Elektronický archív akademických prací na ESF MU

Jiří Poláček, polacek@econ.muni.cz

Abstrakt

Na Ekonomicko-správní fakultě Masarykovy univerzity v Brně se podařilo úspěšně realizovat první fázi zpřístupnění plných textů akademických závěrečných prací v elektronické podobě v rámci internetového prostředí, což je podle dostupných informací počin v České republice zatím ojedinělý. Příspěvek proto shrnuje zkušenosti s nezbytnou administrativou sběru prací od posluchačů, polemiku nad zvoleným prezentačním formátem a jeho možnostmi, a také letmý popis technického řešení webového distribučního systému, skrze nějž mají oprávnění uživatelé k pracím přístup.

Nezbytná administrativa v sobě zahrnuje překážky v zakotvení povinnosti posluchačů odevzdávat práce také v elektronické podobě a vypořádání se s autorskoprávními aspekty odevzdaných prací. V pojednání o prezentačním formátu jsou shrnuty výhody formátu PDF, využití možnosti jeho zabezpečení a digitálního podepisování a také úskalí týkající se převodu prací do tohoto formátu. Popis technického řešení se dotýká zejména tvorby bibliografických záznamů k odevzdaným pracím v jazyce XML a jejich využití při vyhledávání požadované práce prostřednictvím webového prohlížeče. Stejně tak jsou komentovány použité techniky zabezpečení vzniklého elektronického archívu před neoprávněnými přístupy.

Závěrem jsou rozebírány další možné cesty rozvoje a vylepšování stávajícího systému včetně možnosti zprovoznění fulltextového vyhledávání, které úzce souvisí s rozborem, v jaké míře je možné pro takovéto projekty využít volně dostupného softwaru.

Abstract

At Faculty of Economics and Administration, Masaryk University in Brno, there was realized the first phase of the project which deals with accessing academic theses in an electronic form via Internet. To the best of our knowledge, this act is currently unique in the Czech Republic. This paper recapitulates experience with an administration of the collection of students' works, polemic of the selected presentational format and its features, and also a partial description of a web server technical solution, which provides an access to the e-theses to authorized users.

The necessary administration itself includes rule barriers of student's obligation to submit the thesis in an electronic form and also dealing with copyright aspects of theses. Benefits of the Portable Document Format, security possibilities, digital signing, and problems of theses conversion to PDF are discussed in the section dealing with the presentational format. Description of the technical solution coheres with theses bibliographical records creation in XML language and its usage in searching the requested thesis through a web browser. Security principles of the electronic database related to the unauthorized accesses are commented as well.

Finally, there are considered additional paths of the actual system development including the possibility of fulltext searching, which is close to an analysis of the free software usage in similar projects.

Středisku vědeckých informací[1] na Ekonomicko-správní fakultě (dále též SVI ESF) Masarykovy univerzity v Brně se podařilo realizovat veřejně anoncovanou (viz [nekuda]) první fázi sběru, převodu a vystavování elektronických verzí akademických závěrečných prací odevzdaných k obhajobě v roce 2002. Zkušenosti získané a dále v tomto příspěvku rozebírané poměrně přesvědčivě potvrzují obecnou premisu, že základním předpokladem dosáhnutí kýženého výsledku je *vůle celý pracovní proces nastartovat*; nezbytná administrativa a hledání vhodného technického řešení pak již nepředstavují nezdolatelnou překážku.

Administrativa sběru prací v elektronické podobě

Výchozím bodem celého procesu je zakotvení *povinnosti* posluchačů odevzdávat s papírovou verzí závěrečných prací taktéž jejich elektronickou verzi. Toto nařízení samozřejmě může fakultní knihovna maximálně iniciovat; je třeba přesvědčit vedení školy, aby nařízení odsouhlasilo a zařadilo do studijního řádu. Vzhledem k rozmanitosti formátů elektronických dokumentů rozhodně není na škodu již na tomto místě precizovat, co se pod elektronickou formou závěrečné práce rozumí. Na ESF je odpovídající pokyn formulován následovně:

> Práce se odevzdává na zadávající oborovou katedru vždy v písemné a elektronické podobě. V písemné podobě je práce vyvázána v pevné vazbě a odevzdává se ve dvou exemplářích. Exemplář, který obsahuje originál zadání si ponechá katedra pro své potřeby, druhý odevzdá po ukončení SZZ Středisku vědeckých informací.
>
> V elektronické podobě se odevzdává práce na zadávající katedru na podepsané disketě nebo CD v některém z následujících formátů dokumentů (vždy včetně kompletní identifikace, anotace a klíčových slov):
>
> - dokument MS Word ve verzích 95 a 97/2000/XP (obvyklá přípona DOC)
> - Rich Text Format (obvyklá přípona RTF)
> - textový dokument 602 Text z balíčků PC Suite 2000 a PC 2001
> - textový dokument z kancelářského balíku MS Works
> - užití jiných než doporučených formátů (např. TeX) je třeba dohodnout s SVI ESF MU.
>
> V každém případě nechť je práce kompletní v jediném souboru, a to včetně titulní strany, bibliografické identifikace, anotací, prohlášení, obsahu, literatury i příloh. V případě, že některá z povinných částí práce bude chybět, nebude povinnost odevzdání diplomové práce i v elektronické podobě uznána.

Pozastavme se nejdříve u digitálního média, na kterém má být práce odevzdána – disketa nebo cédéčko. Technicky by nebyl problém provést sběr prací modernější cestou, například elektronickou poštou, varianta fyzického média byla zvolena čistě z administrativních důvodů – pro sekretářky kateder je prostě jednodušší vybrat od posluchače dva vyvázané sešity, jednu disketu a vydat mu potvrzení, že vše požadované odevzdal. Z hromady disket odevzdaných v roce 2002 bylo nečitelných pouze pár kusů, inkriminované případy byly uspokojivě dořešeny s konkrétními posluchači prostřednictvím elektronické pošty.

Specifikované formáty dokumentů, ve kterých je možné závěrečnou práci odevzdat, představují pro posluchače docela slušnou možnost výběru z textových editorů. Jelikož odevzdané elektronické verze závěrečných prací jsou před vystavováním zpracovávány a především převáděny do prezentačního formátu (viz odstavec *Formát PDF jako ochranný obal závěrečných prací*), je pro Středisko vědeckých informací podstatná pouze skutečnost, zda má softwarové prostředky pro korektní manipulaci s odevzdanými pracemi – což v případě specifikovaných formátů samozřejmě má, dokonce se jedná o nabídku textových editorů, s kterými mohou posluchači pracovat na fakultních počítačích. Práce odevzdané v roce 2002 byly v drtivé většině dokumenty MS Wordu, jen zhruba deset posluchačů odevzdalo práce vysázené v systému LaTeX[2].

Pro zamezení jiných případných komplikací, které by se mohly týkat zejména grafů, obrázků a speciálních symbolů vkládaných do textu práce, či jiných typů příloh, obsahuje pokyn ještě následující požadavky:

> Kromě souboru s výsledným dokumentem je dále zapotřebí odevzdat:
>
> - veškeré tabulky, obrázky a jiné přílohy, které se v práci vyskytují, a to v jejich původních souborech (tedy například pokud bude práce obsahovat tabulky a grafy vytvořené v Excelu, pak odevzdávat tyto excelové soubory),
> - soubory s nestandardními písmy či symboly, pokud budou v dokumentu použity, aby byl elektronický dokument správně zobrazen i u jiných čtenářů. Za standardní fonty písma se považují písma předinstalovaná ve Windows, tedy zejm. Times New Roman, Arial, Courier New, Wingdings, Tahoma).
>
> Veškeré odevzdané soubory je možné pro úsporu místa zkomprimovat do standardních archivních formátů (ZIP, RAR, ARJ).

Celkově lze konstatovat, že s takto předepsanými pokyny byl sběr závěrečných prací v elektronické podobě dobře realizovatelný a kvantitativně úspěšný. Největším problémem odevzdaných prací byla jejich kvalita z technického pohledu. Mnohé práce byly roztříštěny do pěkné řádky jednotlivých souborů, měly chybně provedené číslování stránek, vzhled nadpisů vypadal v rámci jediného dokumentů naprosto odlišně, nezřídkakdy chyběly některé části práce – přílohy, titulní strany či obsah. V následném zpracování prací tedy byly v rámci možností chybějící části doplňovány a kritické závady opravovány; jakožto poučení do dalších let byl do pokynů přidán požadavek odevzdání práce v jediném souboru se zdůrazněním, že opravdu žádná část práce nesmí chybět. Výhledově tato zkušenost indikuje, že je třeba zvyšovat počítačovou gramotnost posluchačů ESF v oblasti elektronického zpracování textu, a to zejména naučit je základům typografie, pracovat se styly a oddíly.

[1] Středisko vědeckých informací působí zejména jako fakultní knihovna, viz http://www.econ.muni.cz/svi

[2] V takovém případě je třeba kromě výsledného dokumentu v PostScriptu či PDF odevzdat též zdrojový text.

Vystavování plných textů prací versus autorský zákon

Problematika zpracování závěrečných prací v elektronické podobě se ve dvou bodech dotýká autorského zákona. První diskuze se vede ohledně pravomocí školy nakládat se získanými pracemi, které v liteře zmíněného zákona nepochybně figurují jako autorská díla, druhá debata vychází z obav možného zneužití prací, které jsou ve své elektronické podobě snadněji kopírovatelné a šiřitelné.

Při řešení prvního okruhu je vhodné vzít v úvahu tradici, s jakou se nakládá se závěrečnými pracemi v papírové podobě. Jak je patrné z pokynu citovaného ze studijního řádu, jeden výtisk práce se dostává do rukou Střediska vědeckých informací, což v praxi znamená zařazení práce do veřejně přístupného knihovního fondu. V takovém případě, kdy si každý posluchač i zaměstnanec může práci vypůjčit z knihovního regálu, je myšlenkový posun k *virtuálnímu regálu* v rámci lokální počítačové sítě poměrně přirozený. Jinými slovy, škola se posluchačů neptá, zda-li může jejich práce v elektronické podobě zpřístupnit v rámci svého (respektive univerzitního) intranetu – pokládá to na základě předložené paralely za samozřejmost.

Jinou otázkou zůstává, zda-li může škola závěrečné práce zpřístupnit komukoliv na Internetu, aniž by se na to autorů těchto prací někdo ptal. Protože žádná ze známých analýz na tuto otázku nedala jasnou pozitivní odpověď, jsou posluchači na ESF při odevzdávání své závěrečné práce formou podpisového formuláře dotázáni, zda souhlasí se zveřejněním své práce na Internetu, a to buď *hned*, či až po *určité době* – v roce 2002 tato doba představovala tři roky. Výsledek dotazování za rok 2002 dopadl relativně příznivě, 62,9 % posluchačů dalo souhlas ke zveřejnění ihned, 17,9 % posluchačů pak dalo souhlas se zveřejněním po třech letech. I tak bylo rozhodnuto u formulářů pro rok 2003 zkrátit odkladnou dobu na dva roky a u varianty nesouhlasné *požadovat důvod nesouhlasu*. Přitom je na posluchače apelováno, aby svůj souhlas dali, neboť publikování odborných prací je hnací motor vědecké činnosti. O pracích, kterým autoři nedají souhlas s publikováním na Internetu, je tak docela dobře možné se domnívat, že se jedná o díla obsahově nevalné kvality. Výjimku mohou představovat práce posluchačů, kteří ve svých dílech uvádí neveřejné údaje získané například od spolupracující firmy.

Ona snadnost a přímočarost manipulace s elektronickými daty, jednoduchost kopírování a šíření souborů, tedy základní výhody světa výpočetní techniky, nesou u konzervativních přístupů k řešené problematice obavy z možného zneužití získatelných závěrečných prací v elektronické podobě. V zásadě se může jednat o dva typy přestupků:

- opisování částí již existujících prací dalšími generacemi posluchačů, kteří takto stvořené kompilace vydají za vlastní;
- komerční zneužití, zejména získání a prodej těch prací, které nejsou volně dostupné na Internetu.

Dilema mezi konzervativními a liberálními postoji již bylo podrobněji rozepsáno v [nekuda], zde se soustřeďme na odpověď: Jestliže jsme se vydali cestou elektronické distribuce dokumentů, *musíme na tato rizika přistoupit*.

To jistěže neznamená, že dokumenty musí být vydány napospas uživatelům. V každém případě jsou tyto práce stále chráněny autorským zákonem, takže v případě odhalení komerčního zneužití je možné se domáhat práva soudní cestou. Obranou proti opisování bez řádných citací může být rovněž výpočetní technika – pomocí softwaru, který dokáže porovnávat části dokumentů na parciální shodu, lze odhalit případné viníky a vyvodit z toho exemplární potrestání.

Kromě těchto ochranných mechanismů, které popisované excesy řeší *a posteriori*, jsme se v SVI rozhodli též pro *preventivní opatření* – využití interních ochranných mechanismů prezentačního formátu, v kterém jsou závěrečné práce vystavovány. Podrobně jsou tato opatření popsána v následující sekci.

Formát PDF jako ochranný obal závěrečných prací

Výběr vhodného souborového formátu, v kterém by měly být závěrečné práce předkládány uživatelům, jednoznačně upřednostnil otevřený formát PDF (Portable Document Format) stvořený společností Adobe (viz [adobe]). Toto rozhodnutí bylo učiněno zejména z následujících důvodů:

- jedná se o obecně známý, rozšířený a prověřený formát elektronických dokumentů;
- prohlížeče dokumentů ve formátu PDF jsou volně k dostání pro většinu existujících softwarových platforem;
- formát PDF má interní podporu pro digitální podepisování dokumentů;
- na dokument uložený v PDF je možné uplatnit zabezpečovací pravidla jako například zákaz tisku dokumentu či zákaz kopírování vybraných částí dokumentu;
- v porovnání s ostatními formáty vychází PDF jasně nejlépe (viz například [řepišová]);
- SVI získalo licenci komerčního programu Adobe Acrobat 5, s jehož pomocí lze dokumenty do PDF převádět a následně je zabezpečovat a digitálně podepisovat.

Z hlediska obav možného zneužití závěrečných prací se určitě jako velmi zajímavá jeví vlastnost uplatnění zabezpečovacích pravidel. V rámci standardního způsobu zabezpečení je možno omezit následující oprávnění:

- nepovolit tisk
- nepovolit změny dokumentu (těmi je myšleno zejména výmaz či vkládání stránek, retušování textu apod.)
- nepovolit kopírování nebo vyjmutí částí textu či v dokumentu obsažených obrázků, tabulek a grafů
- nepovolit přidávání a změny poznámek a polí formulářů

Nastavení těchto oprávnění je chráněno správcovským heslem, které je šifrováno pomocí 40bitového klíče. V současné době lze heslo považovat za (relativně) bezpečné, pokud je delší jak deset znaků. U závěrečných prací ESF jsou všechna tato omezení využita. To znamená, že práce v elektronické podobě předkládané Střediskem vědeckých

informací není možné pomocí standardních prohlížečů tisknout ani z nich (přes schránku) kopírovat vybrané části textu[3] – z jiného úhlu pohledu lze tuto prevenci také chápat jakožto zdůraznění výhradně studijního účelu předkládaných prací.
To bohužel neznamená, že by se nestandardními prostředky vybrané části textu nedaly zkopírovat, vždy je například možné uložit si na obrazovce zobrazenou část dokumentu jako obrázek[4] a s ním dále manipulovat – třeba i pokusit se jej převést zpět na editovatelný text. Tento poněkud pracnější úkon zhruba odpovídá možnosti zapůjčit si práci v papírové podobě a tu si ofotit či naskenovat; v každém případě však jej lze jednoznačně posoudit jako porušení daných pravidel (viz odstavec *Vstup do archívu prací*).
Středisko vědeckých informací využilo také druhé zmíněné vlastnosti dokumentů PDF – digitálního podepisování. V rámci funkce *Self-Sign Security* programu Acrobat byly vytvořeny podpisové certifikáty, pomocí kterých je každá závěrečná práce digitálně podepsaná. Veřejné části těchto podpisových certifikátů lze získat z webových stránek SVI, jejichž aplikací si čtenář může ověřit, zda podepsané práce skutečně předkládá Ekonomicko-správní fakulta. Zároveň je tímto způsobem předesíláno, že závěrečná práce bez podpisu je s největší pravděpodobností pochybný dokument, za jehož obsah SVI nenese odpovědnost. Ověřování digitálních podpisů v prohlížeči Acrobat Reader je možné až od verze 5.1[5].
Převod odevzdaných závěrečných prací z formátu, v kterém jej posluchači odevzdali, do formátu PDF, se provádí tiskem do souboru na virtuální tiskárnu *Acrobat Distiller*[6]. Vzhledem k tomu, že většina odevzdaných prací je zpracovatelná v programu Microsoft Word, lze pro převod použít makro *PDFMaker*, které umí výsledný dokument obohatit o křížové odkazy a záložky z nadpisů (pokud posluchač ve své práci pro nadpisy použil styly). Pro zachování srovnatelné kvality výsledných dokumentů je nanejvýš vhodné nakonfigurovat odpovídajícím způsobem veškeré parametry převodu, které makro nabízí, SVI si za tímto účelem vypracovalo dvoustránkový prováděcí předpis.
Po převodu práce do formátu PDF již tedy stačí jen jej zabezpečit, podepsat (nutno udělat právě v tomto pořadí) a výsledný uložený dokument je *připraven k distribuci*. Dlužno podotknout, že převod stovek odevzdaných závěrečných prací znamená lehce rutinní, avšak časově náročnou operaci, na kterou je třeba počítat s několika týdny práce.

Vznik a využití bibliografických záznamů o odevzdaných pracích

Souběžně se zpracováním odevzdaných závěrečných prací v elektronické podobě probíhá také katalogizace těchto prací v papírové podobě. Údaje o každé práci jsou zaneseny do elektronického knihovního katalogu a každé práci je přidělena signatura předtím, než se dostane do volného výběru v knihovně. Tímto procesem vznikne ke každé práci bibliografický záznam v elektronické podobě[7], který tvoří základ metadatových informací pro vyhledávání dané práce ve vzniklém elektronickém archívu závěrečných prací.
Odpovídající záznamy jsou z knihovního katalogu vyexportovány a transformovány na záznamy označkované v univerzálním jazyce XML (eXtended Markup Language, viz [xml]), aby mohly býti dále zpracovávány. Jako názvy elementů figurují identifikátory polí převzaté z knihovního katalogu; pro záznamy tedy není využito některého metadatového standardu, jakým je například Dublin Core [dc], v případě potřeby jednotného standardu je však možné požadovaný převod velmi jednoduše učinit.
Po převodu bibliografických údajů do XML je nutno v záznamech doplnit informaci, kterou knihovní katalog neuchovává – souhlas posluchače se zveřejněním práce na Internetu. Tuto informaci je nutno zanést do záznamů ručně na základě podpisových archů, kde posluchači projevili svou vůli ohledně dalšího nakládání školy s prací. Souhlas je zde vyjádřen celým číslem, které odpovídá počtu měsíců, od kdy je po obhajobě práce na Internetu přístupná, nula znamená okamžitý souhlas, pro nesouhlas máme rezervováno číslo `-1`. Na základě takovéto informace již může distribuční systém rozhodovat o tom, které práce jsou či nejsou pro čtenáře přístupné.
Po doplnění (ne)souhlasů posluchačů jsou bibliografické záznamy připraveny k indexaci vyhledávacím systémem, s jehož pomocí je pak možné zprovoznit vyhledávání v metadatových záznamech k pracím prostřednictvím webového prohlížeče. Práce s vytvořeným archívem proto nyní probíhá následovně:

- na úvodní straně archívu je čtenář seznámen s podrobnostmi projektu a poučen o jeho využitelnosti;
- ve webovém formuláři zadá svůj dotaz, na což vyhledávací nástroj odpoví seznamem vyhovujících bibliografických záznamů s odkazy na plné texty;
- čtenář si na stránce s výsledky dohledá práci, která jej zajímá, přičemž pro zobrazení plného textu musí zadat své fakultní uživatelské jméno a heslo.

Za měsíc březen bylo z elektronického archívu požadováno celkem 2 019 prací od 80 uživatelů – převážně posluchačů. Počet čtenářů, kteří si našli cestu do elektronického archívu, tedy zatím není veliký – je to samozřejmě také dáno omezenou nabídkou archívu, který prozatím čítá 256 prací od posluchačů, kteří obhajovali v roce 2002. Vyjma případů s výpadky elektřiny nebyly při provozu archívu pozorovány či nahlášeny žádné významné problémy. Podrobnější informace k procesu práce s archívem závěrečných prací jsou předmětem následujících odstavců.

[3] Celý dokument jakožto počítačový soubor dat samozřejmě kopírovat lze.
[4] Běžná funkce prohlížečů obrázků či dokonce operačních systémů označovaná jako *sejmutí obrazovky*.
[5] V době psaní tohoto článku je Acrobat Reader verze 5.1 k dispozici pouze v anglickém vydání.
[6] Společnost Adobe tento proces trefně nazývá *destilací*.
[7] V SVI aktuálně používáme knihovní systém EOSi T Series.

Vstup do archívu prací

Úvodní webová stránka archívu závěrečných prací je veřejně přístupná[8] a kromě základních informací o projektu a nápovědy k vyhledávání v archívu slouží zejména k právnímu vymezení využitelnosti dostupných prací pouze ke studijním účelům. Konkrétně toto poučení, které čtenář vstupem do archívu potvrzuje, zní:

1. Každá závěrečná práce – ať v papírové či elektronické podobě – je chráněna Zákonem č. 120/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (tzv. autorský zákon) se všemi důsledky ze zákona vyplývajícími. Jakékoli šíření závěrečných prací a zneužívání ke komerčním účelům není přípustné a je trestné!
2. Závěrečné práce užívá vysoká škola nevýdělečně ke své vnitřní potřebě pro školní a studijní potřeby.
3. Přístup do archivu závěrečných prací je monitorován a může být kdykoliv bez udání důvodu zamezen.

Domníváme se, že toto poučení současně s autorským zákonem a studijním řádem jsou dostatečným zázemím pro uplatňování případných právních postihů.
Jako vyhledávací systém jsme zvolili program *Lucene*, volně dostupný indexovací a rychlý vyhledávací nástroj vyvíjený pod hlavičkou The Apache Jakarta Project (viz [jakarta]). Součástí programu je i jeho webová nadstavba a přestože zacházení tohoto programu s českou diakritikou není optimální, lze program na české záznamy docela dobře využít. Bibliografické záznamy jsou indexovány vždy jako celek, program neumožňuje kategorizované indexování podle předmětových polí, aby bylo možné vyhledávat například jenom mezi jmény autorů či v názvu pracích; vzhledem k nízkému počtu malých záznamů a rychlosti programu však toto neznamená žádné zdržení při vyhledávání v archívu. Lucene samo o sobě ovšem také neumí indexovat jiné dokumenty, než prostý text a webové stránky, nehodí se proto k indexování celých textů závěrečných prací. V současné době tedy vyhledávací program splňuje svůj účel, do budoucna však možná bude nutné nahradit jej programem komplexnějším.

Princip fungování distribučního systému pracujícího nad archívem prací

Na základě čtenářova dotazu vyhledaný a vrácený seznam vyhovujících bibliografických záznamů je ještě před zobrazením v prohlížeči „za letu“ zpracován tak, aby čtenář viděl dobře čitelné webové stránky a platné odkazy na plné texty prací. O tuto transformaci metadatových údajů z jazyka XML do jazyka HTML se stará volně dostupný publikační systém *Cocoon* vyvíjený pod hlavičkou The Apache XML Project (viz [apache-xml]). Požadovaná transformace je popsána (naprogramována) v jednom dokumentu[9] a je přizpůsobená potřebám SVI zejména co se týče konstrukce a aktualizace odkazů na výsledné práce v PDF.
V rámci první a tedy víceméně testovací fáze projektu jsou plné texty prací přístupné pouze posluchačům a zaměstnancům Ekonomicko-správní fakulty oproti zadání jejich uživatelského jména a hesla, tj. práce jsou před neoprávněnými přístupy chráněny touto standardní metodou. Proto jsou nyní odkazy na základě metadat v bibliografických záznamech konstruovány tak, aby směřovaly do zabezpečené[10] sekce fakultního webového serveru, který v tomto případě slouží jako úložiště prací. Jednoduchou úpravou v popisu transformací lze v budoucnu práce se souhlasem autorů zpřístupnit veřejnosti.
Dlužno dodat, že základem distribučního systému, který se stará o vyhledávací rozhraní a o propojení bibliografických záznamů s plnými texty prací, je webový server *Tomcat*, který rovněž spadá do otevřeného projektu Apache Jakarta. Výše zmíněné programy Lucene (vyhledávací systém) a Cocoon (publikační systém) jsou aplikacemi serveru Tomcat, tj. tento server[11] pro svůj chod potřebují. Všechny tyto tři programy, kromě toho, že jsou volně dostupné včetně zdrojových kódů, jsou navíc naprogramovány v jazyce Java, díky čemuž nejsou závislé na operačním systému[12].
Tolik k popisu aktuálního stavu rozvíjeného projektu, následující odstavce jsou věnovány polemice plánovaného rozvoje archívu zejména co do komfortu při vyhledávání.

Otevřené problémy pro další fáze projektu

Při výstavbě archívu závěrečných prací bylo jedním z hlavním kriterií jeho zprovoznění bez zbytečného prodlení – což se ke spokojenosti posluchačů ESF povedlo. Víme, že k ideálnímu stavu má současná první fáze projektu ještě hodně daleko, přičemž možná vylepšení jsou předmětem širší diskuse.
Často probíraným tématem je možnost fulltextového vyhledávání v databázi akademických prací. Pomineme-li úvahy, nakolik je takové vyhledávání přínosnější oproti vyhledávání v bibliografických záznamech, zůstává otázka, jak to provést technicky. Nemá smysl indexovat soubory odevzdané posluchači, které mohou být v různých formátech, roztříštěné a těžko by se z indexů nad těmito soubory konstruovaly vazby na výsledné dokumenty. Jako vhodné se tedy jeví indexovat obsah prací přímo v prezentačním formátu PDF. Zde ovšem narážíme na problémy s českou diakritikou – nástroj *Catalog* společnosti Adobe slova s čárkami a háčky neumí indexovat. Nástrojem, který si poradí s jakýmikoliv znaky, je *TextSpy* české společnosti Amos Software (viz [texstspy]), jedná se však o komerční produkt, který nenabízí

[8] Čtenář ji najde na adrese http://www.econ.muni.cz/svi/zp.html
[9] Jedná se o soubor transformačních stylů vyhovujících specifikaci jazyka XSLT, viz [xslt].
[10] Zabezpečené také ve smyslu šifrované komunikace pomocí protokolu SSL.
[11] Tento nebo jiný obdobný, který umí zpracovávat speciální webové stránky známé jako *Java Server Pages*.
[12] V současné době distribuční systém běží na počítači s operačním systémem Windows, do budoucna zvažujeme přechod na Linux.

žádnou vhodnou formu licencování pro účely on-line dostupného elektronického archívu závěrečných prací. Navíc ani Catalog ani TextSpy nedisponují webovou nadstavbou pro vyhledávání v prostředí webového prohlížeče. Hledání vhodného indexačního nástroje souborů PDF je proto součástí příprav k druhé fázi projektu. Rádi bychom, aby se obdobně jako u výše popsaných částí distribučního systému jednalo o volně dostupný program, neboť současná fáze projektu dokazuje, že fungující archív závěrečných prací s pomocí bezplatného softwaru vybudovat lze – jedinou výjimku v tomto případě představuje licence programu Acrobat.
Popsanému řešení může být vytýkáno, že je vystavěno bez jakékoliv domluvy společného standardu s potenciálními řešiteli stejného problému v jiných institucích. Zkušenosti z reálného provozu však mohou přispět k vhodné formulaci takovýchto standardů, přičemž některé komponenty projektu na ESF je možno do standardizované podoby převést automatizovaně (týká se bibliografických záznamů v jazyce XML) či chybějí komponenty doplnit (například vytvoření univerzálního archivačního formátu závěrečných prací).

* * *

Související odkazy a dokumenty

[nekuda] Jaroslav Nekuda: Bakalářské, diplomové a disertační práce v elektronické podobě na Ekonomicko-správní fakultě Masarykovy univerzity v Brně. Ikaros [online]. 2002, č. 12 [cit. 2002-12-01], ISSN 1212-5075. <http://www.ikaros.cz/Clanek.asp?ID=200212007>

[řepišová] Zuzana Řepišová, Hana Vochozková, Petr Sojka, Jana Kovářová, Jaroslav Nekuda: Disertace Masarykovy univerzity on-line. Grantová zpráva. Ústav výpočetní techniky MU Brno, 2002.

[adobe] Informace o formátu PDF a softwarového produktu Acrobat na webových stránkách společnosti Adobe: http://www.adobe.com/

[texstspy] Informace o indexovacím nástroji TextSpy včetně cenové politiky: http://www.textspy.cz/

[xml] Tim Bray, Jean Paoli, C.M. Sperberg-McQueen, Eve Maler: Extensible Markup Language (XML) 1.0 (Second Edition). W3C 2000. <http://www.w3.org/TR/REC-xml>

[xslt] James Clark: XSL Transformations. W3C 1999. <http://www.w3.org/TR/xslt>

[dc] Dublin Core Qualifiers. DCMI 2000. <http://dublincore.org/documents/dcmes-qualifiers/>

[jakarta] Informace o projektu The Apache Jakarta Project: http://jakarta.apache.org/

[apache-xml] Informace o projektu The Apache XML Project: http://xml.apache.org/