

Katalogizace knih pro knihovní systém pomocí otevřených dat

MIROSLAV ZÍKA

Fakulta pedagogická ZČU v Plzni

Knihovní služby v době mimořádných opatření zaznamenaly nesčetné omezení. Dne 13. 3. 2020 bylo přijato usnesení č. 199 o přijetí krizového opatření [1], které zakazovalo přítomnost veřejnosti v knihovnách. V průběhu roku docházelo k vydávání doporučení, přičemž v současné době platí opatření č. 575 [2]. V něm jsou vymezené podmínky pro provoz knihovních a dalších kulturních institucí. Během té doby musely knihovny nabídnout alternativní způsoby zapůjčování literárních děl veřejnosti. Jejich možnosti byly ovšem značně rozdílné kvůli různorodosti jejich zázemí.

Výsledky těchto aktivit mapuje ve čtyřech krajích¹⁾ série článků *Dopad koronavirové situace na knihovny* internetový měsíčník *Čtenář*. Mezi zmíněné možnosti výpůjček patří:

- donáška knih do domu [3],
- bezkontaktní výpůjční systém,
- poskytování e-knih (např. akce *#knihyprotiviru*).

V případě prvních dvou zmíněných možností je nutno počítat s předpokladem, že čtenáři mají k dispozici seznam knih obsažených v knihovním fondu – ať již v podobě elektronického katalogu nebo např. v méně komfortní papírové podobě. Pouze tak mohou čtenáři vybírat z dostupných knížek a následně komunikovat s personálem za účelem výpůjček.

Problematiku zprostředkování knihovního fondu čtenářům jsme řešili také v knihovně Mladého Smolivce, která se i přes menší velikost dá považovat za nedílnou veřejnou službu v obci. Článek tudíž může sloužit jako materiál uvádějící do problematiky využití otevřených dat a jako shrnutí poznatků získaných během vývoje zmíněného knihovního systému.

Podmínky kladené na knihovní systém

Mezi primární podmínku vzniknuvšího systému byla jeho dostupnost. Vzhledem k tehdy platným vládním nařízením se realizační tým přiklonil

¹⁾Středočeský kraj, kraj Vysočina, Ústecký kraj a Liberecký kraj.

k podobě webové stránky disponující vyhledáváním v databázi. Na webový portál je nutno z hlediska funkcionalit pohlížet ze dvou rozdílných rolí:

- *Čtenář*: disponuje přístupem do katalogu a v případě ověřeného účtu také právem rezervace knihy k zapůjčení.
- *Knihovník*: rozšíření role čtenáře, kterému je navíc umožněno potvrzovat výpůjčky, potvrzovat vrácení a vkládat nové knih do systému s případnou změnou vložených údajů.

Dalšími kritérii byla jednoduchá správa výpůjčního procesu a správa knih v knihovním fondu. V počátečním konceptu měl čtenář k dispozici kromě názvu knihy a jeho autora také stručnou anotaci, žánrové zařazení, jazyk textu a miniaturu obálky knihy. K těmto údajům následně přibýly další doplňující údaje, jež v případě výběru nehrají klíčovou roli, ale mohly by je využít případní čtenáři-studenti za účelem citací, tj. nakladatel, číslo vydání, rok vydání a ISBN. Na základě zmíněných údajů bylo možné navrhnout strukturu tabulky, ale hlavně vyvodit vyšší náročnost evidence záznamů, převážně z důvodů delších anotací, neboť i pouhé vyhledávání textu by bylo časově náročné.

Existující knihovní systémy

Anna Stöcklová ve svém článku *Automatizace v knihovnách České republiky* zmiňuje přehled automatizovaných knihovních systémů [4]. Jedná se o systémy disponující kromě základních funkcionalit také modulem katalog, který by umožňoval potřebnou evidenci a prohlížení knihovního fondu. Pravděpodobně nejznámější z představených automatizovaných systémů je knihovní systém ALEPH, který využívá minimálně knihovna Západočeské univerzity v Plzni.

Všechny ve článku uvedené systémy mají nicméně pár nevýhod, především placené využívání a omezené možnosti zásahu do grafického vzhledu, kvůli nimž jsme se nakonec rozhodli pro vlastní implementaci databázového řešení.

Otevřená data

Evidence knižních záznamů v podobě datových entit by dle výše uvedeného řešení byla na správci webu, popřípadě knihovnici se stejnými přístupovými právy. Jak již bylo zmíněno výše, vzhledem k většímu počtu informací vázaných k objektu a počtu knih nutných k evidenci, bylo nutno vyhledat prostředky ke snížení časové náročnosti evidence.

Dostupnost knižního fondu přes webové rozhraní nabízelo využití principů otevřených dat a s nimi spojené webové služby. Otevřená data jsou velice široký pojem, což vyplývá z definice organizace Open Knowledge Foundation [5]: „Otevřená data jsou data, která může kdokoliv volně použít, opakovaně využít a redistribuovat pouze s výhradou požadavků na atributy sdílení.“ Nutno zmínit, že se neřeší ani účel využití těchto dat. Aby data byla ovšem považována za otevřená, musí splňovat kritéria níže [6]:

Úplnost dat: data by měla být poskytována v co největším rozsahu; nemělo by tedy docházet k odstraňování konkrétních záznamů a současně by měly být udržovány v co nejaktuálnějším stavu.

Strojová čitelnost: data jsou ukládána ve formátech, které je možné číst pomocí programových aplikací. Jedná se o formáty volně dostupné, nebo formáty bezplatně do nich převoditelné. Např. html, pdf, jpeg, xml, . . .

Dostupnost: je podmíněno minimálním vynaložením nákladů pro získání otevřených dat, ačkoliv poskytovatel může požadovat o zpoplatnění přístupu ve výši nákladů spojených s jejich zpřístupněním veřejnosti. Podmínky využití musí být jasně definovány pomocí uvedení licence. Je dělena do pěti kategorií.

Na internetu je množství dat ve výše zmiňovaných formátech, nicméně to samotné z nich nečiní data otevřená. Co vytváří otevřená data opravdu specifickými a opravdu otevřenými je právě kombinace výše uvedených vlastností – převážně strojová čitelnost a dostupnost. Dohromady vytvářejí prostředek k vývoji softwarových aplikací a odborných analýz [6]. Jejich dostupnost a jednoduchost čitelného čtení umožňuje nejen programátorům, ale také vědcům a novinářům rozvíjet dále význam těchto dat a pomáhat tím k rozvoji společnosti, což je i případ tematiky tohoto článku.

Webové služby a čerpání otevřených dat

Jedná se o nástroje využívající protokol HTTP(S). Pomocí nich je zajištěna komunikace napříč aplikacemi, přičemž pod pojmem komunikace dochází k podobnému činnosti jako je načítání webové stránky. Na základě požadavku ze strany klienta dochází ke zpracování vstupních dat webovou službou, která zpět odešle odpověď v podobě požadovaných dat, nejčastěji v podobě XML souboru. Vzhledem k prostředí, ve kterém komunikace probíhá, je možné využít ostatní webové technologie k samotné komunikaci – tedy především skriptovací jazyky JavaScript, PHP či jejich kombinaci v podobě AJAXu [7].

Google Books API

Typ webové služby: REST

Google Books je vyhledávací služba od společnosti Google, která vznikla v roce 2004. Jedná se o projekt, jehož cíle je digitalizace tištěných knih a umožnění jejich prohlížení v rozsahu dle autorského zákona. V současné době je do projektu zapojeno více druhů institucí, neboť pomocí této služby je umožněno knihy prohlížet, zapůjčovat, ale i prodávat [8].

Součástí projektu je také webová služba Google Books API, která bez poplatků umožňuje přístup k informacím o evidovaných knižních entitách. Pro plné využívání je nutno službu provázat s Google účtem. Portálu knihovny službu využívalo k získání informací na základě ISBN, nicméně starší knihy nebylo možné takto vyhledávat. V případě novějších knih došlo při vyhledání na základě ISBN i k zobrazení neodpovídajících záznamů. Výčet by bylo možné zredukovat přidáním dalšího vyhledávacího parametru, např. jména autora.

Výsledky zobrazené na obr. 1 byly získány podáním požadavku následující URL adresou: <https://www.googleapis.com/books/v1/volumes?q=9788024711164>

Rostlinolékař	Dorothea Baumjohann, Peter Baumjohann	2007	Zapiš
Tachles, Lustig	Arnošt Lustig, Karel Hvizďala, Jan Malý	2011	Zapiš
Neviditelný	C.D. Payne	2012-01-01	Zapiš
Tvorba informačních systémů	Bruckner Tomáš, Voříšek Jiří, Buchalceová Alena a kolektiv	2012	Zapiš
Připravujeme zahradu na jaro a léto	Pasečný Petr	2012-03-19	Zapiš
Moje první pohádky	Marie Tetourová	2020-01-01	Zapiš
Biosyntéza	Boadella David, Carletonová Jacqueline A., Frankelová Esther, Rodriguesová Eunice	2012	Zapiš
Mimozemšťani útočí! : případ č. 1	Thomas Brezina	2012	Zapiš
Antonín Vitvar – 20 let na Jawě a Nortonu	Vitvar Jan	2012	Zapiš
PHP a XML	Jiří Kosek	2009	Zapiš

Obr. 1 Vykreslené výsledky vyhledávání ISBN 978-80-247-1116-4 (PHP a XML)

Parametr q představuje vyhledávaný řetězec, který je v případě potřeby možné rozšiřovat o doplňující informace konkretizující vyhledávanou knihu. Odpověď webové služby Google Books API, kterou zobrazí i prostý prohlížeč, je objekt typu JSON s jednotlivým záznamy odpovídající krité-

riím vyhledávání. Přístup k atributům je znázorněn na obr. 2.

```
$.getJSON("https://www.googleapis.com/books/v1/volumes?q=" + d, function (c) {
    "undefined" == typeof c.items
    ? print_modal("\u017d\u00e1dn\u00e1 kniha se neshoduje s ISBN.")
    : 1 < c.items.length && 0 == b
    ? print_modal(modal_vypis_knihy(c.items), 1)
    : ($("#nazev-js").val(c.items[a].volumeInfo.title),
      $("#autor-js").val(c.items[a].volumeInfo.authors.join(", ")),
      $("#rok-js").val(c.items[a].volumeInfo.publishedDate.substring(0, 4)),
      "cs" == c.items[a].volumeInfo.language && $("#jazyk-js").val("\u010c\u00e9\u0161tina"),
      c.items[a].volumeInfo.description != null && $("#anotace-js").val(c.items[a].volumeInfo.description),
      c.items[a].volumeInfo.publisher != null && $("#vydavatel-js").val(c.items[a].volumeInfo.publisher));
});
```

Obr. 2 Načtení dat pomocí JS

Kromě základních informací jsou záznamy doplněny o další atributy, např. cenu knihy, počet stránek, zda je výtisk k prodeji aj. Podrobné informace k implementaci webové služby můžete nalézt v [dokumentaci](#).

knihovny.cz

Typ webové služby: Z39.50

Jedná se o portál vycházející z projektu Koncepce rozvoje knihoven ČR na léta 2011–2015, který má za jeden z cílů umožnit přístup do knižních katalogů. Na této spolupráci se podílí již více jak šest desítek knihoven, včetně 4 univerzitních. Kromě toho umožňuje služba pomocí parametru prohledávat více zdrojů [9]:

- souborný katalog ČR,
- výběr ze zahraničních knihoven,
- staré tisky,
- články,
- zaměření dle jazyka:
 - anglicky mluvící země,
 - německy mluvící země,
 - slovensky mluvící země.

Vyhledávání je na webu knihovny realizováno na základě uvedeného názvu díla a jména autora (stačí příjmení). Oproti *Google Books API* dokáže služba dohledat informace i o starších knižních titulek, dochází ale k nalezení více záznamů, což se projevuje v delší době zpracování dat a následného zobrazení výsledků.

Nevýhodou jsou také rozličnosti ve struktuře dat a způsobu jejich plnění, neboť centralizovaná data nejsou pravděpodobně nijak centrálně validována. Aspoň to by mohl být důvod, proč dochází k neúplnosti dat,

k jejich nečitelnosti nebo i k prohození údajů uvnitř struktury (např. cena místo ISBN). Vzhledem k rozličnostem v evidenci byly implementovány dvě funkce – zápis celého záznamu a zápis pouze anotace.

Nalezeno záznamů: 368

Záznam: 1 <input type="button" value="Zapsat"/> <input type="button" value="Zapsat anotaci"/>		
Autor: Dumas Alexandre	ISBN: 46.00	Název: Tři mušketýři.
Nakladatelství: Albatros	Rok: 1987	Vydání: 5. vyd.

Záznam: 2 <input type="button" value="Zapsat"/> <input type="button" value="Zapsat anotaci"/>		
Autor: Dumas Alexandre	ISBN: 33.00	Název: Tři mušketýři po dvaceti letech.
Nakladatelství: Albatros	Rok: 1982	Vydání: 1. vyd.

Záznam: 3 <input type="button" value="Zapsat"/> <input type="button" value="Zapsat anotaci"/>		
Autor: Dumas Alexandre	ISBN: 32.00	Název: Tři mušketýři ještě po deseti letech aneb Vikomt de Bragelone.
Nakladatelství: Albatros	Rok: 1978	Vydání: 1. vyd.

Záznam: 4 <input type="button" value="Zapsat"/> <input type="button" value="Zapsat anotaci"/>		
Autor: Doblado Ana	ISBN: 978-80-7567-819-5	Název: Tři mušketýři /
Nakladatelství: Sun	Rok: 2021	Vydání: 2. vydání
Array		

Obr. 3 Vykreslené výsledky vyhledávání pro „tři mušketýři“ a „Dumas“ (pouze část výsledků)

Data získaná pomocí webové služby *knihovny.cz* se zpracovávají do zobrazitelné podoby pomocí skriptovacího jazyka PHP a funkce pro práci s protokolem Z39.50, viz obr. 4.

```
$yaz = yaz_connect("www.knihovny.cz:9000");
yaz_database($yaz, "cpk_caslin");
yaz_range($yaz, 1, 10);
yaz_search($yaz, "rpn", "&and @attr 1=4 *. $kniha." @attr 1=1003 *. $autor.");
yaz_wait();
$error = yaz_error($yaz);
if (!empty($error)) {
    echo "Chyba: $error\n";
} else {
    $hits = yaz_hits($yaz);
    echo "<pre>Nalezeno záznamů: $hits</pre>";
    for ($p = 1; $p <= $hits; $p++) {
        $kniha = array("id" => "", "isbn" => "", "autor" => "", "nazev" => "", "vydani" => "", "vydavatel" => "", "rok" => "", "anotace" => "", "zann" => "");
        $kniha["id"] = $p;
        $rec = yaz_record($yaz, $p, "xml");
        $reader = simplexml_load_string($rec);
        //print_r($reader);
        if (empty($rec)) break;
        /*
        if(strpos($reader->controlfield[3], "cze") === false){
            continue;
        }
        */
        //echo $reader->controlfield[3];
        //print_r($reader);
        foreach ($reader->datafield as $value){
    }
    }
}
```

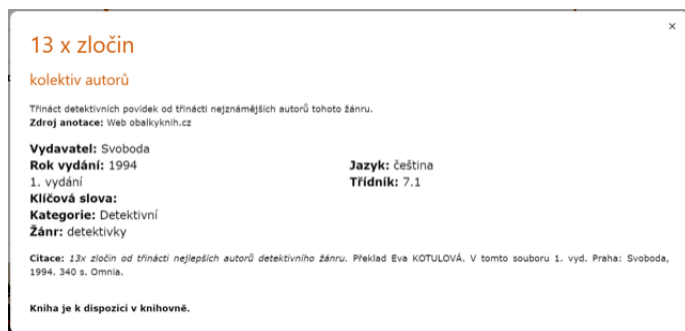
Obr. 4 Zpracování dat ze služby knihovny.cz pomocí PHP

Získání dat je zde realizováno pomocí funkce `yaz_search`, jejíž parametry jsou údaje k připojení k databázi a řetězec definující vyhledávací sekvenci – konkrétněji zde dochází k vyhledávání v závislosti na jménu knihy a jejího autora (zpracování vstupních dat není ve skriptu zobrazeno). V případě nechybového vyhledávání jsou v konstrukcích `for` a `foreach` postupně zpracovány získané záznamy k vykreslení do šablony.

obalkyknih.cz

Typ webové služby: REST

Webová služba vznikla v roce 2008 v rámci projektu VISK3 a stojí za ní, obdobně jako u služby *knihovny.cz*, Moravská zemská knihovna. Primárním cílem této služby byla však centralizace obálek knih za účelem poskytování náhledů do katalogů ostatních knihoven. Na vývoji služeb se postupně podílela Vědecká knihovna Olomouc, Městská knihovna v Praze, Sdružení knihoven ČR a Jihočeská vědecká knihovna v Českých Budějovicích. V současné době databáze obsahuje přes 2,4 miliónů obálek a 650 tisíc obsahů českých i zahraničních publikací. Služba totiž kromě obálek a obsahů poskytuje také anotace, hodnocení knih a citace dle normy ISO 690. Na obrázku níže jsou zobrazeny informace o zvolené knize. Kromě údajů z interní databáze portálu je zde uvedena citace a se zdrojem zvýrazněna anotace, protože je získána pomocí služby *obalkyknih.cz* (pokud interní záznam anotaci nemá) [10].



Obr. 5 Zobrazení informací o knize pomocí modálního okna

Služba poskytovaná serverem *obalkyknih.cz* je podmíněna registrací, kdy je do účtu knihovny nutno uvést IP adresu či URL adresu katalogu, a vytvořením zpětného odkazu na stránky poskytovatele. V případě knihovních stránek je požadovaný odkaz vytvořen na poskytované obálce knihy.

Na obr. 6 je vykreslen výsledek úspěšného vyhledání s parametrem ISBN 80-205-0441-9. Pro knihu *15 pátračů* s ISBN 01-080-87 nebylo nic nalezeno, byl tedy automaticky nahrán obrázek o velikosti 1 × 1 px s transparentním pozadím.



Obr. 6 Vyhledávání v katalogu pro žánr Detektivní

Odeslání požadavku o data ze služby *obalkyknih.cz* je kombinací předchozích dvou služeb. Požadavek je definován URL adresou, která obsahuje parametry vyhledávání, a získaná odpověď je zpracována pomocí PHP. V ukázce se jedná o ISBN a části názvu: [http://cache.obalkyknih.cz/api/books?multi=\[{"isbn":"XXX","part_name":"XXX"}\]](http://cache.obalkyknih.cz/api/books?multi=[{) I zde by bylo možné si výsledek zobrazit v prohlížeči, služba je ale omezena pouze na registrované domény a IP adresy. Z toho důvodu je odpověď ještě zpracována právě pomocí PHP, který je prováděn na serveru a klient získává celistvou vykreslenou stránku se všemi údaji.

Podrobné informace k implementaci webové služby můžete nalézt v [dokumentaci](#).

Zprovoznění systému

Do knihovního systému bylo v rámci testování evidováno 50 knih – nejprve pomocí vyhledávání, následně bez. Časová náročnost evidence, kterou by bylo možné ještě snížit redukcí nepovinných údajů, je zobrazena v tabulce 1.

Styl vkládání	Časová náročnost
Vyhledání údajů	2 hodiny a 12 minut
Manuální vepsání všech údajů	2 hodiny a 39 minut

Tabulka 1: Přehled časové náročnosti evidence knih do systému

Rozdílnost v evidenci daného počtu knih bez vyhledávání a s vyhledáváním je 27 minut. Jedná se o relativně nízké ušetření času, které je pravděpodobně způsobeno následujícími faktory.

- Systém nedovede pomocí skriptů vybírat žánr knihy.
- Je nutno z vyhledaných údajů vybrat správný např. dle roku vydání či vydavatele.
- Po doplnění údajů z otevřených dat je nutno provést kontrolu.
- Je nutno vyplnit unikátní evidenční číslo knihy²⁾ a její třídění dle umístění v knihovně.

Oproti tomu manuální zapsání všech údajů je nejvíce problematické v nalezení informací ve vydavatelské tiráži, žánrového zařazení knihy a případné vepsání stručné anotace. Ze začátku byl žánr vybírán dle posouzení, následně se k tomu využívala Databazeknih.cz, která velkou měrou usnadnila získání údajů o žánru i anotaci.

Závěr

Článek pojednává o využití otevřených dat za účelem komfortnější evidence knižních titulů pro knihovnu Mladý Smolivec. Stručně seznamuje čtenáře s třemi webovými službami (Google Books API, knihovny.cz, obalkyknih.cz) a jejich využitím pro vytvořený portál knihovny, což je doplněno názornými skripty zpracující v odpovědi získaná data. Cílem článku je poskytnout čtenáři náhled do případné automatizace a digitalizace odvětví, které má na první pohled s touto tematikou velmi málo společného. Současně jsou mu také představena *otevřená data*, která jsou obsahově mnohem širší než ve zde zmíněné tematice. O jejich využití je snaha i ze strany vlády České republiky.

První webová služba, která byla v portálu knihovny využívána za účelem efektivní evidence knih, *Google Books API*, z počátku působila jako dostačující. Byly v ní k dispozici všechny potřebné informace o knize – od autora po zkratku země, ve které kniha byla publikovaná. V mnohých případech zde byly k dispozici i náhledy knížek, které by bylo ovšem nutno ukládat do databáze, protože nebylo možné získat pouze jednu konkrétní hodnotu. Tato služba se ovšem ukázala jako nedostačující pro starší knižní tituly, kterých byla většina. Z toho důvodu byly využity ještě dvě webové

²⁾Knihy mají v databázi unikátní číselné ID, ale to nebylo možné využít.

služby, které mají zázemí v České republice – *knihovny.cz* a *obalkyknih.cz*. Ani tyto webové služby nelze vždy efektivně využít za účelem snadného získání údajů ke všem knihám, hlavně v případě získání obálek knih, nicméně obě služby nabízejí širokou škálu dalších informací či doplňujících záznamů, pomocí nichž je možné potřebné údaje „poskládat“ dohromady.

Literatura

- [1] Usnesení vlády České republiky č. 199. Vláda České republiky. [Online] 12. Březen 2020. <https://apps.odok.cz/attachment/-/down/IHOABMNHPPAD>.
- [2] Mimořádné opatření č. 575. Vláda České republiky. [Online] 21. Červen 2021. <https://www.vlada.cz/assets/media-centrum/tiskove-zpravy/w210621a-0575.pdf>.
- [3] *Hradil, J.*: Téma: Dopad koronavirové situace na knihovny: Mimořádná opatření s vlivem na úpravy a rozvoj služeb (výsledky dotazníkového...). Čtenář – měsíčník pro knihovny. [Online] Listopad 2020. <https://www.svkkl.cz/ctenar/clanek/3216>.
- [4] *Stöcklová, A.*: Automatizace v knihovnách České republiky. Ikaros. [Online] 2006. <https://ikaros.cz/automatizace-v-knihovnach-ceske-republiky>.
- [5] *Dietrich, D., a kol.*: Open data handbook. [Online] <http://opendatahandbook.org/guide/en/what-is-open-data/>.
- [6] Ministerstvo vnitra České republiky. Standardy publikace a katalogizace otevřených dat VS ČR. Otevřená data. [Online] 2015. https://opendata.gov.cz/_media/standardy_publikace_a_katalogizace_otevrenych_dat_vs_cr.pdf.
- [7] *Kosek, J.*: PHP a XML. Grada, Praha, 2009.
- [8] Google Books. About Google Books. Google Books. [Online] Google. <https://www.google.com/intl/en/googlebooks/about/index.html>.
- [9] knihovny.cz. O portálu. Knihovny.cz. [Online] Moravská zemská knihovna. <https://www.knihovny.cz/Portal/Page/o-portalu>.
- [10] Obálky knih. O projektu Obálkyknih.cz. Obálky knih. [Online] Moravská zemská knihovna, Jihočeská vědecká knihovna v Českých Budějovicích. <https://www.obalkyknih.cz/about>.