

VEŘEJNÉ INFORMAČNÍ SLUŽBY KNIHOVEN

**Koordinální centrum programu a implementace Koncepce rozvoje knihoven
v České republice**

Obálky knih.cz - rozvoj projektu v roce 2023

**Jihočeská vědecká knihovna v Českých Budějovicích
leden 2024**

ZHODNOCENÍ PROJEKTU

Projekt Obálkyknih.cz sdružuje různé zdroje informací o dokumentech do jedné, snadno použitelné webové služby. Databáze aktuálně obsahuje přes **2,85 miliónu obálek** (nárůst za rok 2023 o cca 220 000 obálek), **858 tisíc obsahů** českých a zahraničních dokumentů (nárůst za rok 2023 o cca 63 000 obsahů) a **18 500** seznamů doporučené literatury (nárůst za rok 2023 o cca 3 500 seznamů). Dále poskytuje přes **756 tisíc anotací, 23,83 miliónu hodnocení u 330 tisíc titulů, 8,6 tisíc komentářů, 62 tisíc fotografií autorit** a cca. **1,5 miliónu vygenerovaných citací** dle normy ISO 690.

API služby projektu využívá většina knihoven v České republice, muzea, archivy, oborové projekty, CPK, aj.

Správce projektu Obálkyknih.cz je Jihočeská vědecká knihovna v Českých Budějovicích a projekt provozuje ve spolupráci s Moravskou zemskou knihovnou v Brně. JVK i MZK do projektu z vlastních zdrojů vkládají nemalé lidské a finanční zdroje. Do spolupráce na projektu se zapojují taktéž desítky knihoven z celé republiky (přispívání obálek a obsahů, kontroly dat, připomínkování vývoje, ...).

Přehled vlastností projektu:

- hlavní servery jsou provozovány v **Jihočeské vědecké knihovně v Českých Budějovicích**, záložní servery jsou v **Moravské zemské knihovně v Brně** a v **Centru vědecko-technických informací SR v Bratislavě**
- v případě výpadku jednoho ze serverů mají knihovní systémy možnost přejít během několika vteřin na záložní stroj bez ztráty dostupnosti služeb pro své čtenáře
- měsíčně hlavní server odbaví průměrně **50 miliónů** požadavků, cca. **2 milióny denně**, průměrně **25 dotazů za vteřinu**
- ve špičkách (9-15:00) odbavují servery **40-80 požadavků za vteřinu**
- denně do databáze je nově nahráno nebo upraveno průměrně **500 dokumentů**
- denní přírůstek dat činí **8 GB**, z nich se následně generují náhledy obálek v různých rozlišeních, PDF dokumenty s obsahy a rozpoznává se text pomocí OCR
- **20 Mbit za vteřinu** je datový tok ven ze serveru a na server což představuje cca. 85% všech dat, které projdou internetovým připojením JVK

Detailní statistiky exportu dokumentů přes skenovacího klienta a kontrolní rozhraní za období leden - prosinec 2023:

Počet odeslaných dokumentů	254 382
Počet uložených obálek (COVER)	242 868
Počet uložených stran obsahů (TOC)	129 110
Počet uložených fotografií autorit	1 009
Počet uložených stran seznamů lit.	17 434

Počty odeslaných stran a titulů přes skenovacího klienta dle jednotlivých knihoven (rok 2023):

STRAN	TITULŮ	SIGLA	NÁZEV
29962	15516	CBA001	Jihočeská vědecká knihovna v Českých Budějovicích
23079	11482	OLD012	Knihovna Univerzity Palackého v Olomouci
19589	5893	ABA013	Národní technická knihovna

17725	3072	BOD010	Masarykova univerzita - Právnická fakulta
15288	10599	ABA001	NKP
14930	7513	BOA001	Moravská zemská knihovna v Brně
11271	5041	ABA004	Národní knihovna
10478	5613	OLA001	Vědecká knihovna v Olomouci
7245	2458	ABA008	Národní lékařská knihovna
7122	2767	ABD001	Knihovna Filozofické fakulty UK
6526	4509	ULG001	Knihovna Ústeckého kraje
5020	2212	KVG001	Krajská knihovna Karlovy Vary
3818	1215	ABA007	Knihovna AV ČR
3557	1197	ABB019	Knihovna Sociologického ústavu AVČR, v. v. i.
3055	1362	CBD007	Akademická knihovna Jihočeské univerzity
2862	2745	BOE451	Knihovna Biskupství brněnského
2851	753	ABA006	Vysoká škola ekonomická v Praze
2675	1180	BOE020	Knihovna Ústavní soud
2356	1195	BOD003	Ústřední knihovna Pedagogické fakulty MU
2282	553	ZLD002	Knihovna UTB ve Zlíně
2256	1043	PNA001	Studijní a vědecká knihovna Plzeňského kraje
2063	1526	HKA001	Studijní a vědecká knihovna v Hradci Králové
2050	420	OSD002	Knihovna VŠB-TUO
1986	1239	LIA001	KVK Liberec
1818	580	BOD031	Ústřední knihovna Fakulty sociálních studií
1790	715	LID001	Knihovna Technické univerzity v Liberci
1538	363	ABD100	ÚK ČVUT
1387	887	HBG001	Krajská knihovna Vysočiny
1103	279	BOD004	Ústřední knihovna Přírodovědecké fakulty MU
1100	866	UHG001	Knihovna Bedřicha Beneše Buchlovana
1088	556	ULD001	Ústřední knihovna UJEP
979	979	ABA010	Knihovna Národního muzea
742	176	BOD022	Středisko vědeckých informací ESF MU
719	405	OSG002	Knihovna města Ostravy
674	352	PNG001	Knihovna města Plzně, p.o.
553	553	ABE313	NIPOS - Národní informační a poradenské středisko pro kulturu
462	107	ABB022	Středisko vědeckých informací Fyziologického ústavu AV ČR, v.v.i.
363	121	ABA003	Knihovnický institut
355	321	ZLG001	Krajská knihovna Františka Bartoše ve Zlíně
333	303	PAG001	Krajská knihovna Pardubice
332	266	OSA001	Moravskoslezská vědecká knihovna v Ostravě
247	169	KMG001	Knihovna Kroměřížska - p.o.
237	131	BOE303	Knihovna Moravské galerie v Brně
235	80	BOD006	Informační centrum, ústřední knihovna Mendelovy univerzity v Brně
222	73	ABA011	Parlamentní knihovna

156	79	KLK001	Středočeská vědecká knihovna v Kladně
142	69	ABD103	Centrum vědeckých informací Fakulty sociálních věd UK
133	105	SOG504	Městská knihovna Chodov
130	32	BOD018	Knihovna FI MU
112	57	ABG312	Knihovna Jabok
77	28	ULE301	Muzeum města Ústí nad Labem
60	17	ABB060	Knihovna Ústavu pro českou literaturu AV ČR
42	18	JID501	Knihovna Univerzitního centra Telč Masarykovy univerzity
19	19	BOD033	Univerzitní knihovna pro studenty se specifickými nároky MU
18	9	ABB001	Knihovna Archeologického ústavu AV ČR, Praha, v.v.i.
10	10	PBG001	Knihovna Jana Drdy Příbram
9	3	ABD010	Knihovna MFF UK - půjčovna skript a učebnic
6	2	ABC023	Česká geologická služba - knihovna
4	4	NAE951	Knihovna (Římskokatolická farnost – děkanství Nové Město nad Metují)
2	1	ABB503	Knihovna ÚACH AVČR, v.v.i.

ÚKOLY ŘEŠENÉ V RÁMCI PROJEKTU V ROCE 2023

V roce 2023 jsme pokračovali v udržovacích pracích, stabilizaci provozu serverů a běhu služeb projektu a vývoji nových funkcionalit, jakožto i v kontrolách dat pomocí zapojených knihovníků.

Jedním z hlavních úkolů byl vývoj systému pro distribuovaný provoz frontend serverů: Cílem bylo umožnit distribuovaný provoz serverů frontend vrstvy v současnosti provozovaných na dvou oddělených hostname cache1 a cache2 na síti Cesnet a na jedním hostname cache3 provozovaném na síti Sanet. Cílem bylo zajistit pod hostname cache.obalkyknih.cz, který byl doposud shodný s hostname cache1, failover mezi servery provozovanými v síti Cesnet. Tj. aby hostname v případě dostupnosti obou frontend serverů rovnoměrně rozkládal zátěž mezi oba dostupné servery a zároveň hostname cache.obalkyknih.cz v případě nedostupnosti jednoho ze serverů byl směrován pouze na momentálně dostupný server.

Ve fázi návrhu bylo počítáno s provozem skupiny jmenných serverů s využitím služby Bind9 a zónovými záznamy domény obalkyknih.cz, kterých obsah by kontroloval v minutových intervalech script arbitra schopný provádět změny v záznamech domény a zajišťoval by požadované funkce loadbalancingu i failover. Jako výsledné řešení bylo následně zvoleno použití služby PowerDNS, který je schopen provozu vícero tzv. backend modulů mimo jiné i Bind, MySQL a GeoIP. Všechny zmíněné moduly se v případě obálek knih využívají a ve výsledku vzniká řešení jmenného serveru, který je možné konfigurovat pomocí standardních doménových záznamů i pomocí databáze MySQL, kdy doménové záznamy jsou importovány do MySQL databáze a ty jsou použité k řízení jmenného serveru. Navíc je umožněno použít skriptovacího jazyka Lua a databáze GeoIP k řízení doménových záznamů.

Loadbalancing i failover hostname cache.obalkyknih.cz jsou řešeny pomocí Lua skriptu využitím funkce ifurlextrup, která je schopna v pravidelných intervalech (řádově jednotek sekund v závislosti od konfigurace) kontrolovat dostupnost cíle, který je v případě obálek knih URL /api/runtime/alive. Dostupnost je kontrolována HTTP návratovým kódem i textem obsaženým v odpovědi. V případě nedostupnosti, nebo špatné odpovědi kontrolního URL je frontend server vyřazen z odpovědi jmenného serveru a je poskytnuta pouze adresa

dostupného serveru. Lua funkce `ifurlextop` navíc umožňuje definovat skupiny serverů, čímž je možné zajistit loadbalancing. Pokud jsou dostupné oba frontend servery jsou poskytovány adresy obou s náhodným rozdělením. Alternativně je možné využít namísto selektoru `random` selektor `pickclosest`, který ve spojení sází GeoIP se pokouší o odpověď IP adresou geograficky nejbližšího frontend serveru.

GeoIP báze (konkrétně MaxMind GeoLite2) se dále používá k řízení dotazů na server `cache3` nacházející se na síti Sanet. Lua script s použitím funkce `countryCode` je schopen definování pravidla pro requesty pocházející z IP adres klientů nacházejících se na území SK, které jsou tímto způsobem směřovány na server `cache3`. Kdežto IP adresy jiných klientů (tedy hlavně z území CZ) jsou směřovány rovnoměrně na servery `cache1` v infrastruktuře JVK a `cache2` v infrastruktuře MZK.

Pro obnovu TLS certifikátů se používá standardní klient `certbot` certifikační autority Let's encrypt, který za využití autentizačního pluginu `certbot-dns-powers` je schopen komunikovat se jmenným serverem PowerDNS za účelem ověření autentizačního požadavku ACME protokolu a jako výsledek vygenerovat wildcard certifikát pro celou doménu obálek knih.

KONTROLY A DALŠÍ PRÁCE

V roce 2023 pokračovaly kontroly záznamů projektu v našem kontrolním rozhraní (<https://servis.obalkyknih.cz/abuse/>). Do kontrol se opětovně zapojovali knihovníci, kteří pracovali z domova v rámci homeoffice a dobrovolníci z UISKu. Aktuálně je v systému zaregistrováno 181 uživatelů (aktivně kontrolovala v roce 2023 cca 15 z nich).

Při kontrolách záznamů dochází k doplňování chybějících identifikátorů (číslo `cnb`, `oclc`, `isbn`), doplňování popisných metadat pro zobrazování na stránkách projektu (název, autor, rok vydání), doplňování či úpravě anotace, výběru nejkvalitnější obálky (rozlišení, bez nálepek, ...). Zároveň se opravovaly chybné záznamy a propojovaly se související záznamy (dílové tituly).

V roce 2023 bylo již zkontrolováno celkem přes 598 000 záznamů (viz tabulka).

Tabulka počtu opravených záznamů dle zapojených institucí (stav k 10.1. 2024):

Přehled počtu zkontrolovaných záznamů dle zapojených knihoven	
Jihočeská vědecká knihovna v Českých Budějovicích	359 345
Městská knihovna v Praze	179 894
UISK	18 875
Knihovna Ústeckého kraje	17 129
Krajská knihovna Karlovy Vary	6 688
Moravskoslezská vědecká knihovna v Ostravě	4 592

Krajská knihovna Františka Bartoše ve Zlíně	3 202
Moravská zemská knihovna v Brně	3 139
KVK Liberec	2 566
NKP	1 202
Krajská knihovna Vysočiny	932
Knihovna města Plzně, p.o.	510
Knihovna Třinec	500
Městská knihovna Tábor	365

Statisticky vychází, že u cca třetiny záznamů dochází k nahrání kvalitnější obálky a u cca. 25% záznamů je doplněna či upravena anotace.

Knihovny takto získaly kvalitnější záznamy s většími obálkami.

Dále byly v roce 2023 prováděny úkoly související s kontrolou anotací a hodnocení, propojení na další vydání titulu a propojení vazeb e-knih.

Jednalo se o práce, které nelze provádět plně automatizovat a je nutné je provádět pomocí zaškolené obsluhy.

Výsledkem práce bylo:

- 1) napojení e-knih na klasické papírové tituly – aktuálně je dostupných více jak 3000 e-knih z produkce Městské knihovny v Praze
- 2) propojení na další vydání titulu – propojení jednotlivých vydání shodného titulu nebo navazující série titulů
- 3) opravy a schvalování anotací (stahovaných automaticky či ručně) k dokumentům – za rok 2023 přibýlo v projektu 96 tisíc anotací, které byly formálně zkontrolovány a doplněny k titulům,

Další úkoly řešené v roce 2023 mimo projekt:

- příprava exportu DNNT dokumentů pro Exlibris – projekt Univerzity Karlovi v Praze
- přidání možnosti hvězdičkové URL adresy pro nastavení práva přístupu k cache serverům
- přepracování systému stahování dat z eshopů, přidání nových 3 eshopů – bylo získáno cca 50 tisíc nových obálek a anotací
- opětovné napojení na OCR server NKP – bylo odstaveno před lety po hackerském útoku na NKP
- spolupráce s Wikimedia na stahování obrázků knihoven a zpřístupnění knihovnám

- stahování obálek a obsahů z Kramériů knihoven ČR – do dohodě s jejich administrátory
- údržba a podpora skenovacího klienta pro nahrávání dat knihovnami do projektu, ladění nové verze s podporou TWAIN
- nastavení individuálních exportů knihovnám pro doplnění obálek a obsahů vlastního fondu (Knihovna Univerzity Tomáše Baťa ve Zlíně, Krajská knihovna Františka Bartoše ve Zlíně, Jihočeská vědecká knihovna v Českých Budějovicích a časopis Týdeník EURO)
- aktualizace webových stránek projektu
- servis a rozšíření servisních stránek projektu <https://servis.obalkyknih.cz/>
- aktualizace metodických pokynů a manuálu pro knihovny a knihovní systémy
- prezentace projektu na knihovnických akcích (Seminář uživatelů ARL, setkání knihoven regionu JČ)
- emailová a telefonická podpora projektu, spolupráce s tvůrci KIS, CPK

Popis řešení a veškeré kódy aplikace jsou volně dostupné jako opensource na adrese <https://github.com/cbvk/obalkyknih/wiki> a <https://www.obalkyknih.cz>.

Ing. Jiří Nechvátal
Jihočeská vědecká knihovna v Českých Budějovicích