

Od obálek knih k obohacování záznamů

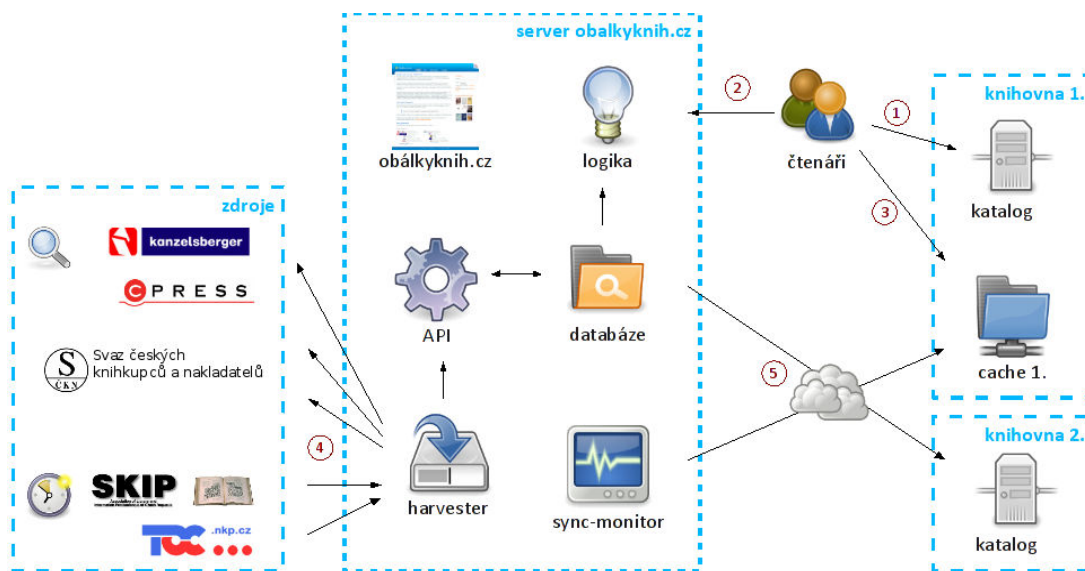
Mgr. Martin Šárfy, Masarykova univerzita v Brně, <sarfy@ics.muni.cz>

Služba obalkyknih.cz vznikla v roce 2008 díky projektu Moravské zemské knihovny v rámci programu VISK3. Jeho původním cílem bylo stát se centrálním repozitářem obrázků obálek knih určených pro integraci do katalogů knihoven. To, jak si ukážeme, se podařilo splnit. V uplynulém roce 2009 navíc projekt prošel dalším intenzivním vývojem, přičemž služba se nově zaměřuje na rozšíření databáze z obálek knih také na další obohacující prvky. O nových službách projektu pojednává tento článek.

1. Databáze obálek knih

Zaměření projektu obalkyknih.cz je vytvoření databáze obálek knih, která by byla knihovnám přístupná jako webová služba umožňující snadnou integraci obálek do knihovního katalogu. Software samotný je přitom vyvíjen jako Open Source Software hostovaný na sourceforge.net [<http://sourceforge.net/projects/obalkyknih>].

Zdrojem pro obálky knih jsou zejména spolupracující nakladatelé a knihkupci, kteří pro knihovny patřící do sdružení SKIP za jistých podmínek souhlasí se zobrazováním obálek jejich knih v knihovních katalozích. Dalším zdrojem obálek jsou knihovny, které aktivně skenují obálky knih z jejich knihovních fondů a tyto vkládají na obálkový server. Obálkový server tak zajišťuje potřebnou infrastrukturu pro evidenci obálek, jejich uchovávání a jednotné zpřístupňování.



Obr. 1.: Celkové schéma architektury systému obalkyknih.cz

Systém je navržen tak, že do stránky webového katalogu stačí vložit krátký JavaScript kód, který pro danou knihu stáhne z webu obalkyknih.cz obohacující informace a ty na stránce zvoleným způsobem vykreslí. Tím se čtenářův prohlížeč (1.) při prohlížení katalogu knihovny pokusí stáhnout informace ze serveru obalkyknih.cz (2.). Aplikační logika dohledá

požadovaný obrázek, skenovaný obsah a další prvky v databázi. Současně si identifikátor knihy zapíše pro pozdější prohledávání zdrojů obsahu. Podle adresy stránky, ze které čtenář přistupuje, určí knihovnu, ze které čtenář přichází a přesměruje prohlížeč čtenáře na nejbližší *cache* server, obsahující kopii souboru (3.).

Současně s poskytováním prvků ve webových katalozích probíhá stahování nových prvků z různých zdrojů (4.). Podle postupu při získávání obálek můžeme tyto zdroje rozdělit do následujících kategorií:

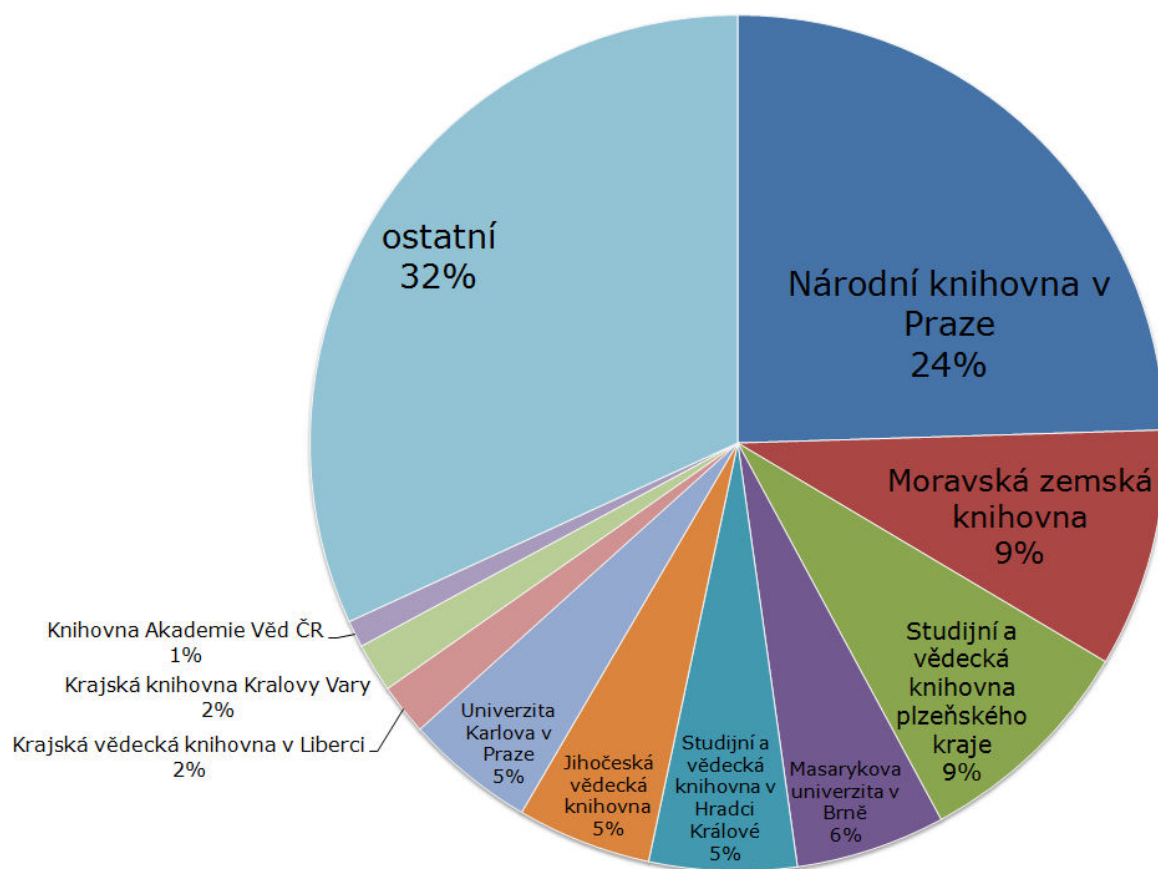
1. Skenované obrázky vkládané "ručně" přes webové rozhraní obálkyknih.cz.
2. Obálky vkládané automatizovaně pomocí webového API spolupracujícími institucemi.
3. Dotazování poskytovatelů na každé neznámé ISBN (*pull* přístup).
4. Pravidelné čtení seznamu nových ISBN vydavatele a jejich stahování (*push* přístup).

Stahovaná data jsou držena v primární databázi na serveru obálkyknih.cz. Pro omezení zátěže je však možné nakonfigurovat několik *cache* serverů, které periodicky stahují nové prvky z primární databáze. Server obálkyknih.cz databáze pravidelně mezi *cache* servery aktualizuje a ověřuje jejich dostupnost (5.).

Výsledkem integrace je oživení webového katalogu knihovny. Čtenáři jsou z různých internetových e-shopů na fotografie produktů zvyknutí, obálky knih na webech knihkupců jsou také samozřejmostí, právem tedy očekávají obálky knih i v knihovních katalozích.

Není proto divu, že už během prvních pár měsíců se využívanost služby vyšvihla na 73 registrovaných knihoven s více než milionem dotazů denně. Nejaktivnějšími knihovnami co do počtu dotazů za měsíc jsou:

Č.	Název knihovny		Počet dotazů
1.	Národní knihovna v Praze	24 %	4902140
2.	Moravská zemská knihovna Studijní a vědecká knihovna plzeňského kraje	9 %	1804571
3.	Masarykova univerzita v Brně Studijní a vědecká knihovna v Hradci Králové	9 %	1715608
4.	Jihočeská vědecká knihovna	6 %	1131600
5.	Univerzita Karlova v Praze	5 %	1115515
6.	Krajská vědecká knihovna v Liberci	5 %	1017048
7.	Krajská knihovna Kralovy Vary	5 %	981232
8.	Krajská vědecká knihovna v Liberci	2 %	386957
9.	Krajská knihovna Kralovy Vary	2 %	371102
10.	Knihovna Akademie Věd ČR	1 %	204142
73.	ostatní	32 %	6370082



V době zátěžových špiček, které kopírují běžnou pracovní dobu, se zátěž pohybuje kolem 30 dotazů za sekundu. Obálkový server však díky vhodným algoritmům a dostatečnému výkonu toto množství dotazů zvládá bez větších problémů. Další vývoj závisí na tom, kolik nových knihoven se do systému zapojí.

2. Nové služby

Vývoj služeb obálkového serveru se však na obálkách knih nezastavil. Na přání uživatelů jsme provedli řadu změn a rozšíření. Nyní si je postupně představíme:

2.1 Identifikátor knihy

Tradičním problémem v knihovnickém světě je jednoznačná identifikace knížky. O tu se snaží různé iniciativy s většími či menšími úspěchy. Obálkový server proto ke knížce eviduje celou řadu identifikátorů -- 13 místné ISBN (EAN kód), OCLC, LCCN či v Národní knihovně čerstvě vznikající národní identifikátor záznamu. Na knihu se pak dá odkázat kterýmkoliv z těchto identifikátorů."

Kromě toho se ke knížce evidují také základní bibliografické údaje (název, autor, nakladatel, rok). Tyto údaje jsou jednak použity při zobrazování informací o knížce čtenáři, výhledově však také počítáme s možností identifikace knihy i pomocí těchto "OpenURL" údajů. Evidence 13 místného EAN kódu místo 10 místného ISBN navíc umožňuje evidovat nejen

obálky knízek, ale případně i jiných doprovodných předmětů -- audiovizuálních nosičů CD/DVD apod.

2.2 Obsahy knih

Užitečnou funkcí čtenářského katalogu je také možnost prohlédnout si obsah knížky. Pro zajištění této služby server obalkyknih.cz navazuje na projekt TOC Národní knihovny v Praze [<http://toc.nkp.cz>], který nabízí spolupracujícím knihovnám jednotnou linku skenování obálek a obsahů knih. Pro obsahy také vytváří textovou OCR vrstvu, stažený PDF soubor s obsahem je tudíž prohledávatelný. Zapojené knihovny takto doposud zpracovaly celkem cca 30.000 knih.

Pro server obalkyknih.cz projekt TOC slouží jako zatím jediný zdroj obsahů knih. I když 30.000 knih tvoří zdánlivě pouze nepatrnou část knižní produkce, jde zejména o nové a často žádané české knihy a díky pečlivému výběru je tak reálné využití prakticky mnohem větší.

Server obalkyknih.cz knihovnímu katalogu ke knížce nabízí jak odkaz na PDF se samotným obsahem knihy, tak náhledový obrázek první stránky obsahu. Dále je možné ke knížkám stáhnout textovou verzi obsahu pro případnou indexaci textu obsahu vyhledávacím mechanismem knihovního katalogu, čímž je čtenáři umožněno vyhledávat knížky v katalogu i zadáním slov z obsahu.

2.3 Recenze, anotace, čtenářská hodnocení

Ke knihám jsou dále evidovány anotace a recenze knih. Zde mezi hlavní zdroje patří databáze SČKN, o zpřístupnění cenné databáze Městské knihovny v Praze se jedná. Projekt obalkyknih.cz se v tomto případě snaží přispět integrací různých zdrojů do jednoho přístupového místa.

Kromě toho server obalkyknih.cz umožňuje čtenářům zadávat i vlastní hodnocení. Knihu lze hodnotit formou známých "hvězdiček" či slovním komentářem knihy. Tyto příspěvky před zobrazením v knihovních katalozích podléhají kontrole ze strany správců databáze, přičemž příspěvky s nevhodným obsahem jsou z nich mazány.

2.4 Prodejci dané knihy

Vedlejším efektem stahování údajů o knížkách z webů nakladatelů a knihkupců je získání informace o možnosti zakoupení dané knížky. Při stahování údajů se proto ukládá i odkaz na knihu na webu knihkupce a také její cena. Knihovna, pokud má zájem, pak může čtenáři nabídnout odkazy na weby knihkupců, kde si čtenář danou knížku může zakoupit. Odkazy jsou také zobrazeny na stránce knížky na serveru obalkyknih.cz. To může čtenáři pomoci například když ve fondu knihovny už není zbývajících volných exemplářů knížky.

3. Integrace do OPACů

A jak postupovat, pokud chcete o obálky knih obohatit i váš katalog?

Je to jednoduché -- stačí se zaregistrovat na domovské stránce projektu [<http://www.obalkyknih.cz>], kde je potřeba vyplnit zejména siglu knihovny a URL adresu

vašeho webového katalogu. Adresa URL katalogu je důležitá -- při dotazu na obálkový server je tato kontrolována a obálky se zobrazují pouze na stránkách registrovaných katalogů.

Server obalkyknih.cz dodavatelům knihovních systémů nabízí dva způsoby integrace do knihovního katalogu -- buď bez programování, kdy stačí pouze do HTML kódu stránky na správná místa vložit speciální značky, nebo s programováním, kdy si knihovna může definovat způsob zobrazení zcela dle svého uvážení a vkusu.

3.1 HTML API

V tomto případě stačí vložit do stránky do sekce zobrazující bibliografické informace o knížce následující HTML kód:

```
<script type="text/javascript" src="http://www.obalkyknih.cz/api/functions.js" />
<script type="text/javascript" src="http://www.obalkyknih.cz/api/custom.js" />

<div class="obalky_book">
  <div class="obalky_permalink">http://aleph.vkol.cz/F?func=find-
c&cccl_term=sys=000557276</div>
  <div class="obalky_bibinfo">
    <div class="obalky_title">S eleganci ježka</div>
    <div class="obalky_authors">Barbery, Muriel</div>
    <div class="obalky_year">2008</div>
    <div class="obalky_isbn">9788072942862</div>
    <div class="obalky_oclc"></div>
  </div>
  <div class="obalky_callback" name="obalky_display_thumbnail"></div>
</div>
```

API zajistí nalezení "obalky_book" sekcí, ve kterých dohledá bibliografické (identifikační) údaje o stránce, dotáže se na příslušnou knížku obálkového serveru a vloží do stránky požadované prvky (v tomto případě obálku).

Důležitým parametrem je "obalky_permalink" -- stálý a neměnný URL odkaz na záznam ve vašem knihovním katalogu. To slouží k jednoznačné identifikaci záznamu z dané knihovny na serveru obalkyknih.cz

3.2 JavaScript API

Tato možnost vyžaduje znalost JavaScript programování. Do stránky se opět přidá odkaz na JavaScript knihovnu uloženou na obalkyknih.cz. Ta disponuje funkcemi, kterým se předají bibliografické údaje knížky a funkce vrátí další obohacující prvky -- rozměry a URL obrázku s obálkou, náhledový obrázek a PDF adresa obsahu, seznam anotací či recenzí apod.

Je pak plně v režii knihovny, jakým způsobem prezentuje získané údaje. Pomocí JavaScriptu je možné stránku s knihovním záznamem pozměnit a vložit do ní získanou obálku, obsah, anotaci apod. Jediným požadavkem ze strany projektu je, aby obrázek obálky byl klikatelný a vedl na vrácenou URL (typicky na stránku o knížce na webu obalkyknih.cz).

Podrobný popis všech JavaScript funkcí včetně ukázkového kódu je uveden na stránce projektu.

3.3 Podpora v knihovních systémech

Pro knihovní systém Aleph společnosti Ex-Libris je na stránkách projektu popsán podrobný návod, jak lze integraci provést.

Dodavatelé dalších knihovních systémů slíbili podporu pro nové rozhraní v nejbližší době.

4. Další vývoj

V dalších fázích projektu plánujeme provést celou řadu dalších zlepšení a rozšíření.

4.1 Systém Doporučování literatury

Systém Doporučování literatury slouží k nabízení odkazů na knížky, které nějakým způsobem souvisí s čtenářem právě prohlíženou knihou. Tyto systémy mají svůj původ ve větších e-shopech, kde se zákazníkovi nabízí zboží, které spolu s právě prohlíženým zbožím koupili předešlí zákazníci.

Doporučování literatury lze opřít buď o údaje o výpůjčkách knih (kdy sledujeme kombinace spolu vypůjčených knih) nebo se k určení podobnosti spokojíme s pouhým zobrazením stránky záznamu o knížce. Pokud si v katalogu více různých čtenářů nezávisle na sobě v rámci svého "sezení" prohlédne stejnou kombinaci knih, tyto knihy spolu pravděpodobně souvisí. Míra pravděpodobnosti je pak dána frekvencí prohlížení této kombinace knih.

Systém obalkyknih.cz nemá údaje o výpůjčkách čtenářů, na druhou stranu má však bohaté údaje o zobrazovaných knihách v knihovních katalozích. Díky tomu může budovat statistický model potřebný pro vytvoření párů souvisejících knih (často společně prohlížených), a tyto knihovnímu katalogu nabídnout pro zobrazení čtenáři. Tuto znalost může budovat napříč katalogy všech zapojených knihoven, doporučovat však může pouze knížky, které má knihovna ve svém v katalogu.

4.2 Deduplikace záznamů

V tuto chvíli se jednotlivé prvky databáze (obálka, obsah, hodnocení, recenze,...) vážou vždy na konkrétní knihu. Pokud má však knížka např. více vydání, některé údaje by bylo vhodné sdílet. Typicky se např. nemění celkové hodnocení díla, obsah bývá také často téměř shodný. Z těchto důvodů plánujeme agregovat více různých vydání díla a v případě, že to není na závadu věci, nabízet obálku, obsah, či hodnocení alespoň z jiného vydání.

S tím souvisí také podpora ISSN identifikátoru a logika zobrazování obálek u seriálových řad, zpracování vícesvazkových děl apod.

4.3 Dostupnost knihy v jiných knihovnách

Deduplikováním bibliografických údajů různých knihoven je možné také zobrazovat odkazy na knížku do webových katalogů jiných knihoven. To má pro čtenáře smysl zejména tehdy, když ve vaší knihovně už není zbývající volný exemplář knížky, případně může informace sloužit pro meziknihovní výpůjční službu.

4.4 Další rozšiřování báze dat

Dlouhodobým úsilím serveru obalkyknih.cz je komunikace s nakladateli a knihkupci za účelem dalšího rozšiřování databáze obalkyknih.cz. Výhoda pro knihkupce je zřejmá -- knihovny, případně server obalkyknih.cz, umožňuje čtenáři knížku u daného knihkupce zakoupit.

Registrace nového knihkupce nebo vydavatele je přitom snadná. Využívá stávající *de-facto* standard agregačních služeb serverů zbozi.cz či heureka.cz, pro které knihkupci už teď nabízí seznam svých knih ve formě stažitelného periodicky aktualizovaného XML souboru s popisem svého zboží. Stejný formát souboru je používán i serverem obalkyknih.cz.

Závěr

Primárním úkolem projektu obalkyknih.cz je agregace různých zdrojů informací o knížkách do jedné snadno použitelné webové služby. Centralizované řešení služby přitom nesmírně zjednodušuje logiku aplikace a ulehčuje její administraci.

Systém obalkyknih.cz se za poslední období proměnil z prosté databáze obálek na platformu, která věrně modeluje knihovnický svět. Eviduje knížky, obálky, obsahy, knihovny a jejich knihovní záznamy, nakladatele či knihkupce a jimi nabízené zboží. Eviduje čtenáře, jimi prohlížené knížky a jejich hodnocení knih.

Služby systému obalkyknih.cz využívá celá řada knihoven všech velikostí i zaměření, včetně většiny velkých knihoven. Navíc trend v zapojování nových knihoven je pořád rostoucí.

Pokud chcete svým čtenářům nabídnout moderní služby bez velkého úsilí, služba obalkyknih.cz tak může být dobrým začátkem i pro vás.