých univerzit v bázi SIGLE představoval jen 6 % počtu disertací zaregistrovaných v německé národní bibliografii v bázi BIBLIODATA. Stejně jednoduchým způsobem nelze posoudit zastoupení disertací z francouzských univerzit, ale pokud je jejich celkový počet jen o málo větší než německých, není pochyb o tom, že ani v tomto případě nepokrývala báze skutečnou produkci. Zajímavý je i pohled z oborového hlediska, kde porovnáním počtu záznamů o disertacích v bázi dat SIGLE a v Chemických abstraktech (Chemical Abstracts) zjistíme, že první obsahuje jen 12,5 % počtu disertací uvedených za stejné časové období v druhé zmíněné bázi. Z toho všeho je zřejmé, že i když disertace představovaly třetinu všech záznamů v SIGLE, nelze v žádném případě předpokládat, že pokrývala evropskou produkci tohoto druhu literatury vyjma Velké Británie a pravděpodobně Nizozemí [ŠILHÁNEK, NOVOTNÁ, 1996].

4.1.8. Závěr

Systém pro registraci šedé literatury SIGLE zanikl koncem roku 2005. Kromě problematického financování patřilo k hlavním důvodům přílišné rozšíření typologie zpracovávaných dokumentů. Rozvoj tímto směrem se ukázal být kontraproduktivní. Zánik způsobil komplikace v zemích, kde nebyla dříve zavedena registrace šedé literatury v jednotné databázi. Bylo tomu tak i v České republice.

Je možné konstatovat, že přes nepochybně sympatický pokus o vytvoření celoevropské báze dat šedé literatury se předpokládané záměry nepodařilo úplně realizovat. Nelze ale přehlížet skutečnost, že současná nabídka sekundárních informačních zdrojů výrazně převyšuje rozumné možnosti jejich využívání. Za připomenutí stojí fakt, že z více než 50 % se vlastně

jednalo o alternativní přístup ke katalogům Britské knihovny [ŠILHÁHEK; NOVOTNÁ, 1996]. Vznik databáze SIGLE byl od počátku motivován možností zařazení částí národních bibliografií mezi standardní nabídku databázových středisek pracujících na ryze komerční bázi. To sice na jedné straně otevíralo jistou možnost ekonomického zhodnocení fondů, ale na druhé straně se jednalo o typicky zprostředkované zpřístupnění předpokládající znalost hesla a účtovacího konta a realizované prakticky výhradně informačními specialisty nebo pracovníky knihoven. Širší odborná veřejnost tak s vlastní bází SIGLE prakticky nepracovala, pouze mohla dostat v té či oné formě výsledky rešerší. Ovšem nic nebránilo tomu, aby vhodné soubory záznamů dokumentů, které byly k dispozici v některém světovém jazyce, ať se jednalo o disertace, nebo asi nejspíše o různé konferenční materiály, byly do databáze dodávány. Roční produkce takových dokumentů byla tak malá, že jejich zpracování by nebyl problém.

V plánovaném projektu se např. nyní pokouší Státní technická knihovna, za pomoci Akademie věd ČR a Ústavu informačních studií a knihovnictví Filosofické fakulty, o tvorbu národní databáze na registraci vysokoškolských kvalifikačních prací.

4.2. Sít'ová knihovna diplomových a disertačních prací (NDLTD)

4.2.1. Úvod

Sít'ová digitální knihovna diplomových a disertačních prací (Networked Digital Library of Theses and Dissertations, NDLTD) je výrazem společné snahy univerzit z celého světa podpořit vznik, archivování, distribuci a přístup k elektronickým diplomovým a disertačním pracím (Electronic Theses and Dissertations, ETD). Systém elektronických diplomových a disertačních prací je provozován Virginským polytechnickým institutem a státní univerzitou (Virginia Polytechnic Institute and State University, <http://www.vt.edu>). Od zahájení činnosti systému v roce 1996 se k této iniciativě připojilo už více než sto univerzit, které zdůrazňují důležitost institucí jako míst, kde se studenti učí psát a publikovat závěrečné práce v elektronické podobě. Výzkumy vyvrcholily **tvorbou jednotné databáze**, která má inspirovat k **vyhledávání diplomových a disertačních prací** ze souborné databáze členských institucí NDLTD. Oficiálně se i na webovém sídle systému NDLTD objevuje příznačnější označení systému: „**federace**”. URL základního sídla NDLTD (<http://www.ndltd.org/>) je velmi často uváděno jako hyperodkaz na stránkách téměř všech prezentací univerzit [BRATKOVÁ, 2003].

4.2.2. Technické zázemí

Databázi zastřešuje systém **Virtua**, který vyvinula firma Visionary Technology in Library Solutions (VTLS). Ve snaze vytvořit most sloužící k překonání nedostatků mezi existujícími distribuovanými archivy institucí a spojení databází diplomových a disertačních prací, byl přijat **speciální metadatový standard** pro tento druh dokumentu. Je rovněž používán univerzitami zapojenými do systému NDLTD k exportu volně dostupných metadat těchto prací pomocí protokolu Open Archives Initiative- Protocol for Metadata Harvesting (**OAI-PMH**). Následný výzkum byl zaměřen na pokročilý vyhledávací mechanismus, sémantickou interoperabilitu, tvorbu a rozvoj vyhledávacích systémů schopných pracovat s více jazyky zároveň a softwarové moduly na podporu rozvoje služeb s vyšší přidanou hodnotou pro uživatele, kteří hledají relevantní elektronické diplomové a disertační práce.

Tyto snahy by mohly mít silný pozitivní vliv na rozvoj informovanosti o vysokoškolských kvalifikačních pracích na univerzitách po celém světě. Jednou z důležitých organizací propagujících rozšíření povědomí o této oblasti je United Nations Educational, Scientific and Cultural Organization (UNESCO), které v roce 2001 zpřístupnilo dokument s názvem **Mezinárodní průvodce pro tvorbu ETDs** [UNESCO, 2001] . Tento manuál i jeho webové sídlo byly zpřístupněny v různých jazycích, jejich cílem bylo pomoci studentům, vysokoškolským učitelům a administrativním pracovníkům podílejících se na zajištění provozu systému NDLTD.

4.2.3. Členové a rozvoj fondu elektronických diplomových a disertačních prací

Nejnovější výčet členů spolupracujících na knihovně NDLTD je možné si prohlédnout webové stránce NDLTD [NDLTD, 2007] . Členové jsou rozděleni do tří kategorií a jedná se pouze o členy, který participují na rozvoji NDLTD a platí roční příspěvky. První jsou **instituce**, kterých je nyní 74 (převážně se jedná o univerzity), další jsou tři **konsorciální členové**, jedná se konkrétně o Radu pro australské univerzitní knihovny (Council of Australian University Libraries, Australia), Síť knihoven Triangle Research (Triangle Research Libraries Network, USA) a Konsorcium akademických knihoven v Utahu (Utah Academic Library Consortium, USA). Poslední skupinu tvoří **fyzické osoby** a zde je zatím pouze jeden člen.

Statistika ze srpna 2001 ukazuje tehdy aktuální členy NDLTD. Za méně než dva a půl roku NDLTD více než zdvojnásobil počet svých registrovaných členů (z 59 v květnu 1999). Tedy v roce 2001 měla 120 členů, 52 univerzit u USA, 52 univerzit ze světa a 16 institucí, regionálních center a dalších organizací (např. UNESCO). Tito různí partneři reprezentují 23

zemí (Austrálii, Brazílii, Kanadu, Čínu, Kolumbii, Německo, Řecko, Hong-Kong, Indii, Itálii, Mexiko, Nizozemí, Norsko, Ruskou federaci, Singapur, Jižní Afriku, Jižní Koreu, Španělsko, Súdán, Švédsko, Taiwan, USA a Velkou Británií). Tito členové se také podílí na růstu globálního zájmu o systém NDLTD jako mezinárodní participace i menších registrovaných členů NDLTD, na nutnosti podrobit se povinným prvkům elektronických disertací. Tab. 8 zobrazuje přístupy do databáze elektronických diplomových prací a disertací na Technické univerzitě ve Virginii (Virginia Tech ETD Collection).

Počet elektronických diplomových a disertačních prací dostupných v systému NDLTD univerzit i institucí rostla vysokou rychlostí. Z několika desítek prací uložených na Technické univerzitě ve Virginii z roku 1996 na počet 4 328 prací z 21 institucí v březnu 2000 a v červenci 2001 bylo nakonec napočítáno konečných 7 268 diplomových a disertačních prací z 25 členských institucí. Tyto statistiky nebraly v úvahu **naskenované disertace**, jež tvořily základ celého systému NDLTD. Existuje 26 naskenovaných dokumentů v New Jersey v Technologickém institutu (New Jersey Institute of Technology), 150 na Univerzitě na Jižní Floridě (University of South Florida), 5 581 na MIT (Massachusetts Institute of Technology) a 12 000 v Národním dokumentačním centru v Řecku. Celkově se jedná o 17 763 naskenovaných disertací v těchto institucích a dost možná tisíce neohlášených v jiných institucích.

Tab. 8: Statistika přístupů do databáze elektronických diplomových prací a disertací na Technické univerzitě ve Virginii [HUSSEIN, 2001].

	1997/98	1998/99	nárůst 1997/98- 1998/99	1999/00	nárůst 1998/99- 1999/00
Požadavky na PDF (Requests for PDF files) (převážně úplné ETD - mostly full ETDs)	221,679	481,038	117.0%	578,152	20.2%
Požadavky na HTML (Requests for HTML files) (převážně obsahy a abstrakty - mostly tables of contents and abstracts)	165,710	215,539	30.1%	260,699	21.0%
Požadavky na multimédiální formát	1,714	4,468	160.7%	12,633	182.7%

(Requests for multimedia)					
Různé požadované soubory (Distinct files requested)	6,419	21,451	234.2%	16,409	-23.5%
Obslouženy různými hosty (Distinct hosts served)	29,816	57,901	94.2%	87,804	51.6%
Denní průměr přenesených dat (Average data transferred daily)	156,089 KMbB	219,132 KMbB	40.4%	382 MbMB	74.4%
Přenesené data (Data transferred)	55,637 MbGB	78,107 MbGB	40.4%	137 GbGB	75.6%

4.2.4. Standardy popisu

4.2.4.1. Úvod

Na podporu budoucího rozvoje byl rovněž specifikován metadatový standard Electronic Thesis and Dissertation Metadata Standard (ETDMS) a dále byl realizován projekt pro kontrolu jmenných autorit.

4.2.4.2. Metadata

Standard ETDMS byl vyvinut v souladu s pravidly systému NDLTD. Jeho specifikace trvala delší dobu, ale nakonec byl stanoven jednoduchý standard pro kódování textu, který má podobu definice typu dokumentu pro formát XML (Document Type Definition for XML, **DTD XML**). Metadatový standard pro elektronické diplomové a disertační práce je flexibilní soubor manuálů pro kódování a sdílení základních metadat sledujících tyto dokumenty mezi institucemi. Samostatné práce dále postupují paralelně podle vyhovující definice typu dokumentu, postaveného na společném rámci pro plný popis diplomových a disertačních prací. **Standard ETDMS** vychází ze základního standardu Dublin Core, ale obsahuje ještě dodatečné elementy specifické pro data disertací. Navzdory svému názvu, byl standard ETDMS navržen v souladu s metadaty pro tištěnou verzi disertací a dále tak, aby bylo možno pracovat s dokumenty v různých jazycích. Standard detailně převádí informace o elektronických diplomových a disertačních pracích do metadatových elementů. Je též podporován jako výstupní formát pro sklízení metadat prostřednictvím protokolu OAI-PMH.

4.2.4.3. Jmenné autority

Každý dokument popisovaný dle standardu ETDMS může obsahovat jméno, ať jméno osoby nebo název instituce, který se pak zaznamenává do příslušného pole formuláře podle

standardu. Proto bylo třeba vytvořit jakousi autoritní databázi, která by sdružovala jména osob a názvy institucí, za účelem sjednocení záznamu tohoto druhu informace. Návrh, který následkem toho vznikl v síťové knihovně NDLTD s názvem „The NDLTD: Authority Linking Proposal“ představil hned několik cílů k tvorbě systému pro jednotný zápis autorit, který se jmenuje „**Linked Authority File (LAF)**“. LAF by měl být volně tvořitelný a sdílený mezi jednotlivými členy. URI vedoucí ke konkrétním bibliografickým záznamům autorit v jednotné databázi by mělo být srozumitelné a užitečné komukoliv i mimo doménu systému NDLTD. Mělo by být trvalé a aktuální. Model LAF by měl být aplikovatelný kdekoliv spolu s otevřeným standardem a s otevřeným zdrojovým kódem softwaru.

Design LAF má další výhody navíc jako např. používání Jmenné autoritní databáze Kongresové knihovny (Library of Congress Name Authority Database). Pouze úroveň spolupráce mezi decentralizovanými účastníky limituje pokrytí kolekce. Protože záznamy jsou tvořeny na základě formátu XML, obsah LAF může být rozšiřován a zužován dle potřeby.

4.2.5. Věcné pořádní

Jako základní termíny pro věcné pořádní elektronických diplomových a disertačních prací si knihovna NDLTD zvolila **předmětová hesla**, která nevybírání z řízeného slovníku, ale **tvoří je volně**. Jazyk předmětových hesel je nástroj pořádní, ve kterém je téma dokumentu vyjádřeno lexikálními jednotkami podle předem daných syntagmatických a syntaktických pravidel už v průběhu indexování [BALÍKOVÁ, 2001].

4.2.5.1. Příklady záznamů a popis použitých předmětových hesel

Tab. 9: Příklad 1

001		vtls000010992
003		VRT
005		20060410200800.0
008		s2001 000 eng d
035		\a OCLCNo/ocm47872384 \z oai:xtcat.oclc.org:OCLCNo/ocm47872384
039	9	\a 200604102008 \b VLOAD \c 200405111937 \d VLOAD \c 200311041557 \d VLOAD \c 200311041355 \d VLOAD \y 200210091513 \z VLOAD
100	1	\a Ramsey, Rachel D.
245	1	\a "A mad intemperance ... of building"
246	1	\a the literary construction of early modern London /

260		\b Morgantown, W. Va. : \c 2001
502		\a () -- .
520		\a Includes bibliographical references (p. 244-265).
520		\a Title from document title page.
548		\a English
653		\a Urbanization
653		\a Urbanization in literature.
653		\a Architecture and literature.
653		\a architecture
653		\a Architecture
653		\a London (England)
653		\a London (England)
655		\a Electronic Thesis or Dissertation
856		\u http://etd.wvu.edu/templates/showETD.cfm?recnum=2059

Příklad 1 je záznam práce na téma architektura a literatura. Z názvu bychom bez použití předmětových hesel v žádném případě nezjistili, čím se práce skutečně zabývá. V tomto případě bylo přiděleno sedm předmětových hesel do duplikovaného pole 653, které je podle standardu MARC 21 [MARC 21, 2006] stanoveno k zápisu předmětového hesla. Jako předmět dokumentu byl zohledněno i místo, ke kterému se práce vztahuje. Dále byl zohledněn rozdíl ve velikosti počátečního písmene u hesel „architecture“ a „Architecture“. V poli 655, které je určeno na popis formy nebo fyzickou charakteristiku popisovaného dokumentu, je zdůrazněno, že se jedná o elektronickou diplomovou práci či disertaci.

Tab. 10: Příklad 2

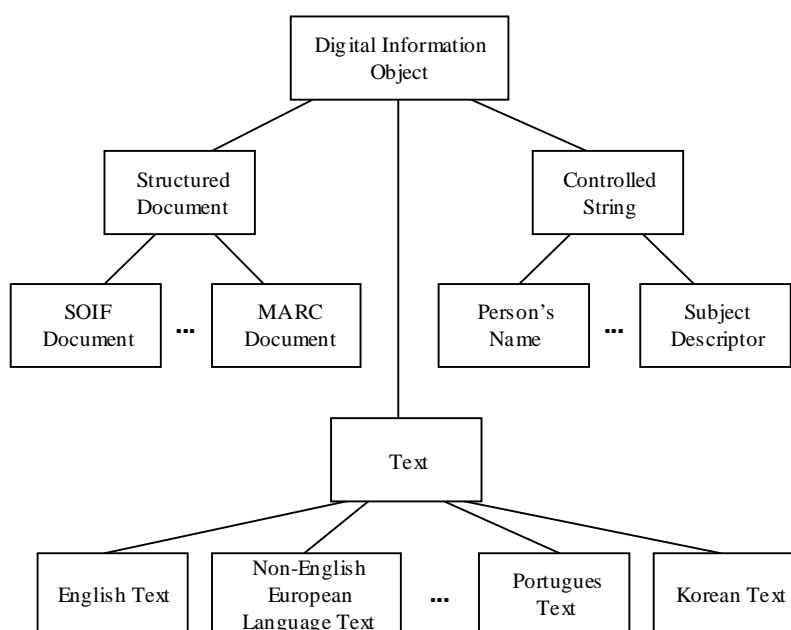
001		vtls000014858
003		VRT
005		20060410232200.0
008		s2001 000 eng d
035		\a uwaterloo.ca-113 \z oai:uwaterloo.ca:113
039	9	\a 200604102322 \b VLOAD \y 200311041315 \z VLOAD
100	1	\a Huang, Yiye
245	1	\a Determining Analytical Potential Energy Functions of Diatomic Molecules by Direct Fitting
260		\b University of Waterloo \c 2001

502		\a Master of Science () -- University of Waterloo.
520		\a The fully quantum mechanical 'direct-potential-fit' (DPF) method has become increasingly widely used in the reduction of diatomic spectra. The central problem of this method is the representation of the potential energy and Born-Oppenheimer breakdown (BOB) correction functions. There are a number of problems associated with the existing method and potential forms. This thesis delineates these problems and finds solutions to some of them. In particular, it is shown that use of a different expansion variable and a new treatment of some of the expansions resolves most of the problems. These techniques have been successfully tested on the ground electronic states of the coinage metal hydrides and the Rb ₂ molecule. To address the problem of representing 'barrier' potential curves, a flexible new functional form, the 'double-exponential long-range' (DELR) potential function, is introduced and applied to the B barrier state of Li ₂ . In addition, the Lambda-doubling level splitting which occurs for singlet Pi electronic states has been taken into account by extending the effective Schrodinger equation. The computer program DSPotFit developed in our laboratory for performing DPF analyses has been extended to incorporate the ability to fit the analytical potential energy functions to tunneling predissociation line widths for quasibound levels. Finally, an attempt is made to investigate whether there exists a hump in the ground state rotationless potential curve of beryllium hydride.
540		\a Copyright: 2001, Huang, Yiye. All rights reserved.
548		\a English
653		\a Chemistry
653		\a spectroscopy
653		\a Potential Energy Functions
653		\a Direct Potential Fit
653		\a Schrodinger Equation
653		\a Born-Oppenheimer Breakdown Correction Functions
653		\a Lambda-Doubling
655		\a Electronic Thesis or Dissertation
710		\b University of Waterloo \b Chemistry and Biochemistry
856		\q application/pdf \u http://etd.uwaterloo.ca/etd/y6huang2001.pdf

V příkladu 2 je popsána disertace z oboru chemie. Je zde vidět velmi pečlivý výběr předmětových hesel indexátorem, kterému pravděpodobně jako největší zdroj posloužil abstrakt obsažený v poli 520. Toto je velmi častá praxe, kdy indexátor vychází při výběru hesel z obsahu a abstraktu obsaženém v dokumentu. V poli 655 je opět uvedena specifikace charakteru práce.

4.2.5.2. Vyhledávání pomocí MARIAN

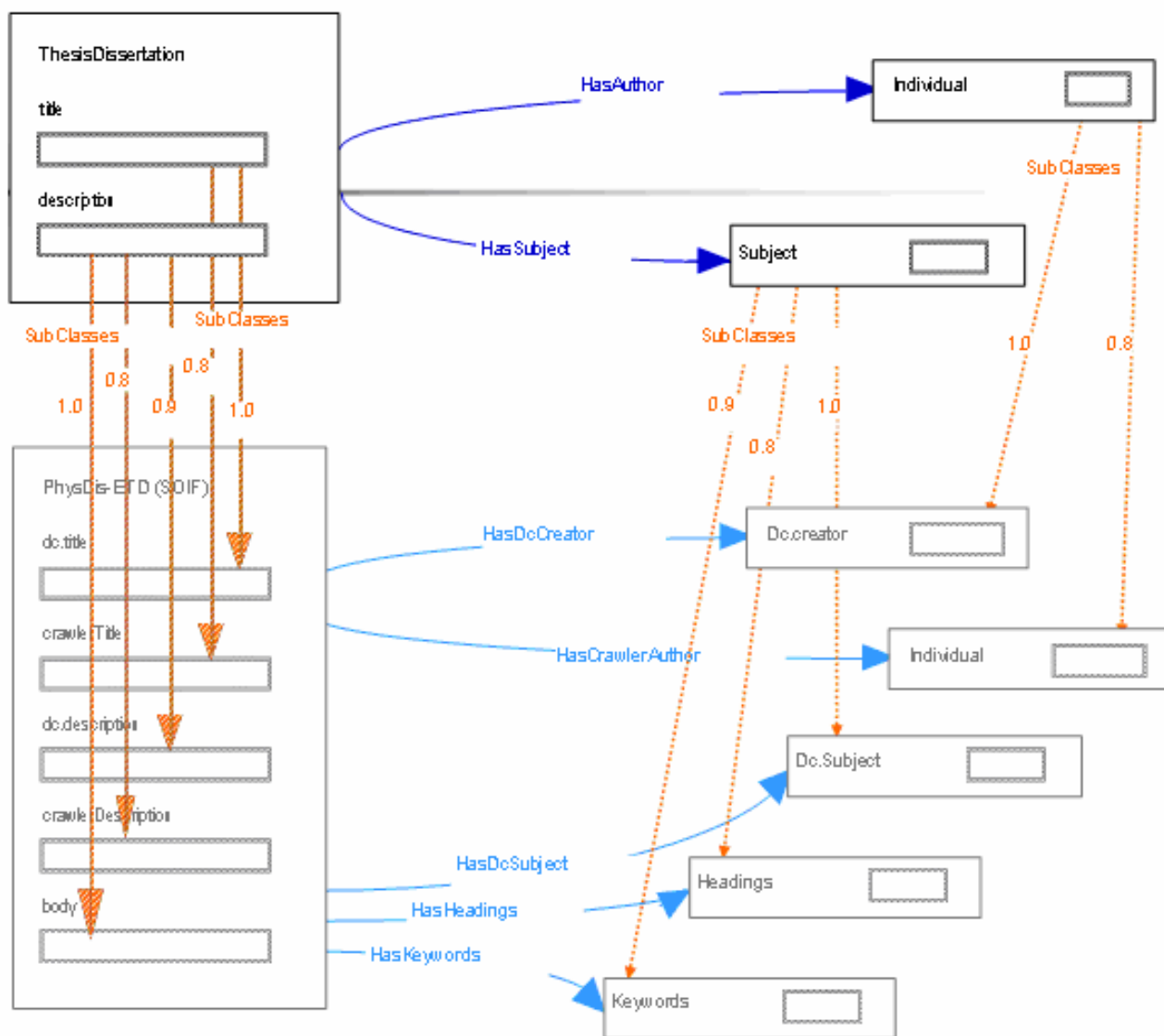
MARIAN patří mezi aktuální rozvíjené snahy stran vyhledávání v elektronických dokumentech. Jedná se o vyhledávací software vyvinutý v laboratořích Technické univerzity ve Virginii (Virginia Tech Digital Library Research Laboratory), který poskytuje globální služby a je kooperativní s mnoha ostatními digitálními systémy. MARIAN podporuje **flexibilní a rozšiřitelné hledání v sítích digitálních informačních objektů**, kteří zahrnují mimo jiné dokumenty či záznamy metadat pro jednotlivce i organizace. Digitální informační objekty a odkazy, které je propojují, jsou uspořádány pomocí předmětově-orientovaných tříd, každý podporuje indexování, vyhledávání a metody moderního prezentování dat (viz obr. 1). Nové třídy informačních objektů lze do MARIAN přidat pomocí manažera pro tvorbu tříd.



Obr. 1: Vyhledávání na základě rozdělení do tříd [GONCALVES, 2000].

V modelu informační sítě MARIANu byl zmapován standard ETDMS, tedy bylo jednoduše nahlédnuto na společnou kolekci zvenčí. Modul je prezentován na obr. 2. Obrázek představuje zdroje informací, dle kterých je možno vyhledávat, tedy název (title), autor (creator), předmět (subject) a popis (description). Model se skládá ze tří tříd objektů – diplomová práce/disertace (thesis/dissertation, jinak také body neboli text) osoby (individual) a předmětu (subject). Třída osoba zahrnuje dvě položky (DC creator a CrawlerAuthor) a sjednocuje vlastně všechny autory, kteří mohou být uvedeni v názvu práce, zatímco třída předmět pokrývá rozličné zpracování obsahu práce podle jejího popisu. Třída předmět v sobě obsahuje dc.subject, hesla a klíčová slova. Ze základních informací uvedených v elektronické diplomové nebo disertační práci jsou extrahována data pro zpracování bibliografického záznamu a ty jsou pak při

vyhledávání rovnou vážena a srovnávána na úrovni podtříd. Výsledky fulltextového prohledání v textu dokumentu jsou porovnávána s výsledky prohledání záznamu dokumentu. Ačkoliv jednoduchá konstrukce syntetických supertříd s váženými podtřídami nemůže zvládnout každou situaci, je tento postup shledáván jako velice efektivní.



Obr. 2: Postup vyhledávání v MARIAN [GONCALVES, 2000].

4.2.5.3. Multijazyčné vyhledávání

Má-li se jednat o mezinárodní federativní sdružení několika institucí, které v jednom společném katalogu zpřístupňují elektronické diplomové a disertační práce svých studentů, je nutné tzv. multijazyčné vyhledávání. V nejlepším případě by měl být **vyhledávač schopen akceptovat tolik jazyků**, jaké jsou užívány v kooperujících institucích. Jednotným spojovacím jazykem je samozřejmě angličtina. Momentálně se v systému ND LTD pracuje na **překladači dotazů**, nazývá se Text Class Manager. Výzkum v oblasti automatického

překladu je založen na práci Akira Maeda a NAIST Multilingual group [MAEDA, 1998], kde zkoumají pomocí statistik podobného výskytu slov váhu (míru) překladů a jejich kombinaci.

4.2.6. Budoucnost NDLTD

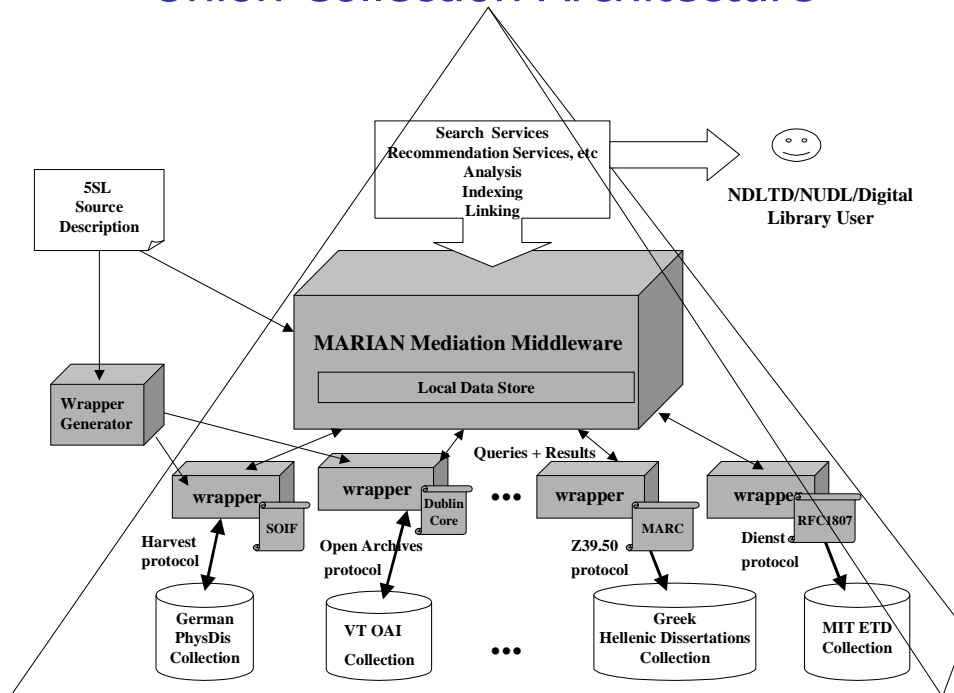
Statistiky ukazují, že produkce a archivování elektronických disertací a diplomových prací nabývá rychle na významu a stává se novou užitečnou součástí informačních zdrojů zpřístupňovaných na univerzitách v digitální době. Síťová knihovna NDLTD podporuje individuální instituce v zakládání jejich vlastních databází elektronických diplomových a disertačních prací. Stejně jako propagaci užívání těchto prací přes webová rozhraní a portály či agregátory individuálních stránek a vytváření jednoduchého vstupu do databáze NDLTD.

4.2.6.1. Rozvoj projektů

Pokračující výzkumy směřují k otázkám řešení problémů, jak spojit aktuálně rozdělené izolované kolekce hostující u různých členů kolekce. Dále se např. v NDLTD zabývají rozvojem projektů, jako je Souborný katalog (The Union Catalog), konstruovaný jako portál ve systému Virtua (viz obr. 3). V projektu souborného katalogu (The Union Catalog Project) bylo přistoupeno k řešení spojení kolekcí digitálních disertací všech spolupracujících institucí a vyhledávání v nich pomocí OAI-PMH protokolu¹⁰. Sdružuje metadata ve formátu EDTMS a zpřístupňuje je prostřednictvím centrálního portálu. Rozdělení mezi slučování kolekcí a poskytování služeb má výhodu v tom, že spojená kolekce může pracovat jako lokální uložiště, které může být prohledáno i jinými prohledávacími systémy, např. Virtua či MARIAN. Navíc vrstvená architektura podporuje více přirozenou integraci a slouží jako pozitivní důkaz správného konceptu pro členy knihovny NDLTD, kteří raději spolupracují jako federace, než aby samostatně pracovali na svém vlastním modelu zpřístupňování elektronických diplomových a disertačních prací.

¹⁰ Protokol OAI-PMH je dotazovací protokol umístěný nad HTTP a umožňuje jednomu serveru sklízet metadata z jiných počítačů. Běžně se tomuto procesu říká sklizení metadat neboli harvestování.

Union Collection Architecture



Obr. 3: Architektura sdruženého katalogu [GONCALVES, 2000].

4.2.7. Závěr

V roce 1999 podepsalo UMI s americkým Úřadem pro copyright dohodu, v rámci níž se UMI stalo oficiálním repozitářem pro ukládání digitálních disertačních a diplomových prací v rámci systému CORDS (Copyright Office electronic Registration, recordation & Deposit System). Záznamy i plné texty zpřístupňuje UMI také v moderním systému "ProQuest Digital Dissertations". I pod tlakem rozvíjejícího se systému NDLTD začalo UMI nabízet volně záznamy disertací a vybrané stránky z textu disertací předložených vždy v posledních dvou letech [BRATKOVÁ, 2003]. Po této krátké zmínce o činnosti UMI je zřejmé, že dosáhnout toho, aby univerzity v USA v co největším počtu změnilы tuto zavedenou tradici, začaly zveřejňovat disertační práce na svých vlastních serverech zdarma a pro celý svět a připojily se tak k federaci NDLTD, nebude, zdá se, vůbec jednoduché. A to i přes to, že existují zjevné přednosti takového systému, které jsou formulovány v jeho cílech a funkcích. Současná nejaktuálnější otázka je okolo tvorby repozitářů, kde se mohou elektronické diplomové a disertační práce skladovat a jejich přístupování na základě vyhledávání a to pokud možno v uživateli nejbližším jazyce.

4.3. Program australských digitálních disertací (ADT)

4.3.1. Úvod

Některé významné australské univerzity brzy následovaly americké univerzity a staly se nejen přímými aktivními účastníky systému NDLTD, ale jejich odborníci začali sami realizovat myšlenku vytvoření vlastního národního systému pro rozšiřování a zpřístupňování digitálních disertací.¹¹ **Národní souborný katalog KINETICA** vykazuje záznamy disertací jenom z některých fondů univerzitních knihoven, a to ne vždy v úplnosti [BRATKOVÁ, 2003]. Program Australské digitální disertace (Australian Digital Theses, ADT, <http://adt.caul.edu.au/>) je realizován pod dohledem **Komise pro politiku** (Policy Committee), která se mimo zástupců z Rady pro koordinaci australských univerzit (Council of Australian University Librarians, CAUL) skládala ještě z reprezentantů Národní knihovny Austrálie (National Library of Australia), firmy ProQuest, Rady australských postgraduálních asociací (Council of Australian Postgraduate Associations, CAPA) a děkanů a ředitelů absolventských škol (Frans and Directors of Graduate Schools, DDOGS).

Angažuje se také jako člen mezinárodního systému NDLTD (Networked Digital Library of Theses and Dissertations). Reprezentanti Programu ADT se v rámci NDLTD podílejí na práci správní komise (Steering Committee), komise pro standardy (Standards Committee), pro strategické plánování (Strategic Planning Committee) a propagační komise (Promotional Committee).

Architektura systému ADT je založena na distribuovaném ukládání digitálních disertací a metadat na serverech jednotlivých účastnických univerzit a na centrálním vyhledávacím systému zajišťovaném kapacitami serveru organizace CAUL včetně rozhraní pro vyhledávání informací (<http://www.caul.edu.au/>). Přičemž disertace ukládají do systému sami studenti – autoři za použití jednoduchých standardizovaných protokolů.

Program ADT má čtyři základní požadavky:

- vytvořit *standardy* pro tvorbu elektronických disertací, jejich ukládání a zpřístupnění,
- vytvořit elektronický *archiv* často požadovaných disertací,
- stanovit postupy pro *povinné vkládání* elektronických disertací přímo studenty,
- zajistit *kompatibilitu* s ostatními školními komunikačními iniciativami za použití pomocné komunikace, archivování a formátových protokolů [CARGNELUTTI, 2004].

¹¹ V Austrálii je pro doktorské disertace užíváno anglického termínu “thesis”, nikoliv „dissertation”.

4.3.2. Členové a rozvoj ADT

Program zakládalo sedm australských univerzit. V čele stojí:

- Knihovna Univerzity Nového Jižního Walesu (Sydney, <http://www.library.unsw.edu.au>) a dalšími členy jsou,
- Univerzitní knihovna Melbournské univerzity (<http://www.lib.unimelb.edu.au/>),
- Knihovna Queenslandské univerzity (Brisbane, <http://www.library.uq.edu.au/index.html>),
- Knihovna Univerzity v Sydney (<http://www.library.usyd.edu.au/Home.html>),
- Knihovna Australské národní univerzity (Canberra, <http://anulib.anu.edu.au/>),
- Knihovna Curtinovy technické univerzity (Perth, <http://lisweb.curtin.edu.au/>) a
- Knihovna Griffithovy univerzity (Brisbane-Gold Coast, <http://www.gu.edu.au/ins/>).

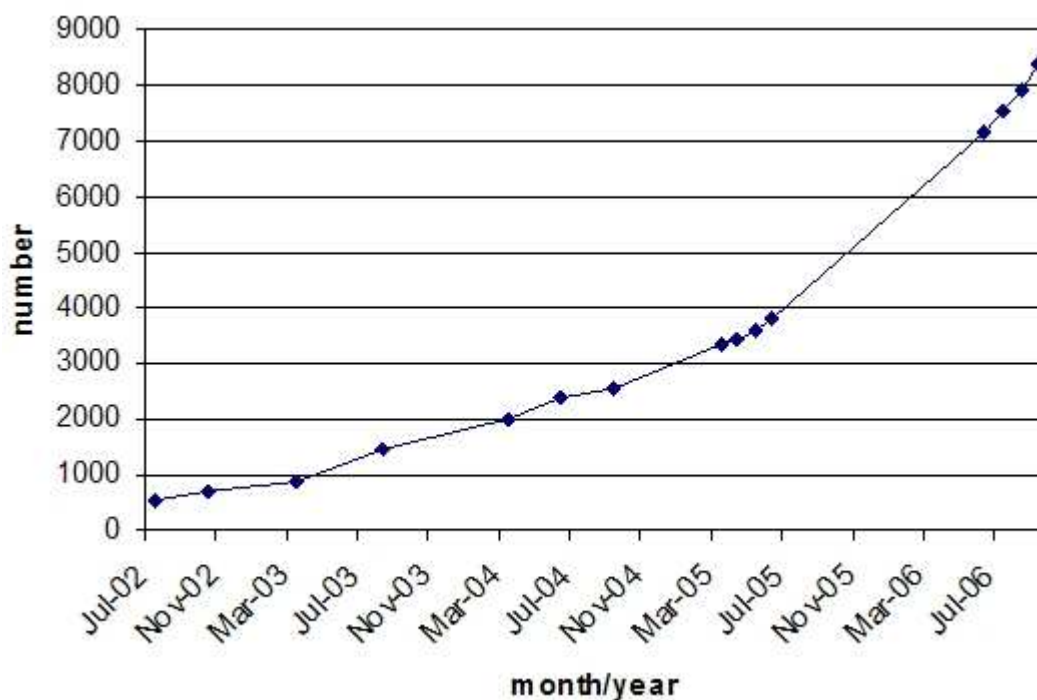
Program byl **iniciován v roce 1998-1999** a k oficiálnímu zahájení došlo v roce 2000. Jeho cílem bylo založit databázi digitálních disertací produkovaných studenty a jejich zpřístupňování prostřednictvím World Wide Web nebo Internetu v souladu se záměrem zveřejňovat výsledky australských výzkumů mezinárodní vědecké komunitě. Původní myšlenky pomohly k vytvoření systému, který by se dal specifikovat jako samostatný, distribuovaný a kolaborativní. Software je vytvořen tak, aby jej bylo možné snadno integrovat do webovských sídlech členských organizací. Po nainstalování mohou studenti pomocí webového rozhraní sami disertace vkládat nebo požádat knihovnu o pomoc. Disertace jsou shromažďovány na lokálním serveru knihovny. **Centrální repozitář metadat ADT** je prohledatelný a je tvořen automaticky, tj. za pomoci stroje, a to z úplného metadatového záznamu dle Dublinského jádra (Dublin Core, DC), který vzniká vyplněním daného formuláře [UNESCO, 2001].

V roce 2002 souhlasila **Rada knihovníků australských univerzit (CAUL)**, jež byla koordinátorem projektu, s přeměnou tohoto projektu na dlouhodobě udržitelný program. Všech 38 australských knihoven se do programu připojilo. V roce 2004 projevila zájem o zapojení do projektu ADT **Rada univerzitních knihoven Nového Zélandu** (Council of New Zealand Univerzity Librarians, CONZUL). Jako první byla přijata Univerzity of Canterbury in Christchurch. Aktivita členů neustále vzrůstají, dvacet pět z třiceti osmi členů aktivně pracují na rozšíření obsahu programu. Všechny univerzity používají stejný formát metadat kvůli zajištění vzájemné kompatibility. Běžný formát pro disertace je PDF (Portable Document Format), ten také zaručuje, že data jsou nezávislá na platformě, na níž byla vytvořena, a umožňuje v případě potřeby snadný tisk.

4.3.3. Analýza systému

V současné době má program ADT 39 aktivních členů. Podíl aktivních členů v organizaci CAUL je 82 % a z Rady CONZUL je to 75 %. Program ADT obsahoval na konci září 2006 136 435 záznamů, 10 963 odkazů na digitální verzi disertací z nichž 8 360 bylo vloženo přímo do systému ADT aktivními členy a 2 603 přes Národní bibliografickou databázi (National Bibliographic Database). Každý měsíc členové vkládají okolo 400 záznamů.

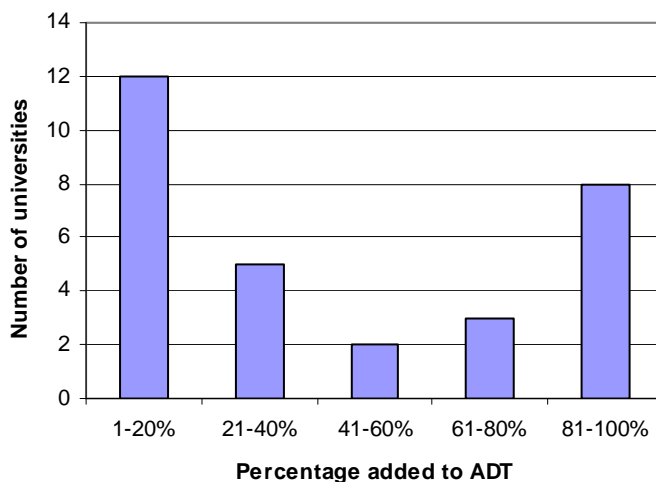
Graf 3: Záznamy získané přímo od členů [RUTHVEN, 2006]



Klíčovou roli hraje fakt, zda je ukládání elektronické disertace pro studenty **povinné nebo nepovinné**. Má-li univerzita povinně stanovené ukládat disertaci v elektronické podobě, pak je do databáze ADT uloženo 95 % obhájených prací. V případě, že tato povinnost není stanovena, dostane se do databáze ADT v průměru 17 % prací.

Graf 3 ukazuje procentuální zastoupení aktivních respondentů v roce 2005. V rozmezí 81 - 100 % se vyskytují studenti univerzit, které jsou zastoupeny v pravém sloupci grafu, předkládají disertace v elektronické podobě povinně.

Graf 4: Počet univerzit podle procentuálního zastoupení [ADT, 2006].



Trend používání digitálních repozitářů se rozšiřuje velmi rychle. Program ADT tento trend rovněž následuje a pomáhá i ostatním univerzitám se zřizováním digitálních repozitářů pro všechny druhy digitálních dokumentů. Podle průzkumu, který byl proveden pro Radu knihovníků australských univerzit (CAUL) je zřejmé že víc než polovina dotázaných univerzit digitální repozitář má, ale pouze čtyři z nich jeji používají pro ukládání disertací [ADT, 2006].

4.3.4. Technické zázemí

Software a základní myšlenky byly **převzaty z projektu NDLTD** rozvíjeném na Technické univerzitě ve Virginii (Virginia Tech). Největší rozdíl mezi softwarem u Virginia Tech a systémem ADT je ve vývoji opravdu distribuovaného národního modelu, který používá běžné nástroje a konvence. K hlavním požadavkům, které měli být řešeny při vývoji modelu ADT patřily:

- Spolupráce na projektu webového formuláře použitelného pro vkládání disertací a software programu, který za ním stojí, aby byl dostatečně flexibilní a funkční pro 40 institucí. Software pro vkládání disertací měl být také schopen rozlišit různé druhy přístupů (úplný, částečný a omezený).
- Dostatečná flexibilita modelu z hlediska možností popisu elektronické i tištěné disertace a dodatečného vložení starších prací.
- Automatické generování metadata na základě vložených specifických dat o disertaci.
- Vytvoření jednotné distribuované databáze zastřešující všechny australské univerzity.

- Stanovení podmínek elektronického dodávání plných textů disertací a s tím souvisejících poplatků.
- Vyvinutí standardů a procesů k udržení kroku se světovými trendy v oblasti zpracování a zpřístupňování elektronických disertací.
- Spojení s Národní australskou knihovnou (The National Library of Australia) a společná práce na řešení společných problematických otázek (např. se jedná o problematiku metadat). [CARGNELUTTI, 2004].

4.3.5. Standardy popisu

4.3.5.1. Úvod

V rámci systému byly připraveny a do praxe zavedeny závazné standardy. Formátem odevzdávaných digitálních a i v retrospektivě digitalizovaných disertací je zatím **formát PDF**, který má podle australských odborníků řadu výhod (snadný pro konvertování, programy pro čtení souborů jsou zdarma k dispozici, zajišťuje vysokou úroveň bezpečnosti dokumentu, je vhodný i pro účely archivace). Následné konverze do **formátu XML** jsou zatím zvažovány. Příslušná pracovní skupina se zabývá i návrhem vlastní definice typu dokumentu (Document Type Definition, DTD) pro elektronické disertace. Na univerzitách je pro doktorandy zajištěna podpora a technická pomoc. Jsou přesně zadány pokyny systému ke strukturaci digitálního dokumentu i pro pojmenování dílčích souborů (vyžívány jsou písmenné i číselné znaky) deponované disertace. Jako povinný se on-line zasílá soubor s pojmenováním **“01front.pdf”** (zahrnuje názvové a autorské údaje, abstrakt, poděkování, obsah, úvod, předmluvu a případně jakýkoliv úvodní text, který není součástí těla dokumentu). Hlavní text práce má pojmenování **“02whole.pdf”** (jde-li o celý dokument). Pokud je elektronický text členěn do více částí, předávají se soubory se jmény **“02chapter1.pdf”, “03chapter2.pdf”, “04appendix.pdf”, “05bibliography.pdf”** apod. Pro potřeby doktorandů je pro vytváření metadat k dispozici pohodlný interaktivní formulář k zápisu a generování metadat. Byly stanoveny přesné pokyny pro zápis a kódování znaků, zejména pro případ zápisu symbolů a diakritiky [BRATKOVÁ, 2003].

4.3.5.2. Metadata

Systém ADT přistoupil na jednoduché metadata podle **standardu Dublin Core**, automaticky seskupené a tvořící metadatový repozitář neboli databázi ADT, do níž přispívají členské instituce. Metadatový vyhledávač kompatibilní s vyhledávacími stroji (search engines) byl vytvořen v Austrálii v Centru pro technologii distribuovaných systém (Distributed Systems Technology Centre, DSTC) a nazývá se **HotMeta** [CARGNELUTTI, 2004]. Centrum pro

technologii distribuovaných systémů (DSTC), rovněž producent softwaru **Metasuite**, který provozoval centrální servis, přestalo fungovat k 30. červnu 2006. Již před svým ukončením podle ujednání (a v souvislosti s očekávaným přesunem správcovských práv) dodalo systému ADT kopie zdrojového kódu. Nezávisle na tom systém ADT uvažoval, že by mělo dojít ke změně zdrojového kódu pro výrobní servis systému v období programu Metasuite.

Zdrojový kód byl nahrán a testován na strojích v prostředí Univerzity Nového jižního Walesu (University of New South Wales). Rozšíření softwaru bude vyžadovat zásadní investici času a zdrojů, aby byl kompatibilní s kódem, tím bude zajištěno, že veškeré změny budou úspěšné. Software měl limitovanou dokumentaci. Je ale možné kontaktovat zaměstnance a vývojové pracovníky centra DSTC, kteří na kódu pracovali. Zpráva o dalším vývoji měla být vydána v polovině roku 2007 včetně platformy softwaru. Na webových stránkách ADT bohužel zatím není.

Disertace od řady členů jsou přidávány do systému ADT z jejich lokálních repozitářů použitím protokolu **OAI-PMH** (Open Archives Initiative – Protocol for Metadata Harvesting). Každé webové sídlo při přechodu na sběr z lokálního repozitáře požaduje testovací fázi, než se přepne ze softwaru VT-ETD. Testy jsou zaměřeny na úspěšně sesbírané všechny záznamy, sesbírání pouze updatovaných záznamů, kontrolu, zda soubor vyzvednutých záznamů obsahuje pouze disertace, zda metadatové prvky jsou platné, zda v záznamu nejsou neplatné znaky a zahrnuje noční sběrný program. Některá webová sídla se nesetkala s problémy ani v jednom kroku, jiné mnohokrát selhala v jednom z kroků vyžadující znovu otestování [RUTHVEN, 2006].

Protokol OAI-PMH je přístupný a schopný zpřístupnit metadata disertací vyhledávacím strojům (search engines), mezi které patří např. Google.

V roce 2004 získal program ADT dotace ze fondu Strategické informační infrastruktury od Oddělení pro vzdělávací vědy a praxi (Department of Education Science and Training, DEST) na **retrospektivní zpracování metadat** pro všechny australské disertace, jejich záznamy jsou dostupné, ačkoliv jejich plné texty nejsou k dispozici on-line [LAFFERTY; EDWARDS; DOVEY, 2005].

4.3.5.3. Jmenný popis

Metadatový formát záznamu obsahuje rovněž pole pro národní jmenné autority. V případě kódů a zkratk jsou autority vybírány z **Národního souborného katalogu** (National Union Catalogue). Jedná-li se o názvy institucí, vybírají se z **Australského mezinárodního souboru sdílených zdrojů** (Australian Interlibrary Resource Sharing Directory, ILRS). Knihovny z Nového Zélandu používají svůj vlastní soubor jmenných autorit, který nese název **Národní**

knihovna zkratek knihoven Nového Zélandu (National Library of New Zealand Library Symbols) [ADT, 2000].

4.3.5.4. Věcný popis

V záznamu elektronické disertace, který se pro jejich ukládání tvoří pomocí metadatového standardu, je zahrnuto i opakovatelné pole pro předmětová hesla. Ačkoliv se toto pole jmenuje keywords (klíčová slova), obsahuje jak **klíčová slova** tak i předmětová hesla. V konkrétním záznamu ze systému ADT je pak pole, které je dle metadatového standardu pojmenováno keywords, označeno popiskou předmětové heslo (subject words), jak je vidět na příkladech 1 a 2. Předmětová hesla jsou doplňována pouze odbornými indexátory, kteří je vybírají z Předmětového hesláře Kongresové knihovny (Library of Congress Subject Headings, **LCSH**). Autor sám nebo i knihovník může do záznamu přidat libovolný počet hesel a klíčových slov. Záznam obsahuje rovněž abstrakt, který může sloužit jako pomůcka při indexování. [ADT, 2000].

Tab. 11: Příklad 1 Záznam č. **132900**

<i>Title</i>	Comparative life cycle energy studies of typical Australian suburban dwellings
<i>Identifier</i> [URI]	http://eprints.unimelb.edu.au/archive/00000382/
<i>Author</i>	Fay, Mark Roger
<i>Autor</i> personalName	Fay, Mark Roger
<i>Subject</i>	<i>words</i> domestic architecture • energy conservation in buildings • architecture and energy conservation • Australian architecture
<i>Abstrakt</i> abstract	Cites, and the buildings of which they are comprised, consume a large proportion of the total energy produced within developed countries such as Australia. Much of this energy, particularly in Australia, is derived from fossil fuels and its consumption results in the emission of greenhouse gases that contribute to an enhanced greenhouse effect causing global warming. The need to reduce the energy consumed by residential and commercial buildings is now widely recognised. This has been acknowledged by state and federal governments within Australia and has resulted in strategies intended to increase the efficiency of building construction and operation...
<i>School/Department</i>	
	University of Melbourne, Faculty of Architecture, Building and Planning

<i>Date</i>	[W3CDTF]	1999
	valid [W3CDTF]	1999
<i>Rights</i>	http://www.utas.edu.au/copyright/copyright_disclaimers.html	
	(c) Copyright 1999 Mark Roger Fay	
<i>Language</i>	[RFC1766]	en

Tab. 12: Příklad 2 **Záznam č. 133425**

<i>Title</i>	Factors influencing the rate of degradation of Amoxicillin sodium and potassium clavulanate in the liquid and frozen states	
<i>Identifier</i>	[URI]	http://adt.curtin.edu.au/theses/available/adt-WCU20020502.125711/
<i>Autor</i>	PersonalName	Laleh Vahdat
<i>Subject</i>		<i>words</i> frozen state, potassium clavulanate, liquid state, amoxicillin sodium, chemical stability, PhD Thesis, PhD
<i>Abstrakt</i>	abstract	Kinetics of the reactions of amoxicillin sodium and potassium clavulanate alone and in combination were investigated in the liquid and frozen states at selected pH values of 2.0, 4.6 and 7.0. A stability indicating HPLC assay was developed to perform simultaneous quantification of these compounds validated under stressed conditions. Amoxicillin and clavulanate degradation obeyed first-order kinetics under all conditions of this study...
<i>School/Department</i>	Curtin University of Technology, School of Pharmacy	
<i>Date</i>	valid [W3CDTF]	2000
<i>Rights</i>	http://www.curtin.edu.au/curtin/policy/	
	Copyright Laleh Vahdat	

Oba příklady mají vyplněné pole pro předmětové heslo (subject words). V záznamu nejsou oddělena volně tvořená předmětová hesla od hesel vybraných z řízeného slovníku LCSH. V zásadě to ale nemá žádný vliv na vyhledávání ani na popis obsahu dokumentu a jeho zařazení dle věcného tématu. Počet hesel, kterými je dokument popsán, není přesně určen. Jako heslo může být použit i formální deskriptor popisující o jaký druh práce se jedná, tj. jaký titul byl prací obhájen (viz Příklad 2).

4.3.5.5. **Předmětová brána**

ADT se připojilo k iniciativě, která se snažila vytvoření jednotného přístupu k elektronickým zdrojům pomocí předmětové brány (subject gateway). Předmětová brána může být definována

jako mechanismus pro přistupování do kolekcí kvalitních zdrojů pro podporu výzkumu, které jsou pečlivě popsány a rozčleněny informačními odborníky podle předmětových hesel. Brána je dostupná ve webovém rozhraní. Zdroje mají výhodu společné architektury založené na distribuované tvorbě a rozvoji metadat s možností centrálního přístupu. Možnosti přístupu přes navigační rozhraní vyžadují transparentnost a přesnost. Aktuálně jsou tři možné formy přístupu:

- hledání dle klíčových slov,
- prohlížení indexů,
- strukturované hledání.

Navigační body jsou obvykle zajištěny používáním nejméně jednoho tezauru, který zajistí spolehlivý popis zdroje a jeho pozdější nalezení. Pro různé soubory zdrojů jsou **používány různé tezaury** v závislosti na jejich obsahovém zaměření. Pokud není možné opatřit záznam zdroje vhodným heslem z tezauru, lze použít volně vytvořené předmětové heslo.

Tezaury, druhý aplikační standard, byly implementovány v okamžiku, kdy tvorba distribuovaných metadat ukázala jejich užitečnost. Na tomto místě lze konstatovat, že většina tezurů není propojena s generátory metadat nebo navigačními systémy, ačkoliv některé projekty se o jejich integraci pokoušejí [CAMPBELL, 2000].

4.3.6. Budoucnost ADT

Program Australských digitálních disertací dokázal (lze předpokládat, že bude dokazovat i dalším vývojem), že byl založen úspěšný australský spolupracující a distribuovaný model. Model, který je flexibilní a škálovatelný. Může být prototypem pro skutečně jednotnou národní databázi elektronických disertací. Pracovní skupina programu si je vědoma, že ne všechny disertace mohou být úplně převedeny do elektronické podoby. Jedná se např. o disertace, které obsahují geologický materiál, unikátní umělecké práce nebo webová sídla. Přesto je skupina přesvědčena, že většinu je možné digitalizovat alespoň z části a ty, které skutečně nelze do elektronické verze převést, je možné alespoň v krátkosti popsat pomocí speciálních metadatových formulářů. Systém ADT má nadále v úmyslu spolupracovat na dalších iniciativách rozvíjených v akademické oblasti. Hodlá podporovat rozvoj elektronických diplomových a disertačních prací (ETD) a spolupracovat s celosvětové snaze o jejich dostupnost bez ohledu na geografické nebo ekonomické okolnosti.

4.3.6.1. Rozvoj projektů

ARROW

Záznamy v systému ADT jsou také sbírány pomocí služby ARROW Discovery Service (<http://search.arrow.edu.au/apps/ArrowUI/>), který testuje možnosti Australské národní služby pro získávání zdrojů (Australian National Resource Discovery Service) v prostředí mezidisciplinárních školních dokumentů, za použití sbírání metadat z australských institucionálních repozitářů. Monashova univerzita (Monash University) se stala testovací webovým sídlem pro sběr disertací z repozitáře ARROW. Toto bude část Monashovy univerzity, která se stane aktivním členem programu ADT. V rámci hlavního projektu se repozitáři ARROW podařilo úspěšně nahrávat záznamy a soubory PDF ze softwaru VT-ETD.

Digitool

Firma Ex Libris předvedla schopnost přenést metadatové záznamy z centrální služby ADT do repozitáře Digitool. Tento postup je testován na Univerzitě Nového Jižního Walesu.

PerX

V září 2006 se záznam ze systému ADT začal objevovat také v **PerX** (Pilot Engineering Repository Xsearch, <http://www.icbl.hw.ac.uk/perx/>) pilotním vyhledávací nad strojírenským repozitářem. Je to projekt Komise pro spojené informační systémy (Joint Information Systems Committee, JISC), který se jmenuje Program pro digitální repozitáře. Byla vyvinuta testovací služba k vysvětlení praktických problémů, které by se mohly vyskytnout v úplném předmětovém rozdělení zdroje zpřístupňovací služby přes sérii repozitářů, které obsahují informace zajímavé pro společnost studia a výzkumu ve strojírenství [ADT, 2006].

Spolupráce programu ADT a společnosti Prouest

Program Australských digitálních disertací spolupracuje s komerční společností ProQuest Informace a vzdělávání (ProQuest Information and Learning, PQIL) na výzkumu praktického řešení a nabídky služeb ke splnění dalšího z cílů programu ADT z oblasti otevřeného přístupu k australskému výzkumu přes vystavení metadat ve webovém prostředí. V rámci pilotního projektu program ADT zkoumá potenciál partnerství se společností PQIL v následujících oblastech:

- retrospektivní skenování a digitalizace disertací,
- představení aspektů kontroly copyrightu (kontrola příloh a dalšího identifikovatelného materiálu disertace, který by mohla spadat pod další ochranu autorského práva),

- poskytnutí přístupu do PQIL digitálního depozitního softwaru (PQIL's Digital Commons Deposit Software),
- poskytnutí testovacího repozitáře, který by uchovával retrospektivně naskenované i elektronicky vytvořené disertace, tj. jejich plné texty a zároveň záznamy v podobě metadat (ve formátu PQIL nebo ve výstupním formátu Dublin Core).

Griffithova univerzita (Griffith University), jedna z institucí participující na programu ADT, přijala výzvy, výhody a výstupy z projektu v jejich perspektivě vývoje a vlivu na lokální operace probíhající v systému ADT. Podrobnější zpráva o průběhu je dostupná na webových stránkách Rady knihovníků australských univerzit v části ADT Meeting papers <http://www.caul.edu.au/meetings/adt.html>.

4.3.7. Závěr

Systém ADT může být vhodnou technologií na podporu přístupu k informacím a k ukládání elektronických disertací v akademických knihovnách. S rozšiřováním profilu těchto knihoven, které obsahují a prosazují disertace, mají potenciál znovu povzbudit knihovny v jejich rostoucí pozici informačního brokera a partnera při výzkumu. Program ADT představuje novou roli knihoven, kterou je tvorba a zpřístupnění národního systému disertačních prací. Vzhledem k aktuálnímu výzkumu probíhajícímu v jednotlivých institucích lze konstatovat, že význam tohoto systému stále roste. Role knihoven je tak posílena – neumožňují pouze přístup k informacím vytvořených třetí stranou, ale samy se stávají vydavateli. Stejně jako v akademických knihovnách může být systém ADT užitečný jako asistent na univerzitách, a to konkrétně v přípravě pracovníků ve výzkumu, vyhodnocování disertací a zdůrazňování hodnoty jejich produkce [LAFFERTY; EDWARDS; DOVEY, 2005].

4.4. Selekční jazyky v STK

4.4.1. Historický vývoj

Na úvod lze konstatovat, že do poloviny 50. let 20. století nebyly všechny dokumenty ve Státní technické knihovně věcně zpracovávány. Od konce 50. let začala být do katalogizačních záznamů doplňována předmětová hesla, znaky MDT pouze výjimečně. Katalogizační záznamy vytvářené v knihovně od 60. let 20. století v souvislosti se zavedením mezinárodního formátu lístků obsahují jak znaky MDT, tak předmětová hesla. Předmětová hesla byla vytvářena v souladu s ČSN 01 0188 Tvorba předmětových hesel [ČSN 01 0188] a to v průběhu věcného zpracování dokumentů. Od roku 1978 byla ve Státní technické

knihovně zavedena nová linka zpracování dokumentů. Znaky MDT byly nadále přidělovány všem dokumentům, zatímco pomocí předmětových hesel byly popisovány pouze dokumenty umístěné pod určitými signaturami – konkrétně se jednalo o signatury A a B (vázané knihy), F (skripta) a M a N (seriály). U dokumentů s ostatními signaturami (např. D, E či K) byly uváděny pouze znaky MDT. Ve specifickém případě firemní literatury byly používány tzv. třídníky – jednalo se v podstatě o výtah znaků MDT. Od roku 1995 jsou katalogizační záznamy ve Státní technické knihovně doplňovány o věcné údaje v elektronické podobě. Od téhož roku jsou všem dokumentům přidělovány znaky MDT, předmětová hesla a také nově hesla PSH. Od roku 2001 nejsou do katalogizačních záznamů nadále přidávána předmětová hesla, využívány jsou již pouze hesla PSH a znaky MDT [Státní technická knihovna, 2002].

V souvislosti s přechodem z předmětových hesel na hesla PSH bylo provedeno porovnání těchto dvou souborů hesel, při němž jako východisko posloužily soubory skutečně použitých hesel. Bylo zjištěno, že předmětová hesla (bez podhesel) byla s hesly PSH totožná v 41,2 % případů. předmětová hesla, která vyjadřovala např. geografické pojmy nebo názvy produktů (nejsou součástí PSH), představovala 9,2 % případů a předmětová hesla, která se v PSH nevyskytovala v identické podobě, 48,1 %. Závěrem bylo konstatováno, že zhruba polovinu hesel by bylo možné přeřadit do první skupiny, další část předmětových hesel pak nahradit dvěma hesly

Od roku 1999 do roku 2004 probíhala ve Státní technické knihovně retrospektivní konverze (retrokonverze) katalogizačních záznamů. Od roku 2005 jsou v rámci redakce katalogů doplňovány znaky MDT, a to částečně pouze na základě katalogizačních lístků, částečně metodou de visu, tj. s knihou v ruce. Od roku 2007 jsou znaky MDT doplňovány již výhradně u záznamů dokumentů reprezentujících živý fond (v tomto případě doplňování probíhá vždy metodou de visu) [ŘÍMANOVÁ, 2007].

4.4.2. Polytematický strukturovaný heslář (PSH)

Polytematický strukturovaný heslář je česko-anglický řízený a měnitelný slovník lexikálních jednotek ze všech základních oblastí lidského poznání. Slouží k vyjádření věcného obsahu dokumentů a ke zpětnému vyhledání dokumentů na základě věcných kritérií.

Heslář byl vytvořen knihovnicí ve spolupráci s odborníky z jednotlivých oborů, které pokrývá. Ve verzi 2.1 je heslář rozčleněn do celkem 44 tematických řad a obsahuje přes 13 000 hesel.

Správci hesláře doporučují jej používat zároveň s dalšími selekčními jazyky, ať již řízenými (např. Mezinárodní desetinné třídění) či neřízenými (např. volně tvořená klíčová slova), aby byl uživateli zajištěn nejvyšší možný komfort při věcném vyhledávání informací.

V hesláři se vyskytují tři základní druhy apriorních pojmových vztahů – vztah hierarchie, vztah ekvivalence a vztah asociace.

Vztahy hierarchie existují mezi hesly vyjadřujícími užší a širší pojmy, ekvivalenční vztahy pak mezi hesly a jejich synonymy, popř. kvazisynonymy. V řadě případů je kvůli přehlednosti hesláře jako celku využito hierarchizace vztahu ekvivalence, tj. některé termíny vyjadřující užší pojmy jsou zařazeny jako termíny nepreferované u hesel vyjadřujících odpovídající širší pojmy. Vztahy asociace jsou využity v případech, kdy je pokládáno za přínosné indexátory a/nebo uživatele upozornit na další příbuzná hesla, která jsou umístěna v jiné části hesláře.

Nová hesla do hesláře přidávají jeho správci na základě návrhů indexátorů, koncových uživatelů a externích spolupracovníků (specialistů v jednotlivých oborech). Jednotlivé instituce používající heslář mohou do hesláře přidávat vlastní hesla (např. podrobněji zachytit oblasti, které jsou v hesláři zastoupeny pouze jedním či několika hesly). S důrazem na návaznost vyšších verzí je nutné přidaná hesla oznámit správcům hesláře. Heslář lze zařadit mezi postkoordinované systémy. Znamená to, že složené pojmy jsou vyjádřeny skládáním jednotlivých hesel (na rozdíl např. od klasických systémů předmětových hesel, které jsou zařazovány mezi prekoordinované systémy).

4.4.3. Termíny

V konkrétním případě Polytematického strukturovaného hesláře sem patří všechny termíny obsažené v hesláři, tj. jak preferované (hesla), tak nepreferované termíny (odkazy typu viz).

Preferované termíny (hesla) jsou výrazy, které se používají pro indexaci. Jedná se o jednoslovné (např. *historie hi*) nebo víceslovné neboli složené (např. *informační prameny in*) termíny, které jsou doplněny dvouznakovými zkratkami, které označují příslušnost hesla do tematické řady.

Nepreferované termíny (odkazy typu viz) jsou výrazy, které se nepoužívají pro indexaci. Indexátorům slouží k vyhledání preferovaných termínů a uživatelům pro účely vyhledávání. V hesláři se nacházejí termíny, které vyjadřují pojmy konkrétní i abstraktní povahy (v následující části textu je využito rozčlenění dle [ČSN 01 0193]):

Konkrétní entity:

- předměty, věci a jejich fyzické části:

- *lesnické stroje ze*
- *balkony au*
- materiály:
 - *sklo sp*
 - *dřevo ze*

Abstraktní entity:

- činnosti a události:
 - *odstředování ch*
 - *odkalování ts*
- abstraktní entity a vlastnosti věcí, materiálů nebo činností
 - *pevnost fy*
 - *hmotnost fy*
- obory nebo vědní disciplíny
 - *fyzika fy*
 - *chemie ch*
- měřicí jednotky [v PSH nejsou zařazeny konkrétní měřicí jednotky, ale pouze obecná označení tříd měřicích jednotek]
 - *fyzikální jednotky fy*

Individuální entity neboli “jednotkové třídy” vyjádřené vlastními jmény

- *Galaxie (Mléčná dráha) as*

Hesla jsou tvořena substantivy, která mohou (např. *mezigalaktická hmota as*), ale nemusí (např. *mlhoviny as*) být rozvíta adjektivy. Jsou-li použita adjektiva, jsou přednostně uváděna před substantivy (viz příklady výše). V ojedinělých případech následují adjektiva až za substantivy. Jedná se mj. o odbornou chemickou terminologii (např. *oxid uhličitý ch*).

V některých případech jsou použita substantiva s předložkovou vazbou (např. *čas v ekonomii ev*), popř. kombinace adjektivní a předložkové vazby (např. *výpočetní metody ve fyzice fy*). Vyskytují se rovněž hesla, v nichž je substantivum rozvíta dvěma adjektivy, z nichž jedno se nachází před substantivem a druhé za substantivem (např. *jazykové obory základní ja*). V případech, kdy byl v předchozích verzích použit odlišný tvar hesla, jsou tato hesla z důvodu

zpětné kompatibility do hesláře zařazena jako nepreferované termíny (např. *měření teploty a měření teplot viz teplota fy*).

Příslovce se v rámci hesel vyskytují ojediněle (např. *rozvodny velmi vysokého napětí en*). Slovesa nejsou do hesláře zařazována vůbec. Zkratky a akronymy se v hesláři vyskytují v pozici samostatných hesel (např. *DNA ch*) nebo jako nepreferované termíny (např. nepreferovaný termín *ASŘ* odkazující na heslo *automatizované systémy řízení ev*).

Pokud se týká užití singuláru a plurálu, počítatelná substantiva se v zásadě vyjadřují plurálem (např. *grafy ob*), ale v některých případech je použit singulár (např. *výpověď ja*). U nepočítatelných substantiv je přednostně užit singulár (např. *pivo pp*). Pokud lze předpokládat, že uživatelé budou pokládat např. určitou látku za třídu, pak je použit plurál (např. *jedy ch*). Lexikální jednotky vyjadřující abstraktní pojmy jsou přednostně uvedeny v singuláru (např. *historie hi*), pouze výjimečně v plurálu (např. *chemické reakce ch*). V případě výrazné odlišnosti singuláru a plurálu je v konkrétních případech přidán nepreferovaný termín (např. *dítě viz děti so*). V případech homografie či polysémie jsou v současnosti v hesláři řešeny tak, že dané (např. dva) pojmy jsou vyjádřeny dvěma různými lexikálními jednotkami (např. *křídlo bi* a *křídla letadla sr*). Perspektivně počítají správci hesláře se zavedením relátorů (kvalifikátorů, závorkového výkladu), které se nyní vyskytují v hesláři pouze ojediněle (např. *strukturalismus (filozofie) fi* a *strukturalismus (jazykověda) ja*). V hesláři jsou používány nejrozšířenější (eventuálně kodifikované) pravopisné tvary lexikálních jednotek. Méně rozšířené termíny (včetně zastaralých), popř. slangové výrazy či žargon, jsou uváděny jako nepreferované termíny (např. *kysličníky viz oxidy ch*).

Pokud je pro daný pojem používán běžný název, je přednostně použit běžný název (např. *jaderná chemie ch*, nepreferovaný termín *nukleární chemie*). Pokud je však v odborné terminologii upřednostňován vědecký název, je i v hesláři použit jako heslo název vědecký (např. *substituce ch*, nepreferovaný termín *záměna*). Formální termíny nejsou součástí hesláře. Hesla typu *mapy gr* či *učebnice pe* je doporučeno používat pouze v případech, kdy jimi lze vyjádřit tematiku dokumentu, nikoliv jeho formu [SKOLKOVÁ, 2007].

4.4.4. Struktura

Polytematický strukturovaný heslář obsahuje tři základní struktury – hierarchie, asociace, ekvivalence.

Pro **hierarchickou** strukturu je charakteristický vztah nadřazenosti a podřazenosti. Ve verzi 2.1 hesláře se na nejvyšších místech hierarchické struktury nachází celkem 44 hesel označujících tematické řady. U hesel na všech hierarchických úrovních jsou za příslušným termínem uvedeny dvouznakové zkratky, které označují příslušnost hesla do dané tematické

řady. Díky těmto zkratkám tak lze i u hesel na nižších hierarchických úrovních snadno zjistit příslušnost k dané řadě.

Asociační struktura naznačuje příbuznost hesla vůči jinému heslu, které není v hesláři uvedeno v pozici nepreferovaného termínu ani není vůči danému heslu v hierarchicky nadřazeném nebo podřazeném postavení (např. *balneologie zd* viz též *minerální prameny vo* a *balneoterapie zd*).

Ekvivalenční struktura představuje vazbu mezi heslem a nepreferovaným termínem (např. *péče o kůži* viz *péče o pleť zd*).

V případě všech výše uvedených struktur se jedná o oboustranné (reciproční) vazby.

V hesláři se vyskytují dva základní druhy vztahů hierarchie:

- generický vztah (např. *savci bi - chobotnatci bi*)
- partitivní vztah (např. dýchací systém bi – plíce bi či psychologické směry ps – analytická psychologie ps)

Vzhledem k tomu, že v hesláři nejsou požity identifikátory, nevyskytují se v něm ani kauzální hierarchické vztahy, tj. případy, kdy je vyjádřen vztah mezi obecným substantivem (třídou) a individuálním případem (instancí) označeným vlastním jménem.

Heslář je uspořádán do stromové struktury. Z ní lze odvozením získat např. abecední seznam hesel, nepreferovaných či asociovaných termínů, v závislosti na využívaném programovém vybavení lze v hesláři také vyhledávat podle řetězců a podřetězců.

4.4.5. Indexace dokumentů hesly PSH

Dokumenty nemusejí být popsány pouze jedním heslem, ve většině případů je vhodné použít více hesel. Důvodem je snaha zachytit co nejlépe tematiku popisovaného dokumentu. Indexace dokumentů hesly PSH by se vždy měla vycházet z potřeb uživatelů informačního systému, v němž jsou hesla využita.

4.4.5.1. Výběr konkrétních hesel

K indexaci dokumentů se používají pouze preferované termíny. Každé heslo je zařazeno pouze na jedno místo v hierarchii; je uplatněn princip tzv. monohierarchie. Při výběru konkrétních hesel je nezbytné brát v úvahu kontext, v němž se vyskytují – kromě nadřazených a podřazených hesel také hesla asociovaná a teprve na základě vyhodnocení příslušného kontextu vybrat heslo či hesla vhodně reprezentující pojem obsažený v indexovaném dokumentu.

Z jedné hierarchické linie by mělo vždy, je-li to možné a účelné, být vybráno pouze jedno heslo.

Významově širší heslo volíme v těchto případech:

- dokument by bylo možné popsat všemi hesly z nižší hierarchické úrovně;
- významově užší heslo není v hesláři obsaženo (např. dokument pojednávající o unitermech bude popsán pomocí hesla selekční jazyky deskriptorové in).

Významově užšímu heslu pak dáváme přednost v případech, kdy:

je dokument převážně věnován užšímu tématu, širší téma je zmíněno pouze okrajově; dokument by bylo možné popsat pouze některými hesly z nižší hierarchické úrovně [SKOLKOVÁ, 2007].

V případě, že konkrétní heslo není v PSH obsaženo, je nejprve potřeba se pokusit pojem vyjádřit jiným způsobem – např. rozložit jej na sémantické složky, k nimž lze v hesláři nalézt odpovídající verbální vyjádření, nebo dát přednost heslu na vyšší úrovni obecnosti. Pokud tyto varianty řešení nelze aplikovat, může být daný výraz navržen jako kandidát na nové heslo.

4.4.5.2. Určení významnosti hesel

Doporučeným pořadím jednotlivých hesel v záznamu dokumentu je pořadí významnosti, tj. hesla, která reprezentují nejdůležitější pojmy obsažené v dokumentu, jsou uváděna jako první, hesla reprezentující méně důležité pojmy pak na dalších místech.

V závislosti na možnostech a účelu konkrétního informačního systému je také možné rozlišovat významnost jednotlivých hesel vzhledem k obsahu dokumentu explicitně, např. formou rozlišení hlavních a vedlejších hesel, resp. témat. V těchto případech lze doporučit umožnit uživatelům vyhledávání podle hlavních a vedlejších hesel a rovněž toto rozčlenění vhodně zobrazit v jednotlivých záznamech (např. v podobě popisů hlavní téma a vedlejší téma).

4.4.5.3. Počet hesel

Počet hesel přidělený jednomu dokumentu není omezen. Je-li to možné, měla by však být zachována zásada proporcionality, tj. rozdíly mezi počty hesel přidělených různým dokumentům by neměly být výrazné. Doporučené počty přidělovaných hesel se obvykle stanovují v indexační politice. Není-li v dané instituci explicitně stanovena, lze za optimální počet hesel přidělených na jeden dokument obecně považovat jedno až šest hesel v závislosti na tematice konkrétního dokumentu a na předpokládaných potřebách uživatelů konkrétního informačního systému, v němž je PSH využíván k vyhledávání. Výše uvedený počet vychází z praxe indexátorů ve Státní technické knihovně.

4.4.5.4. Prekoordinace a postkoordinace

Prekoordinace je uplatňována v případech, kdy je pojem vyjádřen složeným termínem. V případě, že pro daný pojem existuje v hesláři odpovídající složený termín, je mu dána přednost. V případě, že pro daný pojem v hesláři odpovídající složený termín neexistuje, je uplatněn princip postkoordinace, tj. dokumentu jsou přidělena dvě hesla (popř. více hesel), která v souhrnu vyjadřují daný pojem. Vzhledem ke struktuře hesláře, která se blíží spíše tezauru než předmětovým heslům, je silně upřednostňován princip postkoordinace. Podmínkou jeho využití je předpoklad, že při užití kombinace hesel nedojde ke ztrátě původního významu. Princip postkoordinace lze úspěšně využít v případě komplexních témat. Příklady uplatnění principu prekoordinace:

- stavební fyzika st, nikoliv stavebnictví st + fyzika fy
- analytická chemie ch, nikoliv analýza ve + chemie ch

Příklad uplatnění principu postkoordinace:

- téma chemická informatika vyjádřené hesly *chemie ch* a *informační věda in* [SKOLKOVÁ, 2007].

4.4.5.5. Aktualizace PSH

Vzhledem k tomu, že řada oborů se dynamicky vyvíjí, je heslář pravidelně aktualizován svými správci, tj. pracovníky Státní technické knihovny. Návrhy na doplnění hesel či úpravu stávajících hesel jsou velmi vítány. K zasílání návrhů je možno použít elektronickou poštu (psh@stk.cz), klasickou poštu (Státní technická knihovna, oddělení speciálních služeb a PSH, Mariánské nám. 5, pošt. příhrádka 206, 110 01 Praha 1) nebo telefon (+420 221 663 465). Základní informace o hesláři včetně všech kontaktních údajů jsou také k dispozici v rámci webové prezentace hesláře, která je dostupná na adrese <http://www.stk.cz/psh.html>.

4.4.6. Rozvoj

Vizí STK v ohledu na PSH je stát se jedním z klíčových poskytovatelů řešení pro věcné vyhledávání informací v informačně-knihovnické sféře v ČR.

Posláním je pomáhat svým zákazníkům při dosahování jejich cílů v oblasti studia, výuky a výzkumu a vývoje se zaměřením na technické disciplíny. Proto je spravován a rozvíjen Polytematický strukturovaný heslář jako uživatelsky přívětivý prostředek pro vyhledávání dokumentů a informací podle tématu.

Primárním cílem je tedy zefektivnit možnosti vyhledávání informací podle tématu pro koncové uživatele. Primární cílovou skupinu představuje Státní technická knihovna a její uživatelé, sekundární cílovou skupinu pak další instituce využívající heslář a jejich uživatelé.

4.4.7. Oblasti rozvoje

4.4.7.1. *Spolupráce s dalšími subjekty*

Oddělení speciálních služeb a PSH, kde momentálně rozvoj probíhá, se v dalších letech soustředí na komunikaci a spolupráci s těmito subjekty:

- ostatní organizační složky Státní technické knihovny,
- spolupráce na projektu tvorby Digitální knihovny pro šedou literaturu probíhající v STK a začlenění PSH do indexace při bibliografickém popisu elektronických diplomových a disertačních prací,
- další informační (především technicky orientované) instituce v ČR (včetně knihoven Českého vysokého učení technického v Praze) a knihovnicko-informační školy,
- a perspektivní projekty (např. česká verze on-line encyklopedie Wikipedia).

Oddělení speciálních služeb a PSH bude systematicky získávat zpětnou vazbu týkající se hesláře od reprezentantů hlavních i vedlejších cílových skupin. Zpětná vazba bude primárně získávána elektronicky, doplnkově dalšími cestami (např. telefonicky či při osobních setkáních). Bude mít několik podob (např. ankety a dotazníky). Získané poznatky a podněty budou pravidelně vyhodnocovány a na jejich základě budou prováděny odpovídající změny v dosavadních činnostech oddělení speciálních služeb a PSH.

4.4.7.2. *Zpřístupnění hesláře*

Oddělení speciálních služeb a PSH se bude dále zasazovat o zajištění nových způsobů zpřístupnění hesláře pro institucionální (zejména knihovny) i koncové uživatele, a to v souladu s relevantními standardy. Z hlediska institucionálních uživatelů je za podstatný krok pokládáno zpřístupnění hesláře jednak v podobě databáze ve formátu MARC 21 pro authority, které umožní stahování záznamů s hesly pomocí protokolu Z39.50, jednak ve formátu XML (např. podle specifikace Kongresové knihovny ve Washingtonu). Z hlediska koncových uživatelů se jedná zejména o rozvoj nástrojů umožňujících věcné vyhledávání informací v uživatelsky přívětivém prostředí, konkrétně ve webovém rozhraní.

Nezbytností je funkční propojení hesláře (resp. jednotlivých hesel) s dalšími systémy pro organizaci poznatků¹² a prohloubení spolupráce s tvůrci těchto systémů, a to jak na národní,

¹² Terminologické sousloví *systém pro pořádání poznatků* zatím není v české odborné terminologii v oblasti knihovny a informační vědy zavedeno, jedná se o překlad anglického výrazu *knowledge organization systems*, pro který je též používána zkratka *KOS*.

tak na mezinárodní úrovni. Předně se jedná o věcné autority Národní knihovny ČR a perspektivně také o znaky předmětového hesláře Kongresové knihovny. Velmi žádoucí a potřebné je také neustálé vylepšování anglické verze hesláře, která může být využita nejenom pro účely indexace, ale také při překládání termínů z angličtiny do češtiny či z češtiny do angličtiny. V horizontu nejbližších let bude též zaměřena pozornost na uvedení hesláře do souladu s existujícími normativními dokumenty pro oblast tezaurů a na vzorku hesláře bude pilotně proveden pokus o rozšíření hesláře do podoby ontologie.

Klíčové je také vylepšení anglické verze hesláře a obohacení hesláře o notaci, aby byl využitelný ke stavění volně přístupného knihovního fondu.

4.4.8. Oblasti útlumu

Údržba a rozvoj software PSH Manager, který je doposud užíván pro správu hesláře, nebude nadále patřit mezi prioritní činnosti.

PSH nebude proto nadále primárně poskytován v podobě textového souboru, ale ve formátu MARC 21 pro autority a v odvozených formátech (mj. ve formátu XML).

Mezi hlavní omezení rozvoje PSH patří omezení finanční, personální a technologická. Finanční omezení lze překonat zpracováním rozvojových projektů a podáváním žádostí o grant na jejich realizaci, personální omezení pak prohloubením spolupráce s ostatními organizačními složkami Státní technické knihovny, popř. s externisty. S technologickými omezeními je možné se vyrovnat prohlubováním znalostí a dovedností v oblasti informačních a komunikačních technologií a rozvíjením spolupráce s počítačovými odborníky.

4.4.9. Závěr

Koncepce rozvoje vychází z analýzy stavu hesláře k počátku roku 2007. Hlavními myšlenkami koncepce jsou 1) uvedení hesláře do souladu se standardy využívanými v oblasti systémů pro organizaci poznatků, 2) zpřístupnění hesláře ve více formátech, 3) aplikace hesláře na různé druhy zpracovávaných dokumentů a 4) rozvíjení hesláře v návaznosti na vyhodnocení zpětné vazby získané od uživatelů.

5. Věcné pořádání šedé literatury

5.1. Definice

Věcné pořádání informací představuje dílčí proces vstupního zpracování informací, který vychází z obsahové analýzy dokumentu. Výsledkem obsahové analýzy dokumentu je selekční

obraz dokumentu, tj. sestava obsahových údajů, které vypovídají o obsahu, tématu či předmětu dokumentu jako celku, jeho jednotlivých částí nebo dokonce jednotlivých, v dokumentu obsažených informací [KOVÁŘ, 1976].

Cílem věcného pořádní informací je vytvořit takový selekční obraz dokumentu, který by umožnil koncovému uživateli dosáhnout při procesu vyhledávání s podporou systémových nástrojů optimální přesnosti a optimální úplnosti, tedy zpřístupnit požadované relevantní dokumenty efektivním způsobem.

5.2. Selekční jazyk

V bibliografických databázích jsou lexikální jednotky součástí selekčního obrazu dokumentu.

Lexikální jednotky zde slouží ke zpřístupnění obsahu dokumentu v procesu indexování¹³ i k vyhledání¹⁴. V plnotextových systémech lexikální jednotky jsou vždy součástí dokumentu samotného; dokument je vyhledán na základě shody lexikálních jednotek obsažených v uživatelském dotazu a jednotek obsažených v dokumentu za podpory systémových nástrojů informačních technologií. Je-li tato shoda stanovena pouze na podkladě frekvenční analýzy bez obsahového/pojmového porozumění textu, výsledná relevance i pertinence vyhledaných dokumentů je nízká. Lexikální jednotka/preferovaný termín selekčního jazyka má „obecně vzato, vyjadřovat terminologii, kterou nacházíme v literatuře, a to zcela nezávisle na tom, kolik jednotlivých slov je v daném případě zapotřebí k vyjádření pojmu“ [VORÁČEK, 1974]. Při této příležitosti je nutno zmínit jeden z nejdůležitějších aspektů, a to **kompaktnost** (celistvost, jednodušnost) **termínů/lexikálních jednotek**. V procesu organizace informací je žádoucí, aby se obsah lexikálních jednotek (preferovaných termínů) co nejvíce blížil obsahu termínů používaných v odborné literatuře a při odborné komunikaci. Identita formy preferované lexikální jednotky a jejího sémantického obsahu na straně jedné s obsahem termínu používaného v odborné literatuře a obsahem pojmu (denotátem) na straně druhé výraznou měrou ovlivňuje přesnost formulace uživatelského dotazu, přesnost vyhodnocování získaných dokumentů, tudíž i kvalitu procesu pořádní a vyhledávání. **Zásada**, „že deskriptory by měly být minimálními lexikálními jednotkami jak po stránce významové, tak i výrazové, a volit tedy jako základní prvky selekčního jazyka (tezauru) převážně jednoslovné pojmy, které se pak mohou kombinovat do libovolně početně stanovených řetězců schopných vystihnout obsah polytematických dokumentů“ [SLAVÍČKOVÁ, 1988], vede k **formálnímu rozkladu termínu** a jeho náhradě několika

¹³ Formulace obsahu dokumentu prostřednictvím předepsaných lexikálních jednotek.

¹⁴ Formulace dotazu uživatelem prostřednictvím jeho vlastních slov, případně pomocí předepsaných lexikálních jednotek.

lexikálními jednotkami z principu obecnějšího významu. Proces vyhledávání je tak posunut do obecnější roviny, což v plnotextových bázích má za následek větší ohlas irelevantních dokumentů. Proto je nutné posoudit oprávněnost existence víceslovných spojení a stanovit kritéria pro jejich používání. K této jednoznačné identifikaci víceslovných lexikálních jednotek zahrnutých do selekčních jazyků mohou výraznou měrou přispět tradiční pořádací systémy, zvláště pak **tezaury polytematického a univerzálního charakteru**, které představují soubor přesně vymezených termínů, propojených přesně definovanými vztahy. Proces výběru lexikálních jednotek pro účely zařazení do lexika univerzálního selekčního jazyka je tedy složitý. Je nutné:

- posoudit jednoznačnost a systémovost termínu v dominantním oboru,
- posoudit jeho případný výskyt v jiných oborech,
- porovnat definice v jednotlivých oborech,
- posoudit jeho zastupitelnost v nedominantních oborech .

V případě neúspěchu je nutno pojednat termín jako homonymum, tj. navrhnout specifikaci pomocí adjektiva, předložkové vazby, případně pomocí rozlišujícího kvalifikátoru (relátoru) [BALÍKOVÁ, 2001].

5.3. Popis tvorby pořádacích znaků

Pořádací znaky mohou být **slova převzatá z názvu nebo podnázvu** dokumentu, případně slova převzatá z jiného místa dokumentu **z anotace či referátu**, nebo **formulovaná** informačním pracovníkem, a to zcela **volně**, či na podkladě nějakého **řízeného slovníku**. Dříve než informační pracovník zformuluje předmětové heslo, množinu unitermů či deskriptorů, musí nejprve ve svém vědomí provést určitou **redukcí nebo kondenzací textu** ve smyslu souvislé, syntakticky svázané věty nebo stručnějšího či rozsáhlejšího souboru vět. Na podkladě této redukce se pak provádí další redukce, jejímž výsledkem je předmětové heslo, nebo sestava deskriptorů. Referát nebo anotace mohou být podkladem pro věcné pořádání informací, ze kterého informační pracovník (a v posledních letech stále častěji i počítač podle určitého algoritmu) vybírá takové termíny, které mu umožňují vystihnout obsah dokumentu nebo jednotlivých v něm obsažených informací, a které se spolu s ostatními termíny (odjinud z dokumentu) stávají součástí selekčního obrazu dokumentu [KOVÁŘ, 1976].

5.3.1. Moderní principy

V dnešní době rozmachu zpřístupňování velkého množství informací se od vyhledávacích systémů očekává nejlépe už zhuštěné a zhodnocené informace. Zhuštěný text je možné získat **algoritmizací procesu** porozumění textu a jeho následnou komprimací. Při komprimaci textu se používají metody **statistické** (jednotky s vyšší frekvencí výskytu jsou nejzávažnějšími nositeli obsahu), **syntaktické** (obsahově nejzávažnější úseky textu se stanoví na základě porovnání syntaktických struktur textu se slovníkem vzorových syntaktických struktur) a metody **sémantické** (na základě sémantické analýzy textu se stanoví obsahově zatížené prvky: nejjednodušší - slova z názvu, podnázvu, složitější - sémantický rozklad na dílčí významově zatížené úseky promluvy - uplatňuje se lingvistická analýza textu). Některé **programy** pracující těmito metodami umožňují sestavit smysluplný text i z výrazů, které neobsahoval výchozí text, podstatně krátí výchozí text (až o 80 %), sumarizují text, vynechávají redundantní věty, registrují podstatné numerické údaje. Jiné systémy pracují na bázi slovníků a vyhodnocují frekvenci výskytů termínu v analyzovaném dokumentu vzhledem k frekvenci výskytů v celém souboru extrahovaných dokumentů, jsou schopny identifikovat jména osob, organizací, lokalit i ve víceslovných výrazech stejného významu, ale odlišného způsobu zápisu. Identifikují jednotlivé technické termíny, mohou identifikovat zkratky i v jejich rozepsané podobě [BALÍKOVÁ, 2001].

5.3.2. Automatická indexace

Pomocí některých systémů je možno přiřadit souboru analyzovaných textů předdefinované předmětové kategorie (Categorization tool), jiné jsou schopny uspořádat vyhledané texty do shluků podle obsahové podobnosti (Clustering tool) [JONÁK, 1998]. Oba texty pracují s pojmem **inteligentní porozumění textu**: v 70. letech převládal podíl lidské inteligence na procesu pořádání informací, v průběhu 90. let se zdá, že tuto oblast zcela opanují systémy umělé inteligence a zásah lidského intelektu při pořádání webovských zdrojů nebude nutný. Počítačový systém zatím není vybaven schopností porozumět textu. Algoritmy, které by měly umožnit počítači textu porozumět, jsou natolik složité, že jsou prakticky nepoužitelné. Vyhledávače (**search engines**) jsou založeny na automatizovaném sběru dat. Předmětem jejich činnosti je indexace internetových stránek, které poté v pravidelných intervalech ukládají do svých pamětí. Takto vzniklé databáze používají při procesu vyhledávání. Při formulaci dotazu používají pravostranné krácení, booleovské operátory, umožňují hledání podle fráze, částečně umožňují stemming. Dosahují však pouze 20% úspěšnosti, protože při vyhodnocování dotazu pracují nejčastěji s frekvenční strukturou textu. Jsou charakteristické množstvím zpracovaných informací, nezaručují však kvalitu těchto informací [JONÁK,

2001]. Dnes se běžný proces automatické indexace skládá ze dvou fází, fáze přípravy dokumentu a fáze výpočtu a sjednocení vah (tj. vlastního algoritmu indexace). **Kompletní algoritmus indexace** se sestává z výpočtu vah značek pomocí adaptivního přístupu¹⁵ a vah značek na základě analýzy vizualizace¹⁶. Následnou normalizací a sloučením těchto vah jsou vypočteny dva vektory ohodnocení dokumentu, vektor *DW* (ohodnocení značek) a vektor *DS* (ohodnocení vizualizace), na základě tohoto procesu jsou pak automaticky přiřazena vybraná slova [MIHULA, 2005].

5.4. Rozdělení pořadacích systémů

Mezi tradiční pořadací systémy řadíme různá třídění, oborové a polytematické hesláře či tezaury a klíčová slova. Pro tuto práci jsou nejpodstatnější polytematické hesláře, tezaury a klíčová slova, takže zbylé zůstanou ponechány stranou.

5.4.1. Jazyky deskriptorového typu (tezaury)

Nutnost pojmenovat složitá interdisciplinární témata a uživatelská nevstřícnost tradičních pořadacích nástrojů vedla ke vzniku a rozvoji specializovaných **selekčních jazyků na bázi přirozeného jazyka** - oborových tezurů. Jejich vázanost na národní jazyky, tematické omezení na jednotlivé obory, případně příbuzné skupiny oborů, se v souvislosti s postupnou internacionalizací a globalizací informační scény pocítují jako výrazný limitující faktor. Téma dokumentu je vyjádřeno sestavou **izolovaných jednotek**, mezi nimiž **nejsou syntaktické vztahy** explicitně vyjádřeny [BALÍKOVÁ, 2001]. Jazyky deskriptorového typu, jejichž typickým reprezentantem jsou oborové tezaury, jsou jazyky **postkoordinované**. Rozvinuly se zvláště v prostředí elektronických katalogů v souvislosti s nástupem vyhledávacích systémů druhé generace, charakterizovaných možností vyhledávat pomocí kombinace slov. Předmět dokumentu bývá vyjádřen sestavou izolovaných lexikálních jednotek - **deskriptorů**, variantní vstup do systému bývá zajištěn pomocí **nedeskriptorů**. Typickým reprezentantem lexikální jednotky je jednoslovné substantivum, přičemž je možno zařadit do systému i víceslovné lexikální jednotky v přirozeném slovosledu, avšak velmi často dochází k jejich rozkladu. Typické pro tezaury je důsledné vyjádření vztahů mezi lexikálními jednotkami, a to **ekvivalence** (synonymie, homonymie), **hierarchie** (nadřazenost,

¹⁵ Naivní přístup výpočtu vah s následným složením ohodnocení ukazuje, že je vhodné brát v úvahu celou cestu od aktuální značky až ke kořenu dokumentu, a to nejen z pohledu vah, ale také druhů značek.

¹⁶ Hlavní myšlenka algoritmu spočívá v ohodnocení vizualizace značky na základě vybraných parametrů písma, kterým je obsah dané značky zobrazen. Sledované parametry jsou nejprve pro jednotlivé značky vyhodnoceny a poté propagovány do podřízených uzlů ve stromě dokumentu. Ohodnocení vizualizace každého termu v dokumentu je pak odvozeno z ohodnocení nejbližší nadřazené značky.

podřazenost), **asociace**. Při vyjádření komplexního obrazu dokumentu je důsledně uplatněn princip postkoordinace. Tezaurus vzniká obvykle deduktivní metodou (postup “shora”).

Mezi výhody patří:

- princip postkoordinace a z toho pramenící,
- přehlednost selekčního obrazu dokumentu,
- snadná tvorba hierarchických struktur,
- snadná údržba,
- snadná manipulace.

Mezi nevýhody patří:

- rozklad víceslovných jednotek - nerespektování kompaktnosti termínu,
- nedostatečná kontrola registrovaných informací (různorodost zdrojů),
- informační šum způsobený parazitním (náhodným) spojením deskriptorů a v důsledku toho,
- velký ohlas irelevantních dokumentů,
- omezení pouze na tematickou část obsahové charakteristiky dokumentu: systém, deskriptorů a nedeskriptorů zahrnuje pouze tematické termíny; tento nedostatek se odstraňuje připojením podpůrných souborů identifikátorů (personálie, jména korporací, geografické názvy atd.) [BALÍKOVÁ, 2001].

5.4.2. Volně tvořené předmětové termíny

Dalším pořádacím systémem ve většině informačních institucí v ČR, které postupně přistupovaly k nějaké formě věcného pořádání, bylo volné tvoření předmětových termínů. Logický předchůdce indexace hesly z řízených slovníků. Některé knihovny u nás i ve světě u této praxe zůstali (viz kap. 4.2). Považují výhody této metody za převažující a využití celého spektra jazyka, zejména s ohledem na budoucnost, která směřuje k vyhledávání plných textů, pro ně za ideální. Vyhledávání a použití volně tvořených termínů charakterizují podobné znaky jako práci s tezaury. Rozdíl je v tom, že **soubor hesel** používaných k indexaci **nepodléhá specifickým pravidlům a není uzavřený**. Tento systém sloužil v prostředí lístkových (manuálních) katalogů. Předmět dokumentu byl vyjádřen několika předmětovými hesly (tři až šest), přičemž předmětové heslo představovalo strukturu (větu) tvořenou podhesly a uspořádanou podle předem stanovených syntaktických pravidel. Soubor

předmětových hesel se vytvářel **induktivní metodou** (tvorba „zdola“), tedy na základě konkrétního materiálu. Mezi lexikálními jednotkami byly definovány dva typy vztahů: **vylučovací a přidružovací**. V syntaktické rovině byla dominantním principem **prekoordinace**, kdy byla definována komplexní pravidla umožňující vystihnout obsah dokumentu detailním způsobem. Standardizace se řešila povýšením řetězce předmětového hesla do autoritního souboru [BALÍKOVÁ, 2001].

K výhodám patří:

- detailní vyjádření předmětu dokumentu,
- odpovídající specifičnost,
- maximální informační hodnotu řetězce předmětového hesla,
- efektivní servis pro uživatele v tradičním prostředí.

Nevýhody, které se projeví v elektronickém prostředí, jsou:

- rozklad víceslovných lexikálních jednotek,
- délka řetězce předmětového hesla - koncové údaje se nezobrazí vůbec, případně na druhé řádce,
- princip prekoordinace uplatněný v syntaktické rovině a z toho pramenící,
- komplikovaná pravidla aplikační syntaxe a redundantnost informací v bibliografickém záznamu.

5.4.3. Klíčová slova

Volným tvořením termínů je předmětovým heslům podobný systém věcného pořádku pomocí klíčových slov. Použití této funkce umožní uživateli vyhledat informační zdroje, **aniž by bylo nutné přemýšlet nad hierarchickou strukturou** služby a odhadovat, kam bylo hledané téma zařazeno. Hledání prostřednictvím klíčových slov je využitelné také tehdy, pokud není možné ztrácet čas procházením jednotlivých úrovní tematických kategorií, nebo jako doplněk předchozího prohlížení. Pokud je vlastní databáze předmětového katalogu prohledávána klíčovými slovy, je třeba mít vždy na paměti, že jde o hledání v databázi zcela jiného typu než u služeb založených na automatizovaném sběru dat. Je-li použito hledání prostřednictvím klíčových slov u předmětového katalogu, lze snadno vidět, že se zvolená slova vyskytují **v názvech dokumentů**, ve stručných **anotacích** nebo v adresách dokumentů (URL), ale

nikoliv přímo v textech nalezených dokumentů. Nejde zde totiž o plnotextové indexování. V databázi předmětově orientovaných služeb jsou uloženy pouze ty údaje, které byly poskytovatelem informace o zdroji **vyplněny do formuláře** při registraci nového zdroje, případně jsou zde uloženy recenze (nebo anotace), kterými doplňují odkazy na vybrané kvalitní, spíše však populární, zdroje sami provozovatelé dané služby. [TKAČÍKOVÁ, 2005]. Zde se otvírá prostor pro konkrétního **autora** elektronické diplomové práce či disertace. Je již běžným standardem, že autor práce sám v jejím začátku zdůrazní klíčová slova z názvu a abstraktu dokumentu. Bohužel neznalost definice předmětových termínů a klíčových slov je občas zavádí k jejich kombinaci a jedno vydávají za druhé, což znamená nutnou pečlivější kontrolu knihovníkem.

Mezi výhody patří:

- slova vybraná z názvu či abstraktu mají často velkou vypovídající hodnotu o dokumentu,
- autor nejlépe zná náplň dokumentu a může ho vyjádřit vlastním jazykem,
- při vyhledávání si uživatelé často pamatují jen části názvu.

Mezi nevýhody patří:

- přirozený jazyk (skloňování v rámci gramatiky českého jazyka je těžko vystihnout při vyhledávání),
- vypovídající název (anglosaské země mají ve zvyku dávat svým pracím jasně vypovídající název, slovanské země jsou více poetičtější a použití rejstříků KWIC¹⁷ a KWOC¹⁸ může být zavádějící, např. název „Domácí klenoty“ může znamenat faunu, floru či třeba děti.)

5.5. Harmonizace selekčních jazyků

Trendy dnes vedou mimo jiného k harmonizaci věcných selekčních jazyků Existenci mnoha věcných selekčních jazyků lze na první pohled lehce spojit pomocí „**uzákonění**“ jednoho

¹⁷ Keyword-in-Context. Rotovaný rejstřík klíčových slov z názvu, podnázvu nebo textu dokumentů abecedně seřazených ve vertikální úrovni uprostřed sloupce a v jednotlivých řádcích cyklicky doplněných úplným nebo zkráceným názvem dokumentu [TDKIV, termín KWIC].

¹⁸ Keyword-out-of-Context. Rotovaný rejstřík klíčových slov z názvů, podnázvů nebo textů dokumentů seřazených abecedně, přičemž klíčová slova se uvádějí odděleně před nebo nad úplným nebo zkráceným názvem, ze kterého pocházejí [TDKIV, termín KWOC].

národního jazyka. Preference angličtiny je samozřejmá, ovšem prozatím je tento návrh nereálný. Další možnost propojení je možné zavedením druhého **paralelního selekčního jazyka**, kde se zatím jeví jako nejlepší Deweyho desetinné třídění (DDC). Tuto možnost lze obměnit tím, že budou zavedeny dva systematické jazyky Mezinárodní desetinné třídění (MDT) a Deweyho desetinné třídění (DDC). Ani tato varianta není úplně ideální a je tedy spíše přistupováno k harmonizaci používaných selekčních jazyků. Výzkum realizovaný Národní knihovnou ČR v roce 1997 prokázal, že v oblasti věcných selekčních jazyků naprosto převládá používání izolovaných lexikálních jednotek, tj. klíčových slov, nejčastěji tzv. volně tvořených, tedy ve své podstatě neřízených předmětových termínů, a dále deskriptorů oborových, případně polytematických tezaurů a heslářů. Prekoordinovaná předmětová hesla s rozvinutou strukturou jsou v naprosté menšině a navíc se nejčastěji používají pouze jako alternativní selekční jazyk [BALÍKOVÁ, 2001]. Znamená to tedy, že při věcném pořádní je dominantní princip postkoordinace. Za této situace bylo dospěno k závěru, že z výše uvedených variant postupu v oblasti aplikační syntaxe autoritních prvků bude více odpovídat **našemu prostředí varianta druhá**, tedy autorizace jednotlivých prvků předmětového hesla.

5.6. Zhodnocení kvality selekčního jazyka

Kvalita věcného pořádní informací a jejich zpřístupnění je ovlivněna hodnotícími algoritmy při stanovení relevance dokumentů [JONÁK, 2001].

Současné vyhledávací služby pracují na podkladě statisticko-**pravděpodobnostních metod**. Bylo již řečeno, že tyto vyhledávací nástroje nejsou schopny porozumět významu slov, nejsou schopny posoudit význam slova v kontextu. Nejsou také schopny uspokojivě řešit problémy vyvolané synonymií, homonymií a polysémií termínů vyskytujících se v textu.

Při stanovení **relevance** dokumentů používají algoritmy založené:

- na výskytu slov; vyhledané dokumenty jsou uspořádány podle počtu shodných slov vyskytujících se v uživatelském dotazu a ve vyhledaném dokumentu, přičemž výše jsou hodnoceny dokumenty, které obsahují větší počet výskytu jednotlivých slov dotazu,
- na počtu hypertextových odkazů; dokumenty jsou hodnoceny podle počtu směřovaných odkazů; dokumenty s větším počtem odkazů jsou hodnoceny výše.

Kvalitu věcného pořádní informací a jejich zpřístupnění ovlivňuje i **způsob prezentace** věcných údajů.

Při posuzování tohoto jevu vycházíme z těchto předpokladů:

- „koncový uživatel, člověk v komunikaci s informačním systémem či informační službou přímou či zprostředkovanou informační institucí nebo informačním specialistou, je středem celého problému, kolem kterého se odehrávají procesy spojené s vyhledáváním informací v informačních a počítačových systémech“, [PAPÍK, 2001]
- čtení z obrazovky počítačů je o 27 % pomalejší než z papíru,
- koncový uživatel je schopen vnímat a identifikovat údaje v rejstřících nacházejících se na prvních dvou až tří pozicích,
- údaje prezentované v dalších pozicích jsou pro mnohé elektronické katalogy nezobrazitelné („mizí za obrazovkou“), případně se zobrazují na druhé řádce [PAPÍK, 2001].

5.7. Věcný popis elektronických dokumentů

5.7.1. Problematika indexace

V on-line prostředí se jeví jako užitečné původní bohatě strukturované řetězce prekoordinovaných předmětových hesel výrazně zkrátit a ukládat selekční prvky do vstupních pozic. Maximální **informační hodnota řetězce** předmětového hesla je hodnotné zejména při vyhledávání v tištěných zdrojích. V on-line prostředí je nutno tuto zásadu přehodnotit.

Koncový uživatel je stále více konfrontován s existencí hypertextových odkazů i v situacích, kdy je neočekává a které kladou větší nároky na jeho pozornost: prohlížení nebo též listování či browsing. Dlouhá, bohatě strukturovaná informace v této situaci ještě ztěžuje vnímání, a je proto považována za uživatelsky nevstřícnou [BALÍKOVÁ, 2001] .

5.7.2. Možnosti přístupu k indexaci elektronických dokumentů

Obecně užívání elektronických informací vyžaduje účelnější přístup.

U databází není přesně definovaný začátek nebo konec, omezení je pouze u indexů které mohou být zobrazeny na obrazovce v abecedně číselném pořadí. Uživatel hledající informaci z elektronických zdrojů potřebuje mít jasně definován soubor termínů a návrhy ještě před podniknutím úspěšného hledání.

Důležitý rozdíl mezi vyhledáváním v tištěných a elektronických zdrojích je způsob, jakým byla informace indexována s ohledem na ulehčení jejího nalezení. K indexování a vytváření abstraktů tištěných zdrojů se obvykle užívá **řízený slovník**. To znamená, že indexátor může popsat spoustu článků o „absenci“, „pracovním uspokojení“, „pracovním vytížením“, „poradenství“ pod jedním heslem „personální management“. Toto je následek sdružování

podobných článků dohromady k usnadnění jejich nalezení. Občas to může způsobit problémy při hledání specifického tématu, závisícího na stupni detailnosti indexování.

Věcný popis elektronických informačních zdrojů používá přirozený jazyk, takže lze vyhledávat i podle velmi specifických hesel, jaké jsou uvedeny ve výše zmíněném příkladě. Separátní vyhledávání dle konkrétních uvedených hesel nám systém vrátí konkrétní články na zadané témata. Hodně poskytovatelů elektronických informací používá řízený slovník a zároveň umožňuje vyhledávat pomocí **termínů z přirozeného jazyka**. To je zjednodušené vysvětlení jednoho z mnoha očividných problémů v teorii informací, ale existuje mnoho dalších podobných i nepopsaných problémů. Hlavní myšlenka je, že úspěšné nalezení informace z databáze je závislá na uživateli (na termínech, které při vyhledávání použije). Je třeba mít na paměti, že systém udělá pouze to, co mu uživatel řekne. Většina vyhledávacích systémů nepozná, kdy je slovo napsáno špatně nebo kdy se jedná jen o překlep. Překlepy nyní již umí systém, užívaný např. ve vyhledávači Google, rozpoznat a po srovnání s vlastním rejstříkem slov nabízí nejblížeš možnou správnou variantu slova. Jasná výhoda elektronického vyhledávání informací je, že lze potenciálně vyhledávat dle všech slov použitých v databázi. Takže třeba kompletní text novinových článků může být prohledán na základě jména společnosti [FISHER, 1995].

5.7.3. Věcný popis elektronických diplomových a disertačních prací

Klíčovým cílem posunu směrem k elektronickým diplomovým a disertačním pracím je zlepšení povědomí o rostoucí distribuci studentského výzkumu. NDLTD (viz kap. 4.2) o němž se diskutuje od roku 1987, umožnil soustředění na tyto snahy. Byl vyvinut jednotný vyhledávací systém, který umožňoval vyhledávat disertace a diplomové práce z různých databází z mnoha zemí. S cílem zvýšit povědomí těchto elektronických dokumentech přijaly knihovny mnoho opatření a metod k rozšíření možnosti přístupu k jednotlivým záznamům.

5.7.3.1. *Metody zpřístupnění bibliografického záznamu*

Jedna metoda je **integrace záznamu do knihovního katalogu**. Uživatelé naleznou nejen knihy a různé dokumenty, ale i disertace. Tento způsob je ideální pokud uživatel hledá jakékoliv druhy dokumentů na zadané téma. Nejde o nic neobvyklého pro hledání třeba v on-line knihovním katalogu, kde jsou obsaženy hypertextové odkazy na informace o knihách, videu, jiných multimédiích a odkazy, které mají prokliknutou internetovou adresu. Záznam často sám také v sobě obsahuje link na plný text. Druhá metoda je sdružit elektronické diplomové a disertační práce do **samostatné databáze**, nejspíše jako část „virtuální knihovny“. Je více použitelná pro informačně gramotnější uživatele, kteří hledají přímo

disertaci jako typ zdroje informací. Hledají zvláštní druh dokumentu, tedy disertaci. Takovým uživatelům je dobře známa cena, kterou tyto práce mají, neboť jsou výsledky výzkumů, ačkoliv jen studentských. Ve skutečnosti je ale nejlepší používat obě metody společně [FRANK; ROWE, 2004].

5.7.3.2. Aplikace metod

Technická univerzita ve Virginii (Virginia Tech) a Univerzita na Severní Floridě (University of South Florida) na tento ideální způsob užití obou metod přistoupila. Část záznamu disertace použitého pro on-line katalog slouží pro vyhledávač (search engine). Materiál je tedy prohledatelný jako skrytá databáze pomocí search engine. Na první pohled se tento přístup může jevit jako multiplicitní a krokem zpět. Ovšem když nebudou dostupné elektronické verze disertací a diplomových prací, knihovníci je pořád budou schopni nalézt ve svých fondech po staletí [FRANK; ROWE, 2004]. Řada knihoven připravuje záznamy závěrečných prací a ukládá je do fondu. Základní prvky bibliografického popisu jsou předmět, autor, titul, lokace a občas obsah dokumentu nebo abstrakt. Lokace může ještě obsahovat vedoucího práce nebo fakultu, kde byla práce obhájena.

5.7.3.3. Nároky na lidské zdroje vs. užitečnost předmětových hesel

Knihovny mohou použít software, který slouží k rozpoznávání znaků (Optional Character Recognition, OCR), k přístupu do abstraktu a umožnění fulltextově v něm vyhledávat. Tento druh vyhledávání patří dnes mezi nejoblíbenější způsob nalézání informací. Oproti tomu je používání předmětových hesel pracnější a časově náročnější. Přidání hesel z příslušného hesláře nebo tezauru zabírá indexátorům mnoho času a vyžaduje jejich absolutní pozornost a schopnosti. Je tedy otázkou, zda při zlepšování vyhledávacích schopností by bylo lépe šetřit čas indexátorů s tím, že bude systém ochuzen o jeden způsob hledání, neboť fulltextové prohledávání jej může plně nahradit. Rozhodnutí přidat plnohodnotné předmětové hesla je důležité téma zasluhující jistě další diskusi. Časově nejnáročnější část katalogizace diplomových a disertačních prací je tedy analýza textu a následné přidání předmětových hesel. K této práci je nezbytné mít zkušené a profesionální katalogizátory [FRANK; ROWE, 2004].

5.7.3.3.1. Výhody přidávání předmětových hesel do záznamů

Proč tedy přidávat ke klíčovým slovům ještě předmětová hesla, zvláště má-li systém ještě k dispozici fulltextové vyhledávání? Jeden způsob jak uživatelé nalézají disertace je na okraji on-line knihovního katalogu, jinak řečeno náhodně během vyhledávání informací. Během přidání předmětových hesel do záznamu v tab. 13, uživatel hledající v on-line knihovním

katalogu nejspíše nenajde disertaci na téma „sebeúcta“. Název totiž zmiňuje „strategii sebeovládání“, ale žádné termíny jako „stres“ nebo „sebeúcta“, které rovněž patří k tématu disertace. Zahrnutí ověřeného hesla ze souboru LCSH (v tab. 14 pole označené Subject, general), v našem příkladě „sebeúcta u dětí“, do katalogizačního záznamu zajistí, že uživatel najde tuto položku v různých dokumentech a to dle prohlížení pole vyhrazeného pro **specifikaci obsahu dokumentu** pomocí předmětového hesla.

Další dobrý důvod, proč používat předmětová hesla, je zvládnutí **nepřesností jazyka**. Ač je čeština jazyk velmi specifický i v něm se najdou případy jako např. termín „operace“, který může patřit do mnoha oborů a v každém má trochu odlišný význam. Angličtina je na tom v tomto úhlu pohledu o něco hůře. Na příklad termín „pitch“ znamená něco jiného pro muzikanta (ladička), architekta (svah) nebo inženýra (dehtovat) atd. [FRANK; ROWE, 2004]. Pravděpodobnost nalezení správně tematicky definované disertace pak závisí na správné vyplnění pole pro předmětové heslo .

Tab. 13: Příklad katalogizačního záznamu [FRANK; ROWE, 2004].

Santa-Lucia, Raimong C.	
A situational investigation of hassles, uplifts, coping strategies, and adjustment in 3rd - through 5th-grade children/by Raymond C. Santa-Lucia.	
1998	
v, 35 leaves : ill.; 29 cm.	
Thesis (M.A.) - University of South Florida, 1998.	
Includes bibliographical references (leaves 33-35).	
(s=):	
Self-esteem in children.	
Stress (Psychology) in children.	
Adjustment (Psychology) in children.	
Dissertations, Academic-USF-Psychology-Masters.	

Tab. 14: Výsledky hledání za použití klíčových slov „pitch and dissertations“ v on-li katalogu WebLUIIS na Univerzitě na Jižní Floridě [FRANK; ROWE, 2004].

Author, etc.:	Goodwin, Mar A.
Title:	The effectiveness of Pitch Master compared to traditional classroom methods in teaching sign singing to college music students/by Mark A. Goodwin
Published:	1990
Description:	xi, 167 leaves: Ill., music; 28 cm

Notes:	Thesis (Ph. D.)-University of South Florida, 1990. Includes bibliographical references (leaves 135-141).
Subject, general:	Sight-singing-Instruction and study. Music-Programmed instruction. Dissertations, Accademic-USF-Music education-Doctoral
Author, etc.:	Shaw, Jill D. K.
Title:	The relationships in the usage of oral contraceptives and their effects on vocal pitch and vocal quality : a short term study/by Jill D. K. Shaw
Published:	1979
Description:	viii, 42 leaves; 29 cm
Notes:	Thesis (M.S.)-University of South Florida, 1979.
Bibliography:	leaves 34-36
Subject, general:	Oral contraceptives-Side effects. Larynx. Dissertations, Academic-USF-Speech-Language Pathology-Masters.
Author, etc.:	Lutfi, Robert A.
Title:	The effects of uncertain mask intensity and frequency on pitch judgments in the backward recognition masking paradigm/by Robert A. Lutfi.
Published:	1977
Description:	iv, 54 leaves; 29 cm
Notes:	Thesis (M.A.)-University of South Florida, 1977.
Bibliography:	leaves 52-54
Subject, general:	Psychology, Experimental. Human information processing. Dissertations, Academic-USF-Psychology-Masters.

5.7.3.3.2. *Nevýhody přidávání předmětových hesel do záznamů*

Navzdory výhod předmětového pořádání se některé knihovny rozhodly jej nepoužívat, zejména z důvodu **omezeného množství zaměstnanců**, knihovny by pak musely čelit nárůstu množství práce kladené na jednotlivce. Aplikace předmětových hesel se v takových případech může stát skličujícím břemenem. Nicméně bez přidání předmětových hesel do záznamů mohou uživatelé zjistit, že se musí naučit víc používat invenci, aby odhalili všechno, co jim konkrétní instituce nabízí. Tím, co je výše popsáno, se ovšem nesnižuje hodnota fulltextového prohledávání a vyhledávání dle klíčových slov v abstraktu, pokud je v záznamu obsažen. Důležitost přístupu k dokumentům přes předmětová hesla stále vyvolává spoustu diskusí mezi knihovníky [FRANK; ROWE, 2004].

5.7.3.3. Alternativní řešení

Jedna metoda nahrazující předmětová hesla závisí na vstupu spíše autora než knihovníka, který hodnotí obsah dokumentu, jako součást procesu zpracování. Zavedení této praxe je v podstatě pokus o ušetření peněz a času indexátorů. Na první pohled se metoda může zdát jako ideální řešení, ale objevil se jeden závažný problém. Katalogizátoři zjistili, že **hesla přidaná autory** nejsou v souladu s používaných heslářem či tezaurem. Ze zkušeností na univerzitě na Jižní Floridě je patrné, že studenti, kteří s velkým vypětím sil napíší diplomovou nebo disertační práci, nejsou schopni doplnit vhodná předmětová hesla. Zavedli tedy kurzy na používání předmětového hesláře LCSH, ale i tam se potvrdila bezradnost autorů. Po zhodnocení všech snah došli knihovníci k závěru, že ušetřený čas z indexace věnovali školením autorů, takže výsledek nebyl příliš uspokojivý. Nutno ale podotknout, že autoři proškolení v kurzech přidělovali předmětová hesla přesně a správně [FRANK; ROWE, 2004].

5.7.3.4. Závěr

Poskytnout autorům disertačních a diplomových prací přístup do hesláře nebo tezauru, který je v knihovně používán, není evidentně dostačující a je třeba autorům vysvětlit jeho používání. S pomocí moderních technologií, např. vytvořením animované nápovědy ve formátu flash, by bylo možné ušetřit čas pracovníků na kurz, a přesto by autoři mohli být názorně a jednoduše seznámeni s funkcí a přiřazováním předmětových hesel. Metoda založená na spolupráci s autorem na vytvoření bibliografického záznamu kvalifikační práce tedy rozhodně není nepoužitelná a bez budoucnosti.

Studenti se jako autoři mohou podílet na tvorbě záznamu své práce i v jiných oblastech. Mohou vytvořit abstrakt se zdůrazněním některých klíčových slov, vkládat název práce, fakulty či katedry, kde práci obhájí a další konkrétní informace. Studenti jsou obvykle vyzváni, aby dodali klíčová slova, což jsou slova z názvu a abstraktu. Student – autor je sám nejlepším expertem na informační obsah práce a může nejlépe posoudit synonyma a/ nebo alternativní termíny popisující obsah jejich práce. Katalogizátoři pak mohou tyto informace přenést do záznamu v on-line katalogu, případně k nim přihlídnout při přidělování předmětových hesel. Tuto praxi převzali např. na Technické univerzitě ve Virginii (Virginia Tech), kde katalogizátoři metodou „kopíruj a vlož“ obohacují záznamy o abstrakty vytvořené studenty. Jejich programátoři dokonce vytvořili v jazyce PERL skript na mapování předávaných informací do MARC záznamu [FRANK; ROWE, 2004]. Pomocní pracovníci tak mohou zvládnout tuto práci knihovny Technické univerzity ve Virginii a knihovníci tím opět získají další volný čas zabývat se dalšími pracemi, jako jsou přidávání předmětových hesel do záznamů.

5.7.4. Standardy

Byl již zmíněn standardní formát záznamu jménem MARC, ale existují samozřejmě další, např. Dublin Core či Text Encoding Initiative. Tyto standardy se od sebe liší zásadně typem dokumentů, které popisují a svou podobou (tj. jsou-li v tištěné či elektronické podobě). V zásadě se ale v mnohých popisných polích překrývají. Pro člověka, který se nikdy katalogizací nezabýval, se může jevit diskuse o národních a mezinárodních standardech záznamů a vyplňování jednotlivých elementů jako příliš složitá. Nicméně všechny tyto nastavené pravidla jsou výhodou právě pro uživatele, neboť při sdružení jednotlivých databází do jedné společné a pod jeden vyhledávač, se dosažení konzistence záznamů stává pro databázi životně důležité [FRANK; ROWE, 2004].

5.7.5. Využití PSH k indexaci elektronických diplomových a disertačních prací

Pokud se týká katalogizace diplomových a disertačních prací, panuje všeobecná shoda, že jejich úplná předmětová analýza je složitá, časově náročná a i finančně drahá. Většina známých autorů, jako Haris a Huffman, Keller, Lee-Smeltzer a Sapon-White a Hansbrough i další, nicméně udávali, že předmětový přístup k tomuto druhu dokumentu je nezbytný. Samozřejmě jsou i tací, kteří jsou k předmětovému vyhledávání více kritičtí a tvrdí, že jej lze nahradit vyhledáváním dle klíčových slov. K těmto patřili např. Pepp a Glaviano, Velez-Vendrell, Halverson či Salas-Tull [HOOVER, 2005]. Přes všechna možná pro a proti se Státní technická knihovna rozhodla pro svůj projekt Národního uložště šedé literatury (viz kap. 4.1.4.2.3) použít v záznamu elektronické diplomové práce či disertace heslo vybrané z řízeného slovníku a to Polytematického předmětového hesláře. Navazuje tak na tradici, kdy byla správcem záznamů vysokoškolských kvalifikačních prací, které byly posílány do centrální databáze SIGLE. Ačkoliv s malou odlišností, totiž že diplomové a disertační práce se budou přijímat výhradně v elektronické podobě (viz kap. 4.1).

5.7.5.1. *Srovnání předmětového pořádní SIGLE vs. heslář PSH*

K předmětovému pořádní SIGLE je nutno vzhledem k Polytematickému strukturovanému hesláři (PSH) podotknout, že je velmi obecné a podléhá rozdělení vědy jak jsme zvyklí spíše v anglicky mluvících zemích. Heslář PSH je mnohem více strukturovaný a rovněž podrobněji zaměřený na techniku, ačkoliv je polytematický, jak už bylo popsáno v kapitole 2.2 . Jeho stromová struktura v grafickém zobrazení je schopna uživatele jednoduše navést a pomoci mu v orientaci v oboru. Je ideálním pomocníkem při vyhledávání pomocí metody prohlížení neboli browsing.

V příloze č. 2 je zobrazeno porovnání obou systémů věcného pořádku. Detailnější rozdělení je zřejmé. Heslo, které v SIGLE popisovalo např.

- Strojírnoství, řízení výroby, stavebnictví a lodní stavitelství

definuje PSH čtyřmi samostatnými hesly

- Strojírnoství sr
- Řízení ev
- Stavebnictví st
- Vodní stavby vo

tato je možné samozřejmě dále specifikovat. Vzhledem k lehké dostupnosti hesláře a hlavně jeho grafické stromové struktuře by jistě nebyl problém zaučit do jeho užívání i studenty (autory elektronických diplomových a disertačních prací), např. formou vytvoření elektronického průvodce. Odborní indexátoři by pak mohly pouze kontrolovat jejich správnost a relevantnost.

5.8. Vyhledávání v elektronických disertacích a diplomových pracích

5.8.1. Úvod

Pro zachování maximalistického přístupu při vyhledávání, které je ve stávající situaci nejvíce žádoucí v rámci poskytnutí maximálních možností pro uživatele k vyhledání požadované informace, je nutné aplikovat v možnostech pokročilého hledání využití vyhledání z polí předmětových hesel, klíčových slov i přímo fulltextové vyhledávání (viz příklady v příloze 3). Záleží především na informační gramotnosti uživatele.

5.8.2. Uživatelé a jejich schopnosti

Předmětové hledání bylo identifikováno jako typ hledání, který vyvolává většinu problémů u uživatelů. Mnoho jich je zřejmě zaviněných dotazovacím softwarem, rozhraním, systémem a zvolenou politikou indexování, nedostatkem znalostí v části seznamu předmětových hesel, který uživatelé prohlížejí a nedostatky booleovského modelu.

Larson [LARSON, 1991] identifikoval hlavní problémy hledání v on-line knihovního katalogu pomocí předmětových hesel jako nedostatek znalostí uživatelů. Předmětového

hesláře Kongresové knihovny (LCSH) a dále problémy LCSH s mechanickými a koncepčními aspekty formulace dotazů. Jejich výsledkem jsou hledání, které nejsou úspěšná dále naopak ty, které jsou úspěšná až příliš, a ty která nejsou relevantní k tomu, co měl uživatel na mysli. Autoři Drabenstott, Novotny a Sihvonen a Vakkari ukazují jak odborní uživatelé získají lepší výsledky jejich hledání pomocí předmětových hesel než uživatelé laici. [DRABENSTOTT, 2003], [NOVOTNY, 2004], [SILHONEN; VAKKARI, 2004]. Pro odbornou knihovnu je tedy využití předmětových hesel volně tvořených i z řízeného slovníku mnohem užitečnější. Názory na užitečnost předmětového vyhledávání se ovšem liší. Na druhou stranu se ukazuje, že bez ohledu na úroveň znalostí v hledané oblasti a schopností jednotlivých uživatelů, strategie aktuálně používat známé termíny vede k nejčastějšímu hledání dle autora a názvu dokumentu. Wylie provedl průzkum, které pole bibliografického záznamu uživatelé označují za nejužitečnější k vyhledávání. Po vyhodnocení zjistil, že předmětové heslo označilo jen 26 % respondentů [WYLY, 1996]. Vyhledávací nástroje jsou určeny všem uživatelům, nejen informačním specialistům. A tak je snahou jejich provozovatelů poskytnout i těm méně zkušeným větší možnost kontroly nad výsledky vyhledávání, aniž by se uživatelé museli příliš zatěžovat přemýšlením nad správnou formulací svých složitějších dotazů. Proto většina služeb nabízí uživatelům pokročilé rozhraní s možností „zformulovat“ složitější dotaz výběrem příslušných položek z menu [TKAČÍKOVÁ, 2005].

5.8.3. Dotazování vs. prohlížení

Specifické hledání – **dotazování** – se užívá, když uživatel přesně ví, jakou informaci hledá, a dívá se na dostupné termíny v databázi, která by mohla obsahovat požadovaný dokument.

Dotazování je zaměřeno na konkrétní znalosti uživatele. V rámci pokročilého vyhledávání je možné vyplňovat informace, které má uživatel k dispozici z dřívějšího průzkumu, do jednotlivých polí. Pomocí těchto informací jsou prohledávány bibliografické záznamy z databáze či fondy knihovny. Pokročilé vyhledávání běžně obsahuje vyhledávací pole autor, název, předmětové heslo/klíčové slovo, abstrakt/celý text (fulltextové vyhledávání), datum, univerzita/fakulta atd. Ačkoliv je vyhledávání metodou dotazování orientováno na konkrétní vyplnění známých informací, je užitečné využít možnosti roletových menu k polím typu datum, univerzita/fakulta a podobně.

Naopak **prohlížení** souborů přístupových bodů je vhodný způsob, pokud požadované informace nemohou být konkrétně vyjádřeny. Systém musí pomoci uživateli listovat dokumenty a objevit informaci obsaženou v dotazu.

Prohlížení jako vyhledávací strategie vyžaduje snahu na straně uživatele a tím více záleží úspěch hledání na potenciálu systému a jeho uživatelském rozhraní.

Analytické vyhledávací strategie jsou spíše orientovány předmětově, jsou více systematické a formální, proto jsou lépe plánovatelné a efektivnější [VILLÉN-RUEDA, 2007]. Prohlížení podle Marchioniniho analýzy širších prohledávacích aktivit lze rozdělit do tří základních skupin :

- přímé nebo specifické prohlížení,
- částečně přímé prohlížení,
- nepřímé nebo též všeobecné prohlížení

[MARCHIONINI, 1995].

5.8.3.1. Prohlížení ve Státní technické knihovně

Podle předmětových hesel, ať už volně tvořených nebo z řízeného slovníku uživatelé vyhledávají zejména v režimu prohlížení rejstříků. Pro získání konkrétního obrazu o použití těchto hesel bylo provedeno zpracování logů Apache na serveru df.stk.cz, na němž běží knihovní systém STK založený na databázi Dataflex, za období od 16. 6. 2006 do 2. 8. 2007. Jde o vyhledávání skutečných uživatelů, nikoliv robotů (Google, Seznam či Jyxo) přes rozhraní on-line knihovního katalogu.

Z výsledků v tab. 15 je zřejmé, že používání rejstříků předmětových hesel je celkově na třetím místě s 20 141 dotazy. Z toho ve volně tvořených předmětových heslech uživatelé hledají 3x více než v rejstříku hesel PSH. To můžeme jednoznačně zdůvodnit tím, že uživatelé obvykle netuší co konkrétně znamená označení PSH nebo Polytematický strukturovaný heslář. Je tedy jednodušší zůstat pouze u jednoduchého předmětového hesla. Tab. 16 nám dále ukazuje využití rejstříků předmětových hesel při hledání z IP adres dalších velkých akademických institucí. Je zřejmé, že čím víc jsou uživatelé technického zaměření a znalí tedy odborných výrazů, tím je pro ně hledání pomocí prohlížení rejstříků předmětových hesel úspěšnější a použitelnější.

Tab. 15: Celkové prohlížení rejstříků v STK v období 16. 6. 2006 – 2. 8. 2007

Hledání v rejstříku	Počet uskutečněných dotazů
Autor	31 568
Název	57 274
Předmětové heslo PSH	4 189
Předmětové heslo (volně tvořené)	15 952

Ostatní (MDT, vydavatel, ISBN, ISSN, signatura)	64 928
Celkem	173 911

Tab. 16: Prohlížení rejstříků předmětových hesel z jiných institucí v období
16. 6. 2006 – 2. 8. 2007

Instituce	Rejstřík předmětových hesel	Rejstřík předmětových hesel PSH	Celkem
STK	8 104	1 913	10 017
ČVUT	1 243	359	1 602
VŠCHT	210	82	292
UK	282	136	418

5.8.4. Vyhledávání podle autora, názvu, předmětového hesla

Navzdory každodennímu užívání předmětového vyhledávání v on-line prostředí jsou mezi uživateli stále patrné problémy. Podle provedeného výzkumu on-line veřejně přístupných katalogů knihoven (OPAC), který provedla Rada pro knihovní zdroje na začátku 80. let, ukázal předmětové vyhledávání jako velmi hodnotné pro uživatele [MATTHEWS, 1983]. Tato a další pozdější studie udávaly, že přístup skrze předmětové heslo bylo jedním z převládajících oblastí zájmu v rámci on-line knihovního katalogu. Nyní databáze nabyly většího rozsahu, výsledky vyhledávání proto vrací méně přesné odpovědi. Čím dál častěji vrací příliš mnoho odpovědí na dotaz, které pak vyvolávají v uživateli jisté frustrace. Práce v této oblasti navrhuje, že zmíněné dva hlavní problémy (selhání vyhledávání i přemíra vrácených informací) vyžadují pozornost. Jsou také důvodem proč hledávání dle předmětových hesel, ačkoliv hodně ceněné uživateli, je dnes méně využíváno ve srovnání s hledáním dle autora a názvu. Důvody, které způsobují menší množství hledání dle předmětových hesel, mohou být:

- nedostatečný růst seznamů předmětových hesel. Politika přidělování hesel v katalogu. Indexace může být příliš obecná s malým množstvím předmětových hesel, jež dokument popisují,
- nedostatečné využívání řízených slovníků s pevnou strukturou, které by uživatele při vyhledávání vedly,
- nedostatek přívětivosti vůči uživatelům, kteří řízený slovník užívají,

- nevhodný design grafického rozhraní pro realizaci tohoto typu vyhledávání,
- omezené množství příkladů v kurzech pro uživatele, které jsou zaměřeny na zlepšení jejich schopností používat on-line knihovní katalog.

[VILLÉN-RUEDA, 2007].

5.8.5. Fulltextové vyhledávání

Jedná se o vyhledávání informací prohledáváním **databáze plnotextově registrovaných zdrojů** slovy, které jsou obsaženy přímo v dokumentu, prostřednictvím jednoduchého rozhraní nebo metodami pokročilého vyhledávání (buď volbou příslušné funkce z menu nebo přímým zápisem dotazu) [TKAČÍKOVÁ, 2005]. Při používání této metody vyhledávání je nutné přihlídnout k subjektivním aspektům věcného zpracování, neboť na úrovni textu existuje různá terminologie, prezentace pojmů a ideologická východiska. Pokud chce uživatel vyhledávat ve **vícejazyčné dokumenty**, je nutné formulovat několik dotazů. Plnotextové vyhledávání řeší pokročilé metody založené na statistických modelech textu např. latentní sémantické indexování¹⁹.

Problémy s **tvorbou slov při zadávání dotazu** lze účinně řešit krácením, tedy použitím zástupných znaků pravostranného rozšíření * a ?, někdy to však může být zavádějící např. sen/sny/snová/snivec. V takových případech je třeba dobře zvážit použití zástupných znaků. Některé vyhledávače pomáhají svým uživatelům s vytvářením výrazu pro hledání, tito pomocníci se odborně nazývají derivátory slovních tvarů a užívá je např. Parlamentní knihovna, sněmovní tisky, <http://www.psp.cz/sqw/text/dotaz2.sqw?search=1&od=1998&do=denes>.

Dalším aspekt jazyka, který je třeba zohlednit při formulaci dotazu, je **synonymie a homonymie**. Ta je na obecné úrovni obtížně řešitelná, neboť konkrétní význam slov je většinou určen kontextem dotazu (tedy dalšími výrazy v dotazu). Zaměříme-li se na obecní a specifické pojmy, pak konstatujeme, že implicitní paradigmatické vztahy nemusí uživatel znát nebo si je uvědomit např. kuesta = typ vrcholu kopce. V plném textu rovněž nebývá problém s vyhledáním všech specifických výrazů. Novost pojmů je ve fulltextovém vyhledávání plně zajištěna, není třeba provádět aktualizaci souboru hesel jako u předmětového vyhledávání. Jen je nutno zohlednit různou terminologii. Fulltextové vyhledávání je výhodné a použitelné pro **všechny druhy uživatelů** bez rozdílu míry odborného vzdělání. Může se více osvědčit z hlediska úplnosti a v případě vyhledávání specifického výrazu zvyšuje přesnost. K zadávání

¹⁹ Latentní sémantické indexování je statistická metoda vhodná pro odvození podobnosti mezi dokumenty a slovy [Wikipedia, termín latent semantic indexing].

termínů dotazu ve fulltextovém vyhledávání je možné využít jak klíčová slova, tak slova z řízeného slovníku. Jejich využití je třeba zvažovat podle konkrétních potřeb uživatele. Ideálním se jeví použití hybridních systémů používajících plný text i řízený slovník či rozšiřování uživatelského dotazu pomocí řízeného slovníku [SCHWARZ, 2006].

Výhody jsou:

- nezávislost na indexátorovi, stroj si přebere sám velké množství vyhledatelných slov,
- možnost použití přirozeného jazyka,
- snadné použití pro uživatele.

Nevýhody jsou:

- vyhledání jen zadaných termínů, bez vazby na příbuzná slova,
- je nutné znát tvary slov ve všech jazycích,
- v češtině navíc v souvislosti s cizojazyčnými vyhledávači vzniká problém skloňování a časování,
- nedostatky sémantického webu,
- nemožnost využití na naskenované materiály v rámci retrospektivního zpracování.

5.8.6. Závěr

Hledání informací není spojeno s užíváním knihovny, ale spíše s vůlí zjistit umístění dokumentu v ní. Takže existuje spojení mezi dvěma aktivitami. To znamená, že on-line knihovní katalog je dobře konsolidovaný nástroj pro přístup do klíčových informací, jeho role postupně zesiluje s rostoucím objemem informací dostupných v elektronické podobě. Fakt, že uživatel konzultuje s katalogem častěji pomocí prohlížení potvrzuje, že on či ona nemá úplně jasno jakou informaci přesně hledá. Jestliže kognitivní infrastruktura není adekvátní, hledání informací povede k modelu pokus-omyl, bez ohledu na možnosti a dobrou informovanost systému [VILLÉN-RUEDA, 2007]. Před volbou metody vyhledávání je vhodné zvážit základní vlastnosti databáze, ve které uživatel vyhledává. Jedná se o odpovědi na následující otázky

- Jaký prostor např. i internetu daná služba prohledává?
- Jaká je velikost její databáze?

- Jaký používá způsob indexování?
- Jaké nabízí způsoby hledání?
- Jaké nabízí možnosti prezentace výsledků vyhledávání?

Znalost těchto informací o vyhledávacích službách je základem pro volbu vhodného nástroje nejlépe odpovídajícího hledanému tématu [TKAČÍKOVÁ, 2005]. Fugmann tvrdí, že volné (neřízené) a extraktivní indexování a indexování v systémech s plnými texty je výhodnější jenom v případě, že přirozený jazyk poskytuje specifické výrazy pro pojmy, které nás zajímají, nebo v případě, že je pro uživatele, který vyhledává, umí předvídat v okamžiku, kdy sestavuje rešeršní dotaz. V případě, že se jedná o případy, kdy jsou pojmy vyjádřeny pomocí parafráze, je vyhledávání pomocí řízených slovníků efektivnější než jiné typy vyhledávání [FUGMANN, 1993].

5.9. Budoucnost pořádacích systémů

Známá knihovnická třídění a pořádací systémy budou bezpochyby figurovat i v novém informačním prostředí. Přibudou k nim ale patrně i další systémy pořádaní, které mají rysy třídění, ale od knihovnických systémů se liší nejen v rozsahu, ale i předmětem. Mnohé z těchto systémů jsou založeny spíše pojmově než dokumentově - známými příklady jsou chemické vzorce a taxonomie v biologických vědách. V nedávných desetiletích došlo k pronikání myšlenek třídění a kategorizace do on-line databází. Např. průzkum databází ve vyhledávacím systému Dialog odhalil, že velké procento jich používá určitý druh třídění nebo kategorizace, od velmi širokých skupin až po velmi podrobné. Mnohé z nich se ani nenazývají třídění, ale odkazuje se na ně např. jako na **pojmové nebo deskriptorové kódy nebo kódy kategorií**, ale mají při tom vlastnosti třídění, pokud jde o jejich logickou strukturu, a mnohé z nich mají i notaci. Známými příklady zde jsou kódy skupin v databázi ERIC (Education Resources Information Center), stromová struktura v MEDLINE (Medical Literature Analysis and Retrieval System Online), biosystematické kódy v BIOSIS, ale i Mezinárodní patentové třídění, Mezinárodní třídění nemocí, třídění enzymů nebo fasetové třídění zemědělského a lesnického výzkumu používaného v CRIS (Current Research Information System) [CHAN, 1995]. Při současném obrovském množství informací budou uživatelé potřebovat pomoc při výběru a hodnocení zdrojů informací. Nadbytek informací a jejich nespolehlivost se stávají stále závažnějším problémem. Už nestačí jen najít relevantní informace. **Uživatelé také potřebují pomoc** při určování toho, co je v nich nejlepší nejužitečnější. Nyní jsme uprostřed nebo spíše na začátku informační exploze, která je daleko mocnější, než byla ta po druhé světové válce. Aby to uživatele neodradilo od cesty vůbec, ale aby je to navedlo správným

směrem – je v moci věcného pořádní informací. Když pohlížíme do budoucnosti nepotřebujeme se ptát, zda existuje potřeba věcného pořádní informací – již víme, že ano.

6. Závěr

Změny v tradičním vnímání šedé literatury a zejména vysokoškolských kvalifikačních prací byly vyvolány zejména změnou akademického prostředí. Elektronické publikování jim dává nový rozměr. Je třeba zaměřit se na lepší věcný popis obsahu a následné vyhledávání v databázích tohoto typu dokumentů. Současná nejaktuálnější otázka je okolo tvorby repozitářů, kde se mohou elektronické diplomové a disertační práce skladovat, jejich přístupování na základě vyhledávání pokud možno v jazyce nejbližším uživateli.

Podíváme-li se do minulosti, již na srovnání příkladů záznamů ze systému SIGLE a v STK je patrný detailnější věcný popis předmětovými hesly. Vznik databáze SIGLE byl od počátku motivován možností zařazení částí národních bibliografií mezi standardní nabídku databázových středisek pracujících na ryze komerční bázi. To sice na jedné straně otvíralo jistou možnost ekonomického zhodnocení fondů, ale na druhé straně se jednalo o typicky zprostředkované zpřístupnění předpokládající znalost hesla a účtovacího konta a realizované prakticky výhradně informačními specialisty nebo pracovníky knihoven. Širší odborná veřejnost tak s vlastní bází SIGLE prakticky nepracovala, pouze mohla dostat v té či oné formě výsledky rešerší.

Dnes Síťová knihovna NDLTD podporuje individuální instituce v zakládání jejich vlastních databází elektronických diplomových a disertačních prací. Stejně tak jako propagaci užívání těchto prací přes webová rozhraní a portály či agregátory individuálních stránek a vytváření jednoduchého vstupu do databáze NDLTD pro koncové uživatele. Univerzity v USA ve snaze změnit zavedenou tradici, začaly zveřejňovat disertační práce na svých vlastních serverech zdarma a pro celý svět a připojily se tak k federaci NDLTD, což nebude, zdá se, vůbec jednoduché. A to i přes to, že existují zjevné přednosti takového systému, které jsou formulovány v jeho cílech a funkcích.

Program ADT představuje novou roli knihoven, kterou je tvorba a zpřístupnění národního systému disertačních prací. Vzhledem k aktuálnímu výzkumu probíhajícímu v jednotlivých institucích lze konstatovat, že význam tohoto systému stále roste. Role knihoven je tak posílena – neumožňují pouze přístup k informacím vytvořených třetí stranou, ale samy se stávají vydavateli.

Toto si uvědomili i knihovníci ve Státní technické knihovně. V rámci selekčních jazyků STK je používán Polytematický strukturovaný heslář, který má velké ambice na své uplatnění i dnešních trendech. Hlavní myšlenky koncepce PSH plně korespondují s hlavními myšlenkami zpřístupňování a vyhledávání v databázích šedé literatury. S pomocí moderních technologií, např. vytvořením animované nápovědy ve formátu flash, by bylo možné ušetřit čas pracovníků a autoři by tak mohli být názorně a jednoduše seznámeni s funkcí a přiřazováním předmětových hesel. V souvislosti s metodami vyhledávání v dnešním heterogením prostředí se jeví jako nejlepší, v ohledu na uživatele, využití sloučení možností hledání jednak pomocí klíčových slov, jednak předmětových hesel vybraných z řízeného slovníku a fulltextového prohledávání dokumentu. Řízený slovník v tomto maximalistickém přístupu nesmí obsahovat řetězce a naopak je žádoucí detailnější struktury, což PSH splňuje. Praxe navíc ukázala, že při vyhledávání v obsáhlejší databázi nebo v souborných katalozích, použití předmětových hesel a booleových operátorů výrazně šetří čas ve všech ohledech (zatíženost strojů, i čas uživatelů). Z pohledu budoucnosti celosvětového propojení, jak naznačuje ve svých programech IFLA, je žádoucí minimálně bilinguálnost databází, čemuž je PSH schopno rovněž dostát. Tato práce rozhodně nepopírá jasný trend fulltextového vyhledávání. V současnosti však není vyřešeno spousta problémů spojených např. s vyhledáváním v tištěných dokumentech a naskenovaných dokumentech v rámci retrokonverze (nutno přidat do záznamu minimálně abstrakt a předmětová hesla). Nicméně zatím čekáme na jeho zdokonalení.

Po zvážení všech výhod a nevýhod předložených v této práci se jako nejlepší řešení pro projekt Digitální knihovny šedé literatury jeví povinné použití věcného pořádku informací v čele s Polytematickým strukturovaným heslářem a na přechodnou dobu, než bude zdokonaleno fulltextové prohledávání, maximalistický přístup ve vyhledávání.

Seznam grafů

Graf 1: Složení uživatelů ETD v USA [McMILLAN, 1999]

Graf 2: Složení uživatelů ETD v mezinárodním pojetí [McMILLAN, 1999]

Graf 3: Záznamy získané přímo od členů [RUTHVEN, 2006]

Graf 4: Počet univerzit podle procentuálního zastoupení [ADT, 2006].

Seznam tabulek

Tab. 1: Záznamy SIGLE z pohledu geografického k roku 2003 [BRATKOVÁ, 2003]

Tab. 2: Záznamy SIGLE z pohledu obsahového k roku 2003 [BRATKOVÁ, 2003]

Tab. 3: Počet záznamů přiřazených jednotlivým předmětovým kategoriím [SIGLE, 2005]

Tab. 4: Předmětová struktura SIGLE Subject Category List

Tab. 5: Příklad 1

Tab. 6: Příklad 2

Tab. 7: Příklad 3

Tab. 8: Statistika přístupů do databáze elektronických diplomových prací a disertací na Technické univerzitě ve Virginii [HUSSEIN, 2001].

Tab. 9: Příklad 1

Tab. 10: Příklad 2

Tab. 11: Příklad 1 Záznam č. **132900**

Tab. 12: Příklad 2 **Záznam č. 133425**

Tab. 13: Příklad katalogizačního záznamu [FRANK; ROWE, 2004].

Tab. 15: Celkové prohlížení rejstříků v STK v období 16. 6. 2006 – 2. 8. 2007

Tab. 16: Prohlížení rejstříků předmětových hesel z jiných institucí v období 16. 6. 2006 – 2. 8. 2007

Seznam obrázků

Obr. 1: Vyhledávání na základě rozdělení do tříd [GONCALVES, 2000].

Obr. 2: Postup vyhledávání v MARIAN [GONCALVES, 2000].

Obr. 3: Architektura sdruženého katalogu [GONCALVES, 2000].

Seznam použité literatury

V uvedeném seznamu jsou zahrnuty jednak položky, které se týkají přímo věcného pořádní šedé literatury, jednak další tematicky příbuzné položky, které byly využity při zpracovávání diplomové práce. Jsou v něm uvedeny záznamy dokumentů publikovaných i nepublikovaných (tj. rukopisů). Celkově obsahuje 86 položek.

Záznamy jsou uspořádány na první úrovni abecedně podle primární autorské odpovědnosti, na druhé úrovni chronologicky. V případě, že u daného dokumentu nelze určit primární autorskou odpovědnost, je dokument zařazen podle názvu. V záznamech po uvedení primární autorské odpovědnosti následuje časové určení (s případným dalším doplňkem formou malého písmena abecedy sloužícího k rozlišení jednotlivých dokumentů od stejného autora ze stejného roku). Záznamy jsou v souladu s ČSN ISO 690 a ČSN ISO 690-2.

1. ADT. 2000. *Metadata standard* [online]. Canberra : Council of Australian University Librarians, c1997, updated 2006-03-24 [2007-08-08]. Dostupný na World Wide Web: <http://adt.caul.edu.au/downloads/ADT_Metadata_V6.doc>.
2. ADT. 2006. *ADT Membership survey 2006* [online]. ADT Major Documents, 14/06/07 [cit. 2007-08-08]. Dostupný na World Wide Web: <<http://www.caul.edu.au/surveys/adt2006.doc>>.
3. BALÍKOVÁ, M. 2001. Problematika věcného pořádní informací a jejich zpřístupnění. *Národní knihovna* [online]. 2001, roč. 12, č. 3 [cit. 2007-06-29], s. 175-186. Dostupný také na World Wide Web: <<http://full.nkp.cz/nkkr/Nkkr0103/0103175.html>>. ISSN 1214-0678.
4. BAYER, J. 1995. Pokyny pro ukládání dat SIGLE v textovém formátu. 13. 7. 1995. 12 s. Rukopis.
5. BAYER, J.; SKÁLOVÁ, H. 1995. Česká účast v evropském informačním systému o šedé literatuře. *Infocus*. 1995, roč. 1, č. 12, s. 305-307. ISSN 1211-0692.
6. BLOCH, J. (překladatel). 1964. Věda, vláda a informace : odpovědnost technické veřejnosti a vlády za předávání informací : zpráva vědeckého poradního výboru prezidenta. Praha : Státní technická knihovna, 1964, l. 67.

7. BRATKOVÁ, E. 2002. Elektronické disertace a projekty a systémy jejich zpřístupňování. *Národní knihovna* [online]. 2002, roč. 13, č. 4 [cit. 2004-06-29], s. 268-287. Dostupný na World Wide Web: <http://full.nkp.cz/nkkcr/pdf/0204/0204268.pdf> nebo <http://full.nkp.cz/nkkcr/Nkkcr0204/0204268.html>. ISSN 1214-0678.
8. BRATKOVÁ, E. 2003a. Technické a technologické otázky budování systémů elektronických disertací. In *Moderní informační a komunikační technologie v knihovnictví 2003 : sborník příspěvků*. Praha : Státní technická knihovna, 2003, s. 31-38. ISBN 80-86504-09-3.
9. BRATKOVÁ, E. 2003b. *Systémy pro registraci šedé literatury : výběr. Pracovní materiál pro potřeby výuky posluchačů ÚISK FF UK*. Verze 4.3 (aktualiz. a doplň.). Praha : ÚISK FF UK, květen 2003. 25 s. Interní nepublikovaný text.
10. BRATKOVÁ, E. 2004. Systémy pro zpřístupňování elektronických vysokoškolských kvalifikačních prací v České republice : od prvních lokálních systémů k platformě jejich dalšího společného rozvoje. In *Elektronický zborník príspevkov z 11. medzinárodného seminára CASLIN 2004 konaného v dňoch 5.-9. septembra 2004 na Čingove (Slovenský raj)* [online]. Prešov : Univerzitná knižnica Prešovskej univerzity, 2004 [cit. 2004-10-05]. Dostupný na World Wide Web: <http://www.pulib.sk/elpub/UK/Caslin04/bratkova.pdf>.
11. BRATKOVÁ, E.; MACH, J. 2006. Standardizace metadat pro národní registr EVŠKP. In MIKULECKÁ, J.; SEDLÁČEK, J. (ed.). *Digitální knihovny : provoz a budování : sborník konference EUNIS-CZ, Špindlerův Mlýn, 28. až 30. května 2006*. Vyd. 1. s přílohou CD. Hradec Králové : Gaudeamus, 2006, s. 5-12. Dostupný na World Wide Web: <http://eprints.rclis.org/archive/00006594>. ISBN 80-7041-200-3.
12. CAMPBELL, D. 2000. *Australian subject gateways - metadata as an agent of change*. [online]. Mooroolbark : VALA - Libraries, Technology and the Future Inc. c2000 [cit. 2007-08-08]. Dostupný na World Wide Web: <http://www.vala.org.au/vala2000/2000pdf/Campbell.PDF>.

13. CARGNELUTTI, T. 2004. Indexing and accessing electronic theses and dissertations: some concerns for users. *Electronic theses and dissertations : a sourcebook for educators, students, and librarians*. New York : Marcel Dekker, 2004, s. 355-359. ISBN 0-8247-0973-X.
14. COMBEROUSSE, M. 1993. La littérature grise. *Le Bulletin des bibliothèques de France*. 1993, 38, no. 5, s. 60-62. ISSN 0006-2006.
15. ČESKO. 2000. Zákon č. 121 ze dne 7. dubna 2000 o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (Autorský zákon). In *Sbírka zákonů, Česká republika*. 2000, částka 36, s. 1658-1685. Dostupný také na World Wide Web: <<http://www.mvcr.cz/sbirka/2000/sb036-00.pdf>>.
16. ČESKO. 2005. Zákon ze dne 20. prosince 2005, kterým se mění zákon č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a některé další zákony. In *Sbírka zákonů, Česká republika*. 2005, částka 188, s. 10766-10773. Dostupný také na World Wide Web: <<http://web.mvcr.cz/sbirka/2005/sb188-05.pdf>>.
17. ČSN 01 0188. *Tvorba předmětových hesel*. Praha : Vydavatelství Úřadu pro normalizaci a měření, 1982. Schválena 20. 5. 1982, účinnost od 1. 4. 1983. 38 s.
18. ČSN 01 0193. *Pokyny pro vypracování a rozvíjení jednojazyčných tezaurů*. Praha : Český normalizační institut, 1995. 49 s.
19. ČSN ISO 5127-3a. *Informace a dokumentace. Slovník. Část 3a, Akvizice, identifikace a analýza dokumentů a dat*. Praha : Český normalizační institut, 1993. Účin. květen 1993. 45 s.
20. DAHLBERG, I. 1974. Grundlagen universeller Wissensordnung Probleme und Möglichkeiten eines universalen Klassifikationssystemes des Wissens. München : Verlag Dokumentation, 1974, s. 366. ISBN 3-7940-3623-9.

21. DAY, S. 1992. SIGLE on CD-ROM. *Aslib Information*. 1992, vol. 20, no. 7-8, s. 298. ISSN 0305-0033.
22. DEBACHERE, M. C. Problems in Obtaining Grey Literature. *IFLA Journal*. 1995, vol. 21, no. 2, s. 94-98. Pro oprávněné uživatele dostupné také na World Wide Web: <<http://ifl.sagepub.com/cgi/reprint/21/2/94>>. ISSN 0340-0352. DOI: 10.1177/034003529502100205.
23. DRABENSTOTT, K. M. 2003. Do Nondomain Experts Enlist the Strategie of Domain Experts? *Journal of the American Society for Information Science and Technology*, 2003, 54, s. 836-854.
24. EATON, J. L. 2004. Enhancing graduate education through elektronik theses and dissertations. *Electronic theses and dissertations : a sourcebook for educators, students, and librarians*. New York : Marcel Dekker, 2004, s. 1-7. ISBN 0-8247-0973-X.
25. EDT digital library: Networked digital library of theses and dissertations (NDLTD) [online]. Blacksburg (VA, USA) : Virginia Tech, [199-]-. Dostupné na World Wide Web: <<http://www.nldtd.org/>>.
26. ENGINEERS JOINT COUNCIL (editor). 1969. *Thesaurus of engineering and scientific terms*. Washington : American Association of Engineering Societies, 1969, s. 690. ISBN: 978-0685092897.
27. FISHER, S. 1995. Access to information. *Management Decision*, 1995, vol. 33, no. 5. s. 22-28. ISSN 0025-1747.
28. FOX, Edward A. et al. (ed.). 2004. *Electronic theses and dissertations : a sourcebook for educators, students, and librarians*. New York; Basel : Marcel Dekker, c2004. xv, 414 s. Books in library and information science, vol. 65. ISBN 0-8247-0973-X.
29. FRANK, I.; ROWE, W. C. 2004. Indexing and accessing electronic theses and dissertations : some concerns for users. *Electronic theses and dissertations : a*

- sourcebook for educators, students, and librarians*. New York : Marcel Dekker, 2004, s. 343-353. ISBN 0-8247-0973-X.
30. FUGMANN , R. 1993. *Subject analysis and indexing : theoretical foundation and practical advice*. Frankfurt/Main : Indeks Verlag, 1993. XVI, 205 s. Textbooks for knowledge organization, vol. 1. ISBN 3-88672-500-6.
31. GENONI, P.; COWAN, R. 2003. Bibliographic control of Australian higher degree theses : the future role of the Australian digital theses program. *Australian Academic & Research Libraries*. 2003, vol. 34, no. 2. Dostupné také na World Wide Web: <<http://www.alia.org.au/publishing/aarl/34.2/full.text/genoni.cowan.html>>.
32. GONCALVES, A. et al. 2000. *MARIAN : searching and querying across heterogeneous federated digital libraries*. [online]. Aktualizováno 12. prosince 2000 [cit. 2007-07-30]. Dostupné z Worl Wide Web: <<http://docs.ndltd.org:8080/dspace/handle/2340/58>>
33. HAGEN, J. H.; DOBRATZ, S.; SCHIRMBACHER, P. 2003. Electronic theses and dissertations worldwide : highlights of the ETD 2003 symposium. *D-lib Magazine* [online]. 2003, vol 9, no. 7/8. [cit. 2007-07-30]. Dostupné na World Wide Web: <<http://www.dlib.org/dlib/july03/hagen/07hagen.html>>. ISSN 1082-9873.
34. HOOVER, L. 2005. A Biginners' Guide for Subject Analysis of Theses and Dissertations in the Hard Sciences. *Cataloging & Classification Quarterly*, 2005, vol. 41, no. 1, s. 133-159. ISSN 0163-9374.
35. HORÁKOVÁ, L.; SKÁLOVÁ, H. 1995. SIGLE : evropská databáze pro šedou literaturu. *Informace*. 1995, č. 1, s. 4-5. ISSN 1210-8502.
36. CHAN, L. M. 1995. Classification, present and future. *Cataloging & Classification Quarterly*. 1995, vol. 21, no. 2. s. 5 – 17. ISSN 0163-9374.
37. JONÁK, Z. 1998. Inteligentní nástroje pro práci s texty na internetu. *Ikaros* [online]. 1998, č. 9. Dostupný na World Wide Web: <<http://ikaros.ff.cuni.cz/1998/c09/nastroje.htm>>.

38. JONÁK, Z. 2001. Omezení a možnosti zvýšení selekčních schopností internetových robotů. *Daidalos* [online]. 2001, č. 1/2. Dostupný na World Wide Web: <http://daidalos.ff.cuni.cz/2001/leden/zj_sj.php>.
39. KATZ, S. 2004. Innovative hypermedia ETDs and employment in the Humanities. *Electronic theses and dissertations : a sourcebook for educators, students, and librarians*. New York : Marcel Dekker, 2004, s. 9-18. ISBN 0-8247-0973-X.
40. KOVÁŘ, B. 1976. Problémy teorie procesu věcného pořádní informací a selekčních jazyků. Praha : Univerzita Karlova, 1976. 165 s.
41. LAFFERTY, S.; EDWARDS, J.; DOVEY, K. 2005. The Australian Digital Theses Program : a disruptive technology? In: *Educause Australasia Auckland, 5. – 8. 4. 2005* [online]. Auckland : University of Auckland, 2005. [cit. 2007-08-06]. Dostupné na Word Wide Web: <<http://www.educause2005.auckland.ac.nz/interactive/papers/C14.pdf>>.
42. LARSON, R. R. 1991. Classification Clustering, Probabilistic Information Retrieval and the Online Katalog. *The Library Quarterly*, 1991, 61, April, s. 133-173. ISSN 0024-2519.
43. LUZZI, D. 2000. Trends and evolution in the development of grey literature : a review. *International Journal of Grey Literature*. 2000, vol. 1, issue 3, s. 106-117. ISSN 1466-6189
44. MAEDA, A. et al. 1998. Viewing Multilingual Documents on Your Local Web Browser. *Communications of the ACM*. 1998, vol. 41, no. 4, s. 64-65. ISSN 0001-0782.
45. *MARC 21 concise formats*. 2006 concise ed. 2006. Washington D.C. : Library of Congress, 2007-10-05 [cit. 2007-07-30]. Dostupné na World Wide Web: <<http://www.loc.gov/marc/concise/concise.html> >.

46. MARCHIONINI, G. 1995. *Information seeking in Electronic Environments*. Cambridge, NY : Cambridge University Press, 1995, 215 s. ISBN 0-521-58674-7.
47. MATTHEWS, G. S.; LAWRENCE, G.S.; FERGUSON, D. K. 1983. *Using Online Catalogs: a nationwide survey. A report of a study sponsored by the council on Library resources*. New York : Neal-Schuman, 1983, 255 s. ISBN 0-918-21276-6.
48. MIHULA, R. 2005. *Indexování HTML dokumentů..* Praha, 2005. iv, 89 s, 1 CDROM. Diplomová práce. Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Katedra softwarového inženýrství. Vedoucí diplomové práce RNDr. Michal Kopecký, PhD.
49. McMILLAN, G. 1999. Perspectives on electronic theses and dissertations. In *New Frontiers in Grey Literature: 4th International Conference on Grey Literature: GL'99 Proceedings* [online]. Amsterdam : GreyNet, 1999 [cit. 2006-09-18]. Dostupné z World Wide Web: <http://scholar.lib.vt.edu/staff/gailmac/presentations/GL99ETDpaper.pdf>.
50. NÁRODNÍ KNIHOVNA ČR. c2007. *Koncepce trvalého uchování knihovních sbírek tradičních a elektronických dokumentů v knihovnách ČR do roku 2010.* [online]. Aktualizováno 19. 7. 2007, [cit. 2007-08-09]. Dostupné na World Wide Web: http://knihovnam.nkp.cz/sekce.php3?page=03_deklarace.htm.
51. NDLTD. 2007. *Networked digital library of theses and dissertations membership* [online]. Aktualizace 7. července 2007 [cit. 2007-07-30]. Dostupné na World Wide Web: <http://www.ndltd.org/members/index.en.html>.
52. NOVOTNY, E. 2004. I don't Think I Click: a protocol analysis study of use of a library online katalog in the internet age. *College and Research Libraries*. 2004, 65, s. 525-537.
53. PAPÍK, R. 1995. [Šedá literatura]. *Infocus*. 1995, roč. 1, č. 5, s. 142. ISSN 1211-0892.
54. PAPÍK, R. 1996. Zdroje šedé literatury. *Infocus*. 1996, roč. 2, č. 11-12, s. 222-223. ISSN 1211-0892.

55. PAPIK, R. 2001. Vyhledávání informací II : uživatelské rozhraní a vlivy oboru "human-computer interaction". *Národní knihovna*, 2001, roč. 12, č. 2, s. 81. ISSN 0862-7487.
56. PATOU, C. 1994. Mezinárodní přístup k vědeckým a technickým informacím. *I '94*. 1994, roč. 36, č. 5, s. 131-132. ISSN 0862-9382.
57. POLLITT, A. S. 1998. The key role of classification and indexing in view-based searching. *International Cataloguing and Bibliographic Control*, 1998, 27, no.2. s. 37-40. ISSN 1011-8829
58. RIESTHUIS, G. J. A. 2004. *Sociological aspects of classification*. [online]. [cit. 2004-01-27]. Dostupné na World Wide Web: <<http://www.ifla.org/index.htm>>.
59. RUTHVEN, T. 2006. *ADT Program Report – October 2006* [online]. ADT Policy Reference Group, 20/10/06 [cit. 2007-08-08]. Dostupné na World Wide Web: <<http://www.caul.edu.au/adt/adt2006report2.doc>>.
60. ŘÍMANOVÁ, Radka. 2007. Osobní komunikace. 30. 7. 2007.
61. SCHWARZ, J. 1999. *Vývoj teorie a praxe tezaurů v České republice : nástin dějin deskriptorových selekčních jazyků v bývalém Československu se zaměřením na vývoj teoretických, metodických a normativních aspektů tvorby tezaurů*. Diplomová práce ÚISK FF UK. Praha : vlastním nákladem, 1999. IX, 121 s.
62. SCHWARZ, J. 2006. *Současné problémy věcného vyhledávání informací : výběrový seminář*. Aktualiz. verze. Pro potřeby výuky ÚISK připravil Josef Schwarz. Praha : ÚISK FF UK, 2. 11. 2006. 14 s. Text ve formátu Microsoft Power Point. Pro oprávněné uživatele dostupné na World Wide Web: <<http://dl.cuni.cz/cuni/course/view.php?id=281>>.
63. *SIGLE* [online]. Den Haag : Koninklijke Bibliotheek, 1999-2005 [cit. 2007-07-18]. Dostupné na World Wide Web: <<http://web.archive.org/web/20040218211123/www.kb.nl/infolev/eagle/frames.htm>>.

64. *SIGLE Manual. Part 1, SIGLE Cataloguing Rules*. 1997. Ed. A. Smith and M. Payne. 4th ed. Den Haag : EAGLE, October 1997. 72 s.
65. *SIGLE Manual. Part 2, Subjekt Category List, Fourth Edition*. 1999. Compil. M. Payne. 4th ed. Den Haag : EAGLE, April 1999. 136 s.
66. SILHONEN, A.; VAKKARI, P. 2004. Subject Knowledge Improves Interacti ve Query Expansion Assisted by a Thesaurus. *Journal of Documentation*. 2004, 60, s. 673-690. ISSN 0022-0418.
67. SKÁLOVÁ, H. 1995. EAGLE/SIGLE : evropská báze šedé literatury. *Informace*. 1995, č. 2, s. 6. ISSN 1210-8502.
68. SKOLKOVÁ, Linda. 2007. *Polytematický strukturovaný heslář (Polythematic Structured Subject Heading System)*. Praha, 2007. iv, 172 s, 27 s. příl. + 1 CDROM. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce Ing. Miloslav Nič, PhD.
69. SLAVÍČKOVÁ, E. 1988. Problematika víceslovných spojení v tezaurech. In *Lingvistické metody a automatizované informační systémy*. Praha : Dům techniky ČSVTS, 1988, s. 12-20.
70. Státní technická knihovna (Praha, Česko). 1999. *Výroční zpráva Státní technické knihovny 1999*. Odp. red. J. Bayer. Praha : STK, 2000. 41 s.
71. Státní technická knihovna (Praha, Česko). 2002. *Výroční zpráva Státní technické knihovny 2001* [online]. Praha : Státní technická knihovna, 2002 [cit. 2007-03-11]. 74 s. Dostupné na World Wide Web: <http://www.stk.cz/vyr_z-01.pdf>. ISBN 80-86504-06-9.
72. Státní technická knihovna (Praha, Česko). 2007. *Příhláška návrhu programového projektu Ministerstva kultury ČR na rok 2008*. Praha, 13. 7. 2007. 46 s., 2 s. Rukopis.
73. SULEMAN, H. 2001. Networked digital library of theses and dissertations. *D-lib Magazine* [online]. 2001, vol 7, no. 9. [cit. 2007-08-08]. Dostupné na World Wide

- Web: <<http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html>>.ISSN 1082-9873.
74. SULEMAN, H. et al. 2004. Network digital library of theses and dissertations. *Electronic theses and dissertations :a sourcebook for educators, students, and librarians*. New York : Marcel Dekker, 2004, s. 49-72. ISBN 0-8247-0973-X.
75. Systémy pořádání informací. *I '94*. 1996, roč. 38, č. 6, s. 167-168. ISSN 0862-9382.
76. ŠILHÁNEK, J.; NOVOTNÁ, M. 1996. SIGLE - evropská báze dat šedé literatury z pohledu uživatele. *Infocus*. 1996, roč. 2, č. 4, s. 75-77. ISSN 1211-0892.
77. *TDKIV : Česká terminologická databáze z oblasti knihovnictví a informační vědy* [online databáze]. 2003- . Praha : Knihovnický institut NK ČR, 2003- [cit. 2007-08-08]. Databáze vznikla v letech 2001-2002 v rámci projektu podpořeného grantem MK ČR. Dostupná na World Wide Web: <<http://sigma.nkp.cz/cze/ktd>>.
78. TKAČÍKOVÁ, D. 2005. *Jak pracovat s webovými vyhledávacími službami*. [online]. Ostrava : Univerzitní knihovna Vysoké školy báňské – Technické univerzity Ostrava, Aktualizováno 2005-06-18 [cit. 2007-08-08]. Dostupné na World Wide Web: <<http://knihovna.vsb.cz/kurzy/hledani/01.html>> .
79. UNESCO, 2001. *The Guide for electronic theses and dissertations* [online]. Paris : UNESCO, c2001 [cit. 2007-07-22]. Dostupné na World Wide Web: <<http://etdguide.org>>.
80. VILLÉN-RUEDA, L.; SENSO, J. A.; DE MOYA-ANEGÓN, F. 2007. The Use of OPAC in a Large Academic Library: a transactional log analysis study of subject searching. *The Journal of Academic Librarianship*, 2007, vol. 33, no. 3. s. 327-337. ISSN 0099-1333.
81. VORÁČEK, Josef. *Tvorba tezaurů v českém jazyce*. Praha : ÚVTEI - STK, 1974, s. 155.

82. WESSELS, Robert H. A.; HANKOVÁ, Jelena. 1995. Interview s Robertem H. A. Wesselsem. *Infocus*. 1995, roč. 1, č. 12, s. 308. ISSN 1211-0692.
83. WESSELS, Robert. 1994. Význam mezinárodní spolupráce pro dostupnost šedé literatury. *I '94*. 1994, roč. 36, č. 6, s. 149-151. ISSN 0862-9382.
84. WOOD, D. N.; SMITH, A. W. 1991. SIGLE – model mezinárodní spolupráce? *I '91*. 1991, 33, č. 8/9, s. 239-241. ISSN 0862-9382.
85. WYLY, B. J. 1996. From Access Points to Materiále: a transaction log analysis of access point value for online katalog users. *Library Resources and Technical Services*. 1996, 40, s. 211-236.
86. *WIKIPEDIA : the free encyklopedia* [online databáze]. 2001- . San Diego : Wikimedia foundation, 2001- [cit. 2007-08-08]. Dostupná na World Wide Web: <<http://en.wikipedia.org>>.

Přílohy

Příloha č. 1 : Weinbergova zpráva

Příloha č. 2 : Srovnání hesel SIGLE a hesel PSH

Příloha č. 3 : Příklady Advanced Search

Příloha 1 Srovnání hesel SIGLE a hesel PSH

Názvy tříd z pořádní SIGLE (COSATI) v angličtině	Názvy tříd z pořádní SIGLE (COSATI) v češtině (zpracováno s využitím podkladů J. Bayera)	Odpovídající hesla z PSH (verze 2.1) v češtině	Odpovídající hesla z PSH (verze 2.1) v angličtině
Aeronautics	Aeronautika	Letecká doprava do	Air transport
Agriculture, Plant & Veterinary Sciences	Zemědělství, rostlinstvo a zvěrolékařství	Zemědělství ze	Agriculture
		Rostliny bi	Plants
		Veterinářství ze	Veterinary medicine
Environmental Pollution, Protection & Control	Znečištění životního prostředí, ochrana a řízení	Znečištění životního prostředí bi	Environmental pollution
		Ochrana životního prostředí bi	Environmental protection
		Řízení ev	Controlling
Humanities, Psychology & Social Sciences	Humanitní vědy, psychologie a sociální vědy	Humanitní vzdělání pe	Education in humanities
		Psychologie ps	Psychology
		Společenské vědy ve	Social Sciences
Biological & Medical Sciences	Biologie a lékařství	Biologie bi	Biology
		Lékařství zd	Medicine
Chemistry	Chemie	Chemie ch	Chemistry
Earth & Atmospheric Sciences	Vědy o zemi a atmosféře	Země gf	Earth
		Atmosféra gf	Atmosphere (geophysics)
Electronics & Electrical Engineering, Computer Science	Elektronika a elektrotechnika, výpočetní technika	Elektronika el	Electronics
		Elektrotechnika et	Electrotechnics
		Informatika if	Informatics
Energy & Power	Energie a síla	Energetika en	Power engineering
Materials	Materiály	Materiály ob	Materials
Mathematical Sciences	Matematické vědy	Matematika ma	Mathematics
Mechanical, Industrial, Civil & Marine Engineering	Strojírenství, řízení výroby, stavebnictví a lodní stavitelství	Strojírenství sr	Mechanical engineering
		Řízení ev	Controlling
		Stavebnictví st	Civil Engineering
		Vodní stavby vo	Hydraulic structures
Methods & Equipment	Výrobní postupy a zařízení	Metody ob	Methods
		Zařízení ob	Equipment
Military Science	Vojenství	Vojenství vv	Military affairs

Missile Technology	Raketová technologie	Raketové zbraně vv	Rocket weapons
Navigation, Communication, Detection & Countermeasures	Navigace, spojení, detekce a protiopatření	Navigační systémy sj	Navigation systems
		Spoje sj	Communications
		Detekce signálu if	Detection of signal
		Rušení sj	Interference (communication)
Ordnance	Arzenály	Zbraně vv	Weapons
Physics	Fyzika	Fyzika fy	Physics
Propulsion & Fuels	Pohon a paliva	Elektrické pohony et	Electric drives
		Pohonné hmoty do	Fuels (transport)
Space Technology	Kosmické technologie	Vesmír as	Universe
		Technologie ve	Technology

Příloha č. 3 : Příklady advanced search

NDLTD <http://www.ndltd.org/serviceproviders/scirus>

Advanced Search

All of the words in Title

AND

Any of the words in Abstract

AND

Exact phrase in Keywords

Only show results published between

1900 and 2012

Only show results in

<input checked="" type="checkbox"/> All subject areas	<input type="checkbox"/> Agriculture and Biological Sciences
<input type="checkbox"/> Astronomy	<input type="checkbox"/> Chemistry and Chemical Engineering
<input type="checkbox"/> Earth and Planetary Sciences	<input type="checkbox"/> Economics, Business and Management
<input type="checkbox"/> Engineering, Energy and Technology	<input type="checkbox"/> Environmental Sciences
<input type="checkbox"/> Languages and Linguistics	<input type="checkbox"/> Law
<input type="checkbox"/> Life Sciences	<input type="checkbox"/> Materials Sciences
<input type="checkbox"/> Mathematics	<input type="checkbox"/> Medicine
<input type="checkbox"/> Neuroscience	<input type="checkbox"/> Pharmacology
<input type="checkbox"/> Physics	<input type="checkbox"/> Psychology
<input type="checkbox"/> Social and Behavioral Sciences	<input type="checkbox"/> Sociology

Only search in the following collection

<input checked="" type="checkbox"/> ETDs	<input type="checkbox"/> The Scholarly Web
--	--

Advanced Search

ADT Distributed Database

Search the complete ADT Distributed Database

All fields	<input type="checkbox"/> Includes word/s anywhere	<input checked="" type="checkbox"/> AND <input type="checkbox"/> OR	
Title	<input type="checkbox"/> Includes word/s anywhere	<input checked="" type="checkbox"/> AND <input type="checkbox"/> OR	
Author	<input type="checkbox"/> Includes word/s anywhere	<input checked="" type="checkbox"/> AND <input type="checkbox"/> OR	
Subject words	<input type="checkbox"/> Includes word/s anywhere	<input checked="" type="checkbox"/> AND <input type="checkbox"/> OR	
School/Dept.	<input type="checkbox"/> Includes word/s anywhere	<input checked="" type="checkbox"/> AND <input type="checkbox"/> OR	
Date	<input type="checkbox"/> Includes word/s anywhere	<input checked="" type="checkbox"/> AND <input type="checkbox"/> OR	

Digital only

The ADT Distributed Database is powered by **HOTmeta**

Advanced Search

Free text search

- Search for words
 Search for phrase or word forms

Fielded Search

Author:

Drop the accent to search for authors with accented characters in their name, e.g. use Bjork to search for an author named Björk.

Keywords and Title:

- Search for words
 Search for phrase or word forms

JEL-Code:

New or updated items

Modified last

Date is Creation/revision

- of item
 of metadata

The words in each textbox are NEARed by default unless you search for a phrase or word forms. Use OR, AND or NEAR together with parenthesis to denote that combinations of words should be present.

[Tips for searching](#)

Search:

- Working Papers
 Journal Articles
 Books & Chapters
 Software Items
written in

Any Language

- Registered Authors
 Online Items Only

Return

a maximum of

hits in pages of

Sort by

- Rank
 Date modified

Combine terms by

- ANDing textboxes
 ORing textboxes

MATHDISS DATABASE SEARCH ENGINE

Advanced Search

Category:	<input type="text" value="Text"/>	Word(s):	<input type="text"/>
<input type="checkbox"/> AND	Category:	<input type="text" value="Title"/>	Word(s): <input type="text"/>
<input type="checkbox"/> AND	Category:	<input type="text" value="Keywords"/>	Word(s): <input type="text"/>
Display:	<input type="text" value="Title"/>	Sort:	<input type="text" value="by Title"/>
<input type="button" value="Search"/>		<input type="button" value="Reset"/>	

There is also a [simple search](#) available.

Notes:

- ◆ This service can only be used from a forms-capable browser.
- ◆ To search in the categories [Subject Class](#), [Language](#) and [Country](#) you have to enter the respective abbreviations.
- ◆ Except for Keywords you can use the Booleans **and**, **or** and **not**. Without these Booleans a search with **and** will be performed.
- ◆ You can also use **question marks** to match only the beginnings of words - you cannot put question marks at the front or inside of words.

