

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



Knihovny.cz

1. ROČNÍK NAVAZUJÍCÍHO MAGISTRSKÉHO STUDIA

Agáta Dařbujanová

Brno, 26. 11. 2016

Poděkování

Ráda bych poděkovala Ing. Petrovi Žabičkovi za jeho ochotu poskytnout informace o projektu Knihovny.cz.

1 Co jsou Knihovny.cz

Knihovny.cz¹ jsou internetovým portálem, který je vyvíjen Moravskou zemskou knihovnou za podpory Ministerstva kultury. Jeho hlavním cílem je umožnit uživatelům odkudkoli a kdykoli nahlédnout do knihovních katalogů zapojených knihoven v České republice, a to pomocí jednoho rozhraní. Zatím se jedná o zakládající knihovny, tj. národní, krajské a dvě okresní. Tento portál nabízí také kromě obvyčejného vyhledávání titulů i správu čtenářských kont, bez nutnosti přesměrování na web cílové knihovny. Návštěvníkům je zde také zobrazeno několik náhodně vygenerovaných knih pro inspiraci nebo si zde mohou stáhnout vybrané e-knihy zdarma.

O portálu se začalo uvažovat již v roce 2010 a následně byly tyto úvahy vztahy v potaz při plánování Koncepce rozvoje knihoven v České republice na léta 2011–2015. V roce 2012 již vznikl Centrální portál knihoven, jehož úkolem bylo poskytnout „...jednotné rozhraní systému knihoven s cílem zprostředkování všech nabízených služeb z každé knihovny“[1]. Na projekt dohlížela Rada Centrálního projektu knihoven, která hodnotila stav, ve kterém se portál nachází a určovala další oblasti vývoje. Členové této rady tvořili zástupci jednotlivých zapojených knihoven. Původním záměrem bylo navržení portálu a přenechání technické realizaci externí firmě. Tento úkol byl ale nakonec svěřen Moravské zemské knihovně v Brně. O organizaci portálu se nyní stará Sdružení knihoven ČR, které je jedním z jejích zakladatelů. Ke koordinátorům patří i Martin Lhoták z Knihovny Akademie věd ČR. [2]

1. URL: <https://www.knihovny.cz/>

2 Aktuální stav

Přestože portál Knihovny.cz, který je možné najít také pod pracovním označením Centrální portál knihoven, byl slavnostně spuštěn 26. října 2016, není ještě zcela vyladěný. Zpoždění projektu bylo způsobeno nepřipraveností knihoven, resp. jejich systémů, poskytnout portálu metadata jednotlivých exemplářů a zajištění kompatibility přes jejich API, které musely dané knihovny vyvinout. Dalším ovlivňujícím faktorem byl problém s financováním v prvním pololetí roku 2016.

Při vyhledávání je možné dotaz různě omezovat, např. na vyhledání konkrétních knih které jsou uživateli fyzicky nejbližší. Jsou sbírány také plné texty s OCR (optické rozpoznávání znaků) z digitálních knihoven, které sice nejsou uživateli zpřístupněny, je však možné v nich vyhledávat a zobrazit úryvek relevantní dotazu.

Mezi další hlavní funkce API patří možnost půjčování, rezervace, prodlužování výpůjček či objednávání ze skladu. To lze provést prostřednictvím jednoho účtu, se kterým je možné propojit ostatní účty, které má uživatel u zapojených knihoven.

Uživatelské rozhraní je navrženo na základě UX analýz. Byl zde odhalen mj. problém s propojením účtů, kdy služby, které jsou provozovány jinde, mohou také potřebovat informaci o tomto propojení. Je tomu tak kvůli absenci vhodného API. Tvůrci portálu nyní zvažují řešení Perun¹ od CESNETu.

Informace z knihoven jsou v portálu aktualizovány nejpozději do jednoho dne. Tato aktualizace většinou trvá jen pár hodin.

V současnosti se také pracuje na funkci doporučení zdroje, který je podobný aktuálně prohlíženému, protože stávající řešení není dostačující. Tímto se mj. zabývá Vysoké učení technické v Brně, a to za pomoci strojového učení.

1. Perun je systém, který umožňuje uživatelům přístup k různým službám prostřednictvím jednoho uživatelského účtu. Více informací na <http://perun.cesnet.cz>.

3 Jak to funguje

Jádrem celého systému je nástroj VuFind¹, což je open source vyvíjený pro finskou národní knihovnu. Je to modulární systém, takže je možné používat pouze vybrané části a ty dle potřeby modifikovat [3]. Tento nástroj je již využit pro jiné knihovní portály v České republice, které by mohly být považovány za předchůdce Knihoven.cz [4].

VuFind je napsán pro jiné než české prostředí, takže bylo potřeba modifikovat jej tak, aby byl schopný pracovat s českou znakovou sadou a knihovním systémem. Je napsán v programovacím jazyce PHP. Pracovníky MZK byl upraven a přepsán do Javy [5].

V současné době je problém s absencí API pro VuFind. Tato situace je již řešena ve Finsku. O mobilní aplikaci MZK zatím neuvažuje, neboť prioritou je nyní odladění portálu a zkrácení reakční doby.

Metadata či plné texty z různých knihoven se sklízí pomocí open source nástroje, který vyvíjejí zaměstnanci Moravské zemské knihovny (dále MZK). Jedná se o tzv. správce zdrojů (Record Manager)², který je napsán v Javě. Tento správce zajišťuje stahování a aktualizaci záznamů, jejich normalizaci a deduplikaci, zpřesňování a sjednocování [6]. V současné době bylo dokončeno jednoduché webové rozhraní, které je určeno pro interní použití.

Výpočetně nejnáročnější operací je deduplikace, tj. identifikace (pravděpodobně) totožných záznamů a jejich spojení. To, jestli se jedná o stejné záznamy, je určeno na základě několika kritérií, které se pro různé druhy zdrojů mohou lišit (např. pro periodika). Mezi tato kritéria většinou patří stejné ISBN³, ČNB⁴, délka všech společných názvů alespoň 16 znaků a rozdíl stran menší než 5 % či 10 stran.

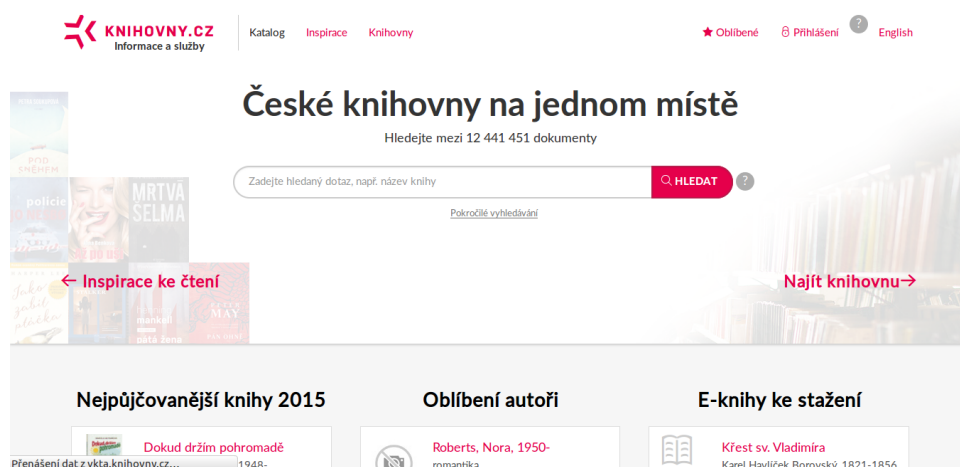
Neočekává se přesná shoda těchto dat, a to z důvodu chyb, které

1. <http://vufind-org.github.io/vufind/>

2. Tento nástroj byl použit mj. také pro projekt Kramerius.

3. Mezinárodní standardní číslo knihy

4. Česká národní bibliografie



Obrázek 3.1: Knihovny.cz: úvodní strana portálu

vznikají při přepisu údajů lidmi do počítače či v nakladatelství. Navíc staré knihy které nemají ISBN dostávají při digitalizaci pouze číslo ČNB a ne všechny knihovny mají tento údaj v metadatech svých exemplářů. Pro určení stejného dokumentu, který byl ale zaspán mírně odlišně, se používá Levenshteinova vzdálenost, která udává počet povolených změn znaků, které vedou ke stejnému řetězci. To je ale nepoužitelné pro krátké názvy.

Databáze obsahuje záznamy zatím nezdeduplikované a sloučené. Sloučené záznamy jsou tvořeny takovými, které byly identifikovány jako stejné.

4 Zhodnocení

Při zadání klíčového slova do vyhledávacího pole je nabídnuta nápověda, kde se daný výraz vyskytuje. Po odeslání složitějšího vyhledávacího dotazu může být nutné nějakou dobu čekat na odpověď, jejíž zobrazení trvá několik sekund, což se uživatelům může zdát velmi nepříjemné.

Jak uvedl i jeden z tvůrců portálu, Petr Žabička, po zadání klíčového slova či fráze nejsou jako první nabídnuty zdroje, které by uživatel mohl očekávat. Pro relevantnější odkazy může být nutné prohlédnout několik stran s výsledky. Pokud je klíčové slovo podobné jménu autora, mohou být uživateli nabídnuty autorovy publikace namísto zdrojů o daném tématu. Výše zmíněné nedostatky jsou aktuálně řešeny, např. nyní se testuje vyladění relevance ve vyhledávání.

Jako výchozí je nastaveno vyhledávání ve všech zapojených knihovnách, které je možné upravit pomocí zaškrťovacího pole. Výběr je třeba potvrdit tlačítkem *Použít filtr*. Z pohledu uživatele se nejedná o nejlepší řešení. Pokud by se ale dotaz spustil okamžitě po změně hodnoty zaškrťovacího pole, velmi by vzrostla doba odpovědi jako je tomu např. při výběru oboru, do kterého hledaná publikace patří.

Protože se jedná o právě spuštěný portál, je třeba počítat s tím, že zde budou nějaké problémy, které je nutné odstranit. Ovšem přínos projektu jako takového může být velký. Umožnit uživatelům provádnout vše skrze jediné rozhraní je pro ně tím nejlepším řešením. Není nutné učit se orientovat v novém prostředí nebo si pamatovat různé přihlašovací údaje, stačí naučit se ovládat pouze jeden portál. Po odladění může být vyhledávání na Knihovnách.cz mnohem rychlejší a jednodušší pro každého uživatele, který nenavštěvuje pouze jednu knihovnu.

5 Metadata v Dublin Core

```
<rdf:Description rdf:about="http://dublincore.org/documents/2001/09/20dcmesxml/»  
<dc:title>Knihovny.cz</dc:title>  
<dc:description>Esej o portálu Knihovny.cz, kde je krátce předsta-  
ven tento projekt a jeho aktuální stav. Je zde také zmíněno jádro  
celého systému, VuFind, popsán Record Manager a princip dedupli-  
kace.</dc:description>  
<dc:date>20161126</dc:date>  
<dc:creator>Agáta Dařbujanová</dc:creator>  
<dc:format>text/pdf</dc:format>  
<dc:language>cs</dc:language>  
<dc:publisher>Masarykova univerzita</dc:publisher>  
<dcq:replacesrdf:resource="https://github.com/moravianlibrary/RecordManager2"/>  
<dcq:replacesrdf:resource="http://vufind-org.github.io/vufind"/>  
<dcq:replacesrdf:resource="https://github.com/moravianlibrary/VuFind-  
2.x"/>  
<dcq:replacesrdf:resource="http://full.nkp.cz/nkkr/knihovna131/13179.htm"/>  
</rdf:Description>
```


Bibliografie

1. LHOTÁK, Martin. KNIHOVNY.CZ – Centrální portál českých knihoven. *Informace* [online]. 2012, č. 3 [cit. 2016-12-01]. ISSN 1805-2800. Dostupné z: <https://www.lib.cas.cz/casopis-informace/knihovny-cz-centralni-portal-ceskych-knihoven/>.
2. *Historie portálu Knihovny.cz*. Dostupné také z: <https://www.knihovny.cz/Portal/Page/historie-portalu-knihovnycz>.
3. *VuFind* [online] [cit. 2016-11-25]. Dostupné z: <http://vufind-org.github.io/vufind/>.
4. ROSECKÝ Václav a Žabička, Petr. Požadavky na výměnu dat mezi knihovním a discovery systémem na příkladu implementace systému VuFind v Moravské zemské knihovně. *Knihovna* [online]. 2003, roč. 24, č. 1, s. 79–88 [cit. 2016-11-25]. ISSN 1801-3252. Dostupné z: <http://full.nkp.cz/nkkr/knihovna131/13179.htm>.
5. *Moravian Library: VuFind* [online] [cit. 2016-11-25]. Dostupné z: <https://github.com/moravianlibrary/VuFind-2.x>.
6. *Moravian Library: Record Manager 2* [online] [cit. 2016-11-25]. Dostupné z: <https://github.com/moravianlibrary/RecordManager2>.