

Masarykova Universita

Fakulta Informatiky



Vyhledávací služby pro OAI

[\(http://www.openarchives.org/\)](http://www.openarchives.org/)

Esej předmětu PV070

Vypracoval: Tomáš Drn

UČO: 387044

Datum: 3. 12. 2010

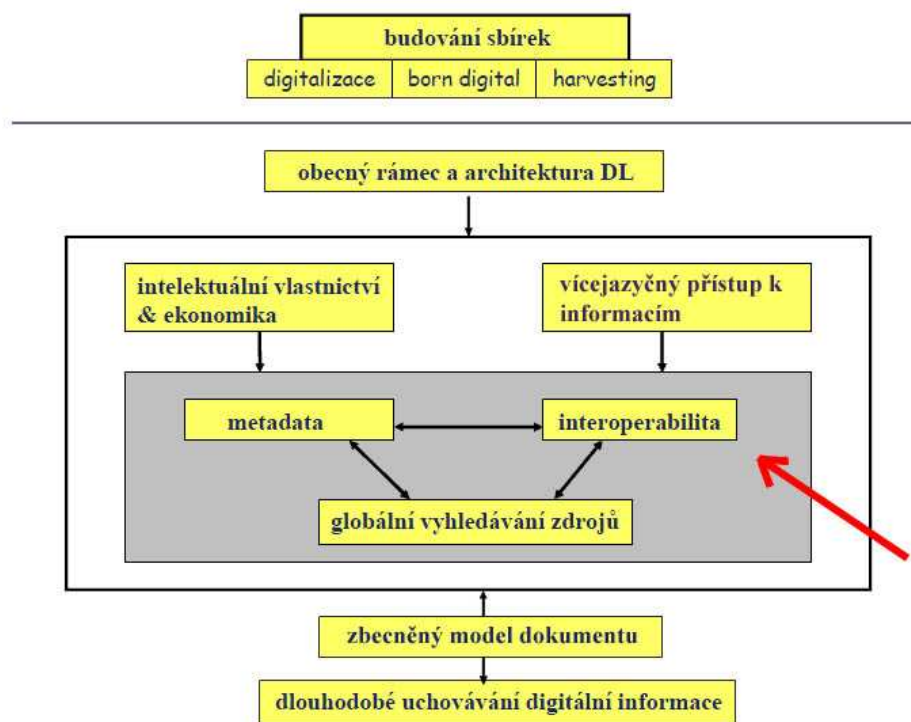
1. Úvod

Tato esej má za úkol seznámit čtenáře s tématem: „Vyhledávací služby pro OAI“. Před řešením samotné problematiky bude stručně zmíněno o co v záležitosti OAI – Open Archives Initiative (Iniciativa otevřených archivů) a úzce spjatým vyhledávacím protokolem OAI-PMH (Protocol for Metadata Harvesting – protokol pro sběr metadat) jde. Cílem práce je provést rozbor, přehled situace v oblasti vyhledávacích služeb OAI a seznámit s případnými novinkami, které se v problematice objevili.

2. Účast v Digitálních knihovnách

Interoperabilita

Obecně je to schopnost systémů, služeb a organizací pracovat společně nad problémy nebo úkoly vedoucí ke společným nebo rozdílným cílům. V technické oblasti je interoperabilita podporována otevřenými standardy vzájemné komunikace mezi systémy a standardy pro popis zdrojových objektů. [1]



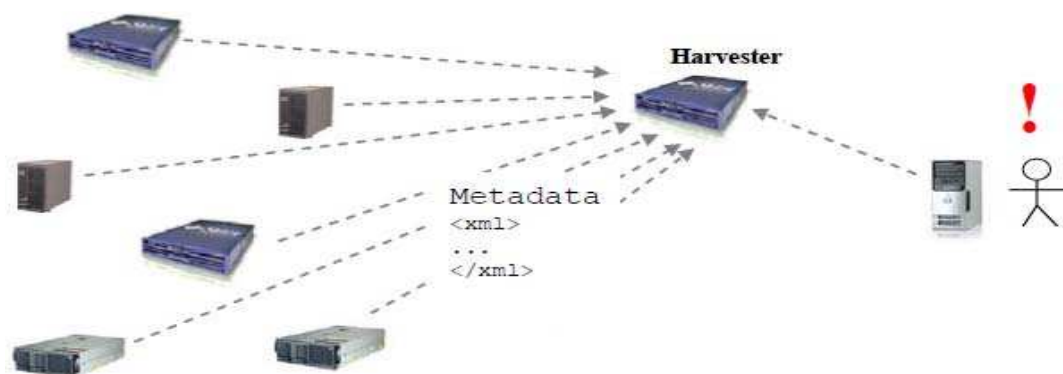
Obr. 1 – Interoperabilita v DL [2]

Iniciativa otevřených archivů (OAI) vznikla pro rozvoj a podporu standardů interoperability, tzn. usnadnění výměny informací digitální povahy. Výraz otevřený se zde vztahuje k principu volného získávání metadat z různých informačních zdrojů a archivem je v tomto případě myšleno skladiště vědeckých prací, nebo jakýkoliv „repozitář“ pro ukládání

informací na webu. Toto vedlo k myšlence, že archivy elektronických tisků jsou perspektivní alternativou k způsobům vědecké komunikace a že je proto nutné jejich vzájemné propojování (iniciativy OAI). Dalším krokem pak byl vznik protokolu pro sběr metadat, známý pod zkratkou OAI-PMH. [5]

OAI – PMH

Je to jednoduchý protokol založený na HTTP a XML, definuje mechanismus pro sklízení metadatových záznamů z repozitářů. Metadata mohou být v jakémkoliv formátu, na kterém se určitá skupina poskytovatelů dat a služeb domluví. Tento protokol neposkytuje možnost prohledávání nad těmito daty, ale umožňuje sběr metadat do jednoho místa. V případě poskytování služeb, sklízení metadat musí být doplněno dalšími prostředky. [4]



Obr. 2 – Princip harvestingu metadat [6]

3. Vyhledávače

Souborný katalog WorldCat

WorldCat obsahuje dokumenty z katalogů knihoven celého světa a je celosvětově neobsáhlejším a nejúplnějším katalogem obsahující 110 milionů bibliografických záznamů. U každého záznamu je možné snadno zjistit, ve které nejbližší knihovně se dokument nachází. Záznamy jsou katalogizovány v mnoha jazycích. WorldCat obsahuje také záznamy více než 40 katalogů národních knihoven. [11]

3.1 OAIster

Link: <http://oaister.worldcat.org/>

Byl vyvinut na Universitě v Michiganu uskupením **Digital Library Production**, tento nástroj zpřístupňuje vědecky orientované datové zdroje, které byly dříve běžnému uživateli skryté nebo těžce dostupné. Takovéto zdroje nebyly procházeny roboty vyhledávacích

společností, které by je následně zobrazily na Internetu. Projektu je založen na využití protokolu OAI, díky němuž jsou tyto zdroje zpřístupněny. Tzn. uživatel nezíská pouze jejich metadata, ale celé zdroje. [8]

OAIster nadále roste a expanduje. Aby byla tato expanze podpořena, byla databáze OAIster přiřazena do katalogu WorldCat v červenci 2010. Výhody spojení: [10]

- Samoobslužné harvestování, což znamená, že si můžete nastavit a naplánovat vlastní sklizně a přizpůsobit sbírku metadat.
- Vaše držené knihovny mohou být nastaveny ve WorldCat,
- Vaše unikátní kolekce získají maximální viditelnost přes WorldCat a WorldCat Gateway, která je k dispozici zdarma.
- Vaše metadata budou sdružené všude tam, kde je publikován WorldCat.

Vyhledávání v OAIster

OAIster má jednoduché vyhledávání, kde jednoduše napíšeme název hledené publikace a zbytek ponecháme na něm a pokročilé vyhledávání, kde můžeme blíže specifikovat hledanou informaci a posílit úspěšnost vyhledávání. Takto je to i u všech dalších zmiňovaných vyhledávačů, proto bude vždy rozebírána pouze část pokročilého vyhledávání, v té se vybrané vyhledávače liší.

Údaje pro vyhledávání: Pro vyhledávání zde máme 2 sekce.

1. Zadájí se 3 vyhledávací termíny z výběru *“titulek, klíčová slova, autor, typ obsahu, vydavatel, jazyk, formát (pdf, HTML, a.j.) apod.”*.
2. Zadáávají se omezení výsledku *Roky od – do, Jazyk a Typ zdroje (“video, audio, obraz, atd.”, celkem 5)*, tuto druhou sekci není nutné vyplnit.

Výsledek vyhledávání: Jsou vám zobrazeny všechny výsledky, na kterých se vyskytují zadané informace pro vyhledávání. S výsledky je možné pracovat, dále je třídit, prohlížet si náhledy atd. a tak rozpoznat hledaný zdroj. Pokud se zaměříme na hledaný zdroj, jsou nám sděleny všechny dostupné informace včetně odkazu na celou práci v elektronické podobě, pokud ji zdroj obsahuje.

3.2 BASE

Link: <http://www.base-search.net/>

Bielefeld Academic Search je jedním ze světově nejobsáhlejší vyhledávačů především pro akademické zdroje otevřeného přístupu. BASE je provozován Bielefeldskou univerzitní knihovnou. Je charakterizován následujícími funkcemi: [9]

- Intelektuální vybraní zdrojů.
- Pouze materiály, které splňují specifické požadavky akademické kvality.
- Transparentnost při vyhledávání.
- Zveřejnění internetových zdrojů "Deep Web", které jsou ignorovány komerčními vyhledávači.
- Zobrazení výsledků vyhledávání zahrnuje přesné bibliografické údaje.
- Několik možností pro třídění výsledků.

Vyhledávání v BASE

Údaje pro vyhledávání: V BASE je možno výběr specifikovat v 5 sekcích.

1. Základní informace – Titulek, Autor, Název, Vydavatel, URL (i částečné)
2. Zda chceme případně námi zadaná klíčová slova použít i v jiných tvarech.
3. Obsah zdroje – jazyk, oblast
4. Roky vydání *PŘED* a *PO*
5. Typ dokumentu (“*audio*, *video*, mapa, kniha, časopis, teze, esej, a.j., celkem 12”)

Výsledek vyhledávání: Podobný OAIster, ale nedá se dále třídít, méně přehledný způsob zobrazení výsledku.

3.3 OpenDOAR

Link: <http://www.opendoar.org/>

The Directory of Open Access Repositories jsou autoritativní adresáře s uložištiemi s otevřeným přístupem využívající OAI-PMH protokol. Jde o adresář typu akademických repozitářů. V tomto projektu se nespolehají na automatizované prohlížení a škatulkování záznamů, ale všechny materiály prohlíží lidský faktor (sklizeň metadat). Projekt umožňuje vyhledávání samotných repozitářů, nebo jednotlivé části obsahu repozitářů.[12]

Vyhledávání v OpenDOAR

Údaje pro vyhledávání: Umožňuje prohledávání repozitářů podle zemí, vyhledávání v repozitářích podle klíčových slov, jazyka, použitého softwaru.

Výsledek vyhledávání: Výsledky lze dále třídit, náhledy obsahují 8 informací, podle kterých se dají výsledky snáze rozpoznávat.

3.4 DOAJ

Link: <http://www.doaj.org/>

The Directory of Open Access Journals je služba zabývající se otevřeným přístupem nad fulltextovými, vědeckými a odbornými časopisy. Uživatelé mají právo číst, stahovat, kopírovat, tisknout, prohledávat, nebo na odkaz s plnými texty těchto článků. [13]

Vyhledávání v DOAJ

Údaje pro vyhledávání: Zde je první velká odlišnost od ostatních vyhledávačů.

- **Vyhledávání časopisů** pouze zadáním názvu, klíčových slov, nebo abecedy.
- **Vyhledávání časopisu** ručním nastavováním odvětví a podkategoriemi např. *ARCHITEKTURA* -> *HISTORIE VYSOKÝCH STAVEB* apod. dokud se nezobrazí časopisy pod námi zvoleným odvětvím a kategorií.
- **Vyhledávání článku**, zde se nachází dvě textová pole, která se dají spojovat operátory **AND**, **OR**, **NOT** do kterých můžeme zadat 2 z následujících 6 údajů (autor, titulek, "Journal title", ISSN, klíčová slova, abstrakt)

Výsledek vyhledávání: Ve výsledku se nám zobrazí všechny časopisy dostupné pro dotaz, či odvětví s podkategoriemi s podrobnými informacemi, jako ISSN, jazyk ve kterém se vydávají, vydavatel, země vydání a.j.

3.5 Scirus

Link: <http://www.scirus.com/>

Je komplexní vědecko-specifický internetový vyhledávač. Běžné internetové vyhledávače vyhledávají podle klíčových slov. Vyhledají všechny stránky, na kterých se daná klíčová slova vyskytují. Scirus vyhledává obsah pouze na stránkách, které jsou určeny vědě a vědním odvětvím. [14]

Vyhledávání v Scirus

Údaje pro vyhledávání: Scirus má nejrozsáhlejší možnosti vyhledávání ze všech zmíněných vyhledávačů, lze opravdu podrobně specifikovat hledaný zdroj. Nejprve zadat do dvou polí výběr ze sedmi z následujících údajů (autor, autorova příslušnost, "Article

title”, “Journal title“, ISSN, klíčová slova, URL – i část). Tyto pole se dají libovolně propojovat operátory **AND, OR, NOT**. Tyto údaje lze rozšířit o následující podúdaje:

- Typ informací (abstrakt, teze, dizertační práce, kniha, časopis, a.j. celkem 12)
- Formát dokumentu (pdf, ppt, ps, HTML, a.j. celkem 12)
- Typ zdroje
 - Journal – 15 podkategorií
 - Web source – 20 podkategorií
- Oblasti (Architektura, Informatika, Matematika, a.j., celkem 20)

Výsledek vyhledávání: Výsledky lze dále třídit, náhledy obsahují méně informací, málo přehledné. Pro zjištění, zda byl nalezen požadovaný zdroj je nutné nalezené výsledky otevřít a důkladněji prozkoumat.

3.6 Další vyhledávače a služby využívající OAI-PMH

- Fedora, <http://www.fedora-commons.org/>
- GESIS, <http://www.gesis.org/>
- Solr, <http://lucene.apache.org/solr/>
- PUMA, <http://puma.isti.cnr.it/>
- ROAR, <http://roar.eprints.org/>
- DRIVER, <http://www.driver-repository.eu/>
- VuFind, <http://vufind.org/>
- AquaBrowser, <http://www.serialssolutions.com/aquabrowser/>
- Encore, www.endeca.com/
- Enterprise, <http://www.librarytechnology.org/>

4. Závěr

Už v dřívější době se prokázalo, že myšlenka Open Access a s ní spojená metodika Open Archives Initiative je velice dobrá. Význam těchto a podobných projektů spočívá v systematickém zpřístupňování velkého množství informací na Internetu. Uživatel má velmi jednoduchý přístup k informacím z mnoha různých tematických oblastí a takřka odkudkoliv. To samé se dá říci o autorovi nově vzniklých dat, který má k zprostředkování své práce stejně jednoduchý přístup k poskytovatelům těchto služeb, jako již zmíněný uživatel. Co se týče

jednotlivých projektů, setkáváme se s graficky příjemným prostředím, které se dá jednoduše ovládat a obsahuje dostatečnou nápovědu s podporou více jazyků.

Při vypracovávání této eseje jsem zjistil, že existuje opravdu mnoho služeb a projektů používající protokol OAI-PMH a metodika jako taková se mohutně rozrůstá. Bohužel málo kdy jsem se setkal s dobrým materiálem v českém jazyce, což by mohlo znamenat, že u nás nejsou tyto myšlenky tolik rozvíjeny, jako v jiných vyspělých zemích.

5. Zdroje:

- [1] Wikipedie otevřená encyklopedie. Interoperabilita [on-line]. [cit. 2010-11-30]. Dostupný z WWW: < <http://cs.wikipedia.org/wiki/Interoperabilita> >.
- [2] BARTOŠEK, Miroslav. Digitální knihovny - dlfi10-1.pdf, Strom Digitální knihovny. [on-line]. [cit. 2010-11-30]. Dostupný z WWW: < <http://www.ics.muni.cz/mba/dl-fi2010/dlfi10-1.pdf> >.
- [3] BARTOŠEK, Miroslav. Digitální knihovny - dlfi10-7.pdf, Interoperabilita. [on-line]. [cit. 2010-11-30]. Dostupný z WWW: < <http://www.ics.muni.cz/mba/dl-fi2010/dlfi10-7.pdf> >.
- [4] HANOUSEK, Tomáš. OAI-PMH pro začátečníky. [on-line]. [cit. 2010-11-30]. Dostupný z WWW: < www.nacr.cz/Z-files/moznosti_06.pdf >.
- [5] PLANKOVÁ, Jindra. Technická řešení pro otevřený přístup. [on-line]. [cit. 2010-11-30]. Dostupný z WWW: < <http://www.ikaros.cz/node/4555> >.
- [6] MICHÁLEK, René. Harvesting metadat jako nový integrační koncept v portálech nové generace. [on-line]. [cit. 2010-12-7]. Dostupný z WWW: < www.qbizm.cz/miranda2/export/sites/Systemova_integrace2007.pdf >.
- [7] BUREŠOVÁ Iva, Martina. Trendy a vývoj otevřeného přístupu. [on-line]. [cit. 2010-12-7]. Dostupný z WWW: < <http://knihovna.nkp.cz/knihovnaplus81/bures.htm> >.
- [8] NĚMA Jan. OAIster – virtuální katalog zdrojů. [on-line]. [cit. 2010-12-7]. Dostupný z WWW: < http://dspace.muni.cz/handle/ics_muni_cz/76 >.
- [9] BASE. [on-line]. [cit. 2010-12-7]. Dostupný z WWW: < <http://base.ub.uni-bielefeld.de/en/index.php> >.
- [10] OASter. [on-line]. [cit. 2010-12-7]. Dostupný z WWW: < <http://www.oclc.org/oaister/> >.
- [11] Národní knihovna ČR a Ústav výpočetní techniky Univerzity Karlovy v Praze. WorldCat. [on-line]. [cit. 2010-12-7]. Dostupný z WWW: < <http://info.jib.cz/news/souborny-katalog-worldcat> >.
- [12] OpenDOAR. [on-line]. [cit. 2010-12-30]. Dostupný z WWW: < <http://www.opendoar.org/> >.
- [13] DOAJ. [on-line]. [cit. 2010-12-30]. Dostupný z WWW: < <http://www.doaj.org/> >.
- [14] Scirus. [on-line]. [cit. 2010-12-30]. Dostupný z WWW: < <http://www.scirus.com/> >.

6. Metadata v DC

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Vyhledávací služby pro OAI" />
<meta name="DC.Creator.personalName" content="Tomáš Drn" />
<meta name="DC.Subject" content="OAI" />
<meta name="DC.Subject" content="OAI-PMH" />
<meta name="DC.Subject" content="harvesting" />
<meta name="DC.Description.abstract" content="Tato esej má za úkol seznámit čtenáře s
tématem: „Vyhledávací služby pro OAI“. />
<meta name="DC.Date" content="3.12.2010" />
<meta name="DC.Type" scheme="DCMIType" content="Text" />
<meta name="DC.Type" scheme="PHNK" content="Esej" />
<meta name="DC.Format" scheme="IMT" content="application/pdf" />
<meta name="DC.Format.extent" content="10 stran" />
<meta name="DC.Identifier" scheme="URN" content="URN:NBN:cz-nk20102638" />
<meta name="DC.Source" scheme="URL" content="http://www.openarchives.org" />
<meta name="DC.Language" scheme="RFC3066" content="cze" />
<meta name="DC.Relation.isVersionOf" scheme="URL"
content="http://www.openarchives.org" />
```