



VISK 3

Informační centra veřejných knihoven - ICEKNI

Obálky knih.cz - rozvoj projektu v roce 2016

Zpráva o plnění cílů projektu VISK 3

Jihočeská vědecká knihovna v Českých Budějovicích

2016

Zhodnocení projektu

Slovní zhodnocení projektu a rekapitulace případných změn projektu provedených a náležitě oznámených během roku.

Projekt ObalkyKnih.cz sdružuje různé zdroje informací o dokumentech do jedné, snadno použitelné webové služby. V současnosti pro knihovny a jiné katalogy poskytujeme:

- **Obálky knih a periodik** - ale také obálky speciálních dokumentů - map, hudebnin, CD, DVD, ... , aktuálně přes **1,42 miliónu obálek**, nárůst za rok 2016 o cca. **300 tisíc** obálek
- **Obsahy knih a periodik** - naskenované obsahy zpřístupněné v podobě PDF souborů, aktuálně přes **238 tisíc obsahů**, nárůst za rok 2016 o cca. **78 tisíc** obsahů
- **Fulltexty obsahů** - převedené obsahy na text pomocí technologie OCR a zpřístupněné pro indexaci knihovním systémem
- **Anotace** - nakladatelské, autorské i knihovnické anotace získané z různých zdrojů, aktuálně přes **300 tisíc anotací**
- **Komentáře a hodnocení** - možnost stažení dostupných komentářů a zároveň i možnost exportu komentářů z jiných systémů na obalkyknih.cz a tím zpřístupnění ostatním knihovnám, aktuálně cca. **60 tisíc hodnocení, 14 tisíc komentářů**
- **Fotografie autorit** - nově poskytujeme portrétové fotografie autorů a jiných osobností z Autoritní databáze Národní knihovny ČR (<http://aut.nkp.cz>), aktuálně přes **47 tisíc fotografií**

Služeb projektu Obalkyknih.cz využívá většina knihoven v České republice. Dále pak muzea, archivy, veřejné projekty aj.

Správce projektu Obalkyknih.cz je Jihočeská vědecká knihovna v Českých Budějovicích (JVK) společně s Moravskou zemskou knihovnou (MZK).

Přehled vlastností projektu:

- hlavní servery jsou provozovány v Jihočeské vědecké knihovně v Českých Budějovicích, záložní server je umístěn v Moravské zemské knihovně v Brně
- denně do databáze je nově nahráno nebo je upraveno průměrně 500 dokumentů
- další dokumenty se automaticky sklízí z externích zdrojů - nakladatelé, vydavatelé, webové portály, ...
- denní přírůstek dat činí 6 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

Statistiky využití služby:

V roce 2016 servery projektu odbavily více jak 1 miliardu požadavků. Viz tabulka a graf níže.

Server "backend" využívají knihovní systémy s implementovanou starší verzí API (API 1 a API 2). Servery "cache1" a "cache2" poskytují data přes aktuální API 3, dostupné již od roku 2013. Umožňuje oproti starším verzím rozšířit knihovní systémy o další prvky (obsahy, anotace, hodnocení, komentáře, texty obsahů k indexaci, citace,...), dotazování pomocí více identifikátorů,

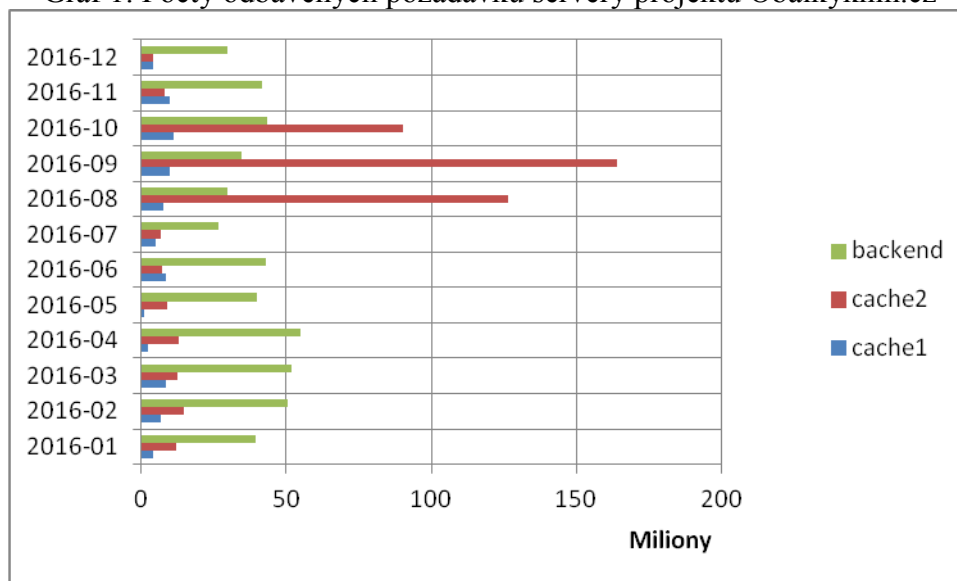
rychlejší načítání obálek, odolnost proti výpadkům, stahování fotografií autorit, API 3 pracuje podstatně efektivněji a nevyžaduje tolik dotazů klientů vůči serverům.

V měsíci srpnu - září 2016 jsme spolupracovali s projektem Centrální portál knihoven (<https://www.knihovny.cz/>). V rámci vývoje Centrálního portálu došlo k opakovanému sklizení dat z projektu Obálkyknih.cz a testovacímu zatížení serverů na 10ti násobek běžného stavu. Servery jsou schopny obsloužit i takovou zátěž.

Tabulka 1: Počty odbavených požadavků servery projektu Obálkyknih.cz

	cache1	cache2	backend
2016-01	4 035 741	12 269 331	39 666 258
2016-02	6 738 552	14 533 628	50 702 887
2016-03	8 386 012	12 370 734	51 897 812
2016-04	2 579 154	13 087 740	54 963 200
2016-05	1 271 111	9 052 701	39 866 411
2016-06	8 451 079	7 444 354	43 158 588
2016-07	5 134 933	6 703 262	26 606 110
2016-08	7 885 712	126 375 728	29 696 325
2016-09	9 733 147	164 139 437	34 526 958
2016-10	11 355 299	90 464 796	43 476 917
2016-11	9 984 521	7 958 728	41 863 335
2016-12	3 981 310	4 202 661	29 543 726
	79 536 571	468 603 100	485 968 527
			1 034 108 198

Graf 1: Počty odbavených požadavků servery projektu Obálkyknih.cz



Statistiky přispívání přes skenovacího klienta za období leden - prosinec 2016:

Počet odeslaných dokumentů **99 748**
 Počet uložených obálek (COVER) **86 182**
 Počet uložených stran obsahu **155 369**
 Počet uložených obrázků autorit **653**

Počty odeslaných stran a titulů přes skenovacího klienta dle jednotlivých knihoven (rok 2016):

STRAN	TITULŮ	SIGLA	NÁZEV
69 904	33 238	CBA001	Jihočeská vědecká knihovna v Českých Budějovicích

27 505	4 825	BOD001	FF MU
18 688	8 831	ABA001	Národní knihovna ČR
16 070	6 776	ABA004	Slovanská knihovna
15 416	7 575	OLA001	Vědecká knihovna Olomouc
13 204	7 936	BOA001	Moravská zemská knihovna
8 512	2 935	ABB019	Knihovna Sociologického ústavu AV ČR, v.v.i.
7 521	2 516	ABA008	Národní lékařská knihovna
7 333	2 273	CBD005	Teologická fakulta JCU
6 924	1 754	BOD010	Masarykova univerzita - Právnická fakulta
4 335	1 040	BOD004	Ústřední knihovna Přírodovědecké fakulty MU
4 224	4 224	ABD001	Knihovna Ústavu Dálného východu Filozofické fakulty UK
4 149	1 464	BOE020	Knihovna Ústavního soudu
3 538	1 132	BOD031	Masarykova univerzita, Fakulta sociálních studií, Ústřední knih.
3 073	775	ZLD002	Univerzita Tomáše Bati ve Zlíně
2 994	1 534	PNA001	Studijní a vědecká knihovna Plzeňského kraje
2 637	684	ABA006	Vysoká škola ekonomická v Praze
2 607	561	ABA013	Národní technická knihovna
2 560	733	ABA007	Knihovna Akademie věd
2 315	1 265	LIA001	Krajská vědecká knihovna v Liberci
2 311	637	LID001	Technická univerzita v Liberci, Univerzitní knihovna
2 178	465	ABB503	Knihovna ÚACH AVČR, v.v.i.
1 590	886	HBG001	Krajská knihovna Vysočiny
1 464	430	ABD103	SVI FSV
1 320	699	KLK001	Středočeská vědecká knihovna v Kladně
1 248	357	ABD100	ÚK ČVUT
914	677	PAG001	Krajská knihovna v Pardubicích
833	465	ULG001	Severočeská vědecká knihovna v Ústí nad Labem
472	472	TAG001	Městská knihovna Tábor
452	368	PTG001	Městská knihovna Prachatice
414	136	BOD006	Informační centrum, ústřední knih. Mendelovy univerzity v Brně
340	201	ULD001	Ústřední knihovna UJEP
339	158	JID501	Knihovna Univerzitního centra Telč Masarykovy univerzity
306	202	ZLG001	Krajská knihovna Františka Bartoše ve Zlíně
302	255	OSA001	MSVK v Ostravě
266	75	BOD018	Masarykova univerzita - Fakulta informatiky
131	51	ABE367	Archiv hlavního města Prahy
63	49	SOG504	Městská knihovna Chodov
60	20	ABB001	Knihovna Archeologického ústavu AV ČR, Praha, v.v.i.
47	12	ABG312	JABOK - Vyšší odborná škola sociálně pedagogická a teologická
31	31	BOD033	Univerzitní knihovna pro studenty se specifickými nároky MU
12	7	KVG001	Krajská knihovna Karlovy Vary
12	12	ABA100	Všenorská knihovna a informační centrum Berounka
10	5	PBG001	Knihovna Jana Drdy Příbram
3	2	ABA009	Knihovna Antonína Švehly

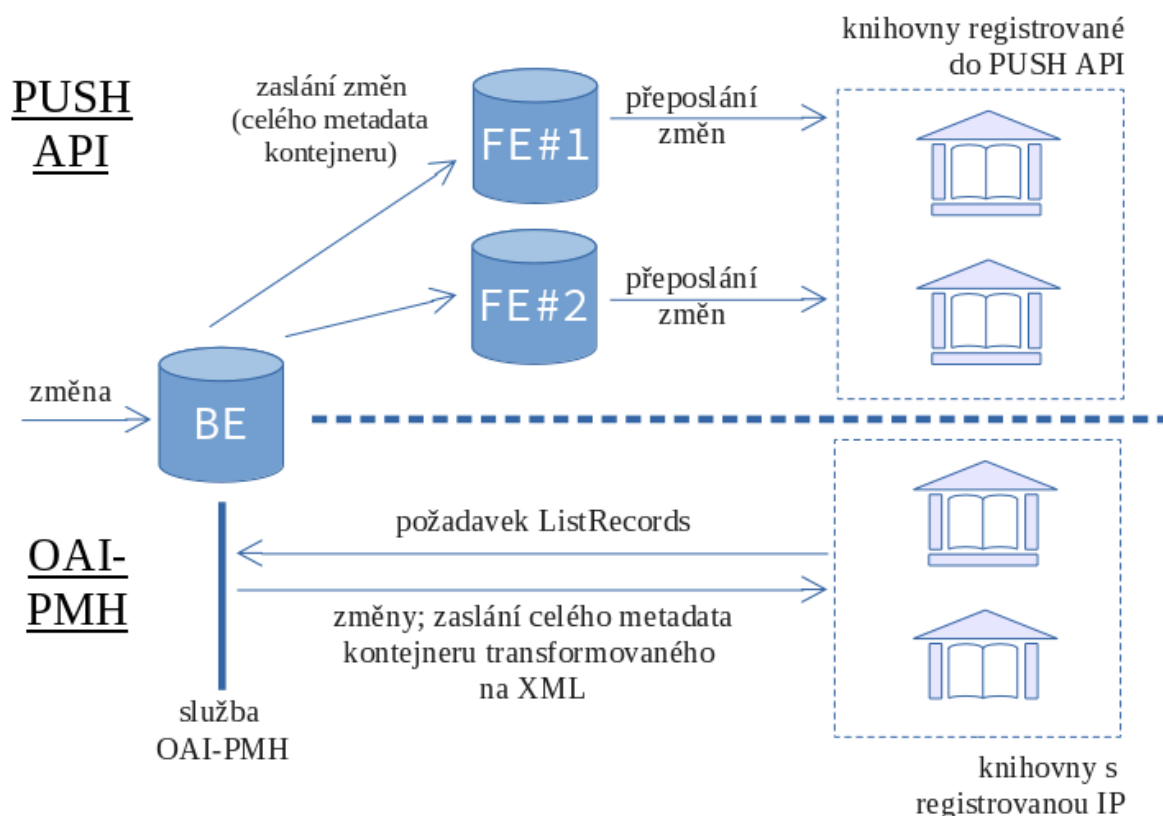
Úkoly řešené v rámci projektu v roce 2016:

Push API

Cílem úkolu bylo zrychlit načítání (aktualizaci) dostupných informací obohacujících knižní katalogy. Jedná se o podporu urychlení přímo pro implementace knižních katalogů. Push API bude sloužit k proaktivnímu informování ze strany infrastruktury projektu Obálky knih.cz (front-end serverů (FE), případně back-end (BE) serveru) směrem k serverům knižních katalogů, které budou okamžitě na tuto informaci reagovat a budou schopny informace ve svých bázích aktualizovat. Kladný dopad se projeví i pro vyhledávání ve fulltextech obsahů, poskytovaných projektem

OAI-PMH představuje opačný postup o informování o změnách, kde bude o zaslání změn žádat samotný knihovní systém. Databázový model byl doplněn o timestamp, který se mění s každou změnou záznamu (tabulky books + auth). Na BE bude spuštěna služba OAI serveru, která bude schopna vyhledávat v hodnotách timestampů těchto 2 tabulek a bude podporovat funkci stránkování pomocí „resumptionToken“. Metadatový kontejner, v odpovědi OAI-PMH serveru bude transformován na formát XML.

Architektura řešení (viz. <https://github.com/cbvk/obalkyknih/wiki/>):



Automatizované sklizení obsahů a obálek ze systému Kramerius

V rámci úkolu bylo řešeno obohacování datového modelu obalky knih.cz o sklizení obálek, TOC digitalizovaných objektů a URL na fulltext titulu v systému Kramerius navázaný na siglu knihovny (pro možný výběr zdroje). Úloha umožní vytěžování dokumentů pořizovaných v rámci projektu NDK a dalších projektů (např. krajské digitalizace). Technické řešení je realizováno cestou OAI-PMH. Jsou prohledávány modely Krameria: monograph, periodicalitem, soundrecording, archive, graphic, sheetmusic, manuscript.

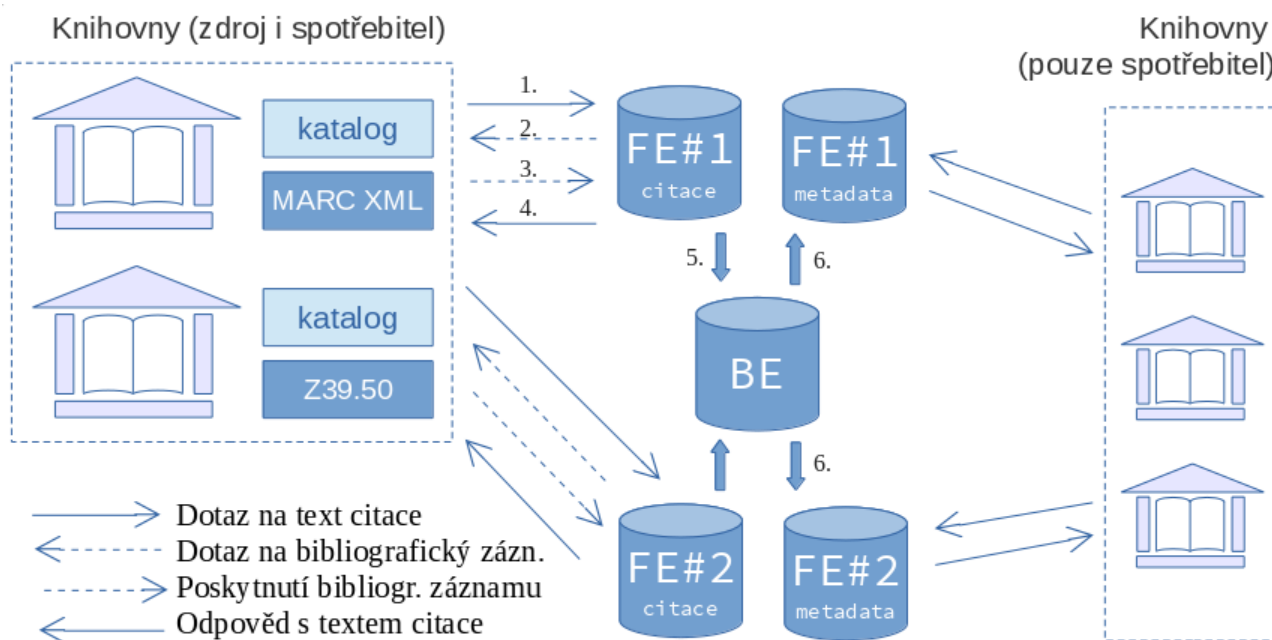
Citace

Implementace citací podle normy ČSN ISO 690 je řešeno dvěma způsoby:

1) Sklizení bibliografických záznamů ze Souborného katalogu ČR a následným vytvořením citace. Uloženy budou výsledky transformace - výsledné citace, které jsou poskytovány pomocí API obálky knih.cz. Tyto citace jsou určeny pro menší knihovny s minimální možností úpravy svého katalogu. Postačuje na požadované místo knihovných stránek vložit IFRAME. Zdrojem dat pro generování textu citace jsou jiné katalogy, zejména NKP a knihovny zapojené do projektu jako přispěvatelé. Nevýhodou je, že text citace nebude generován pro všechny typy dokumentů a pro všechny záznamy.

2) Umožnění citování i jiných typů dokumentů, jako např. články, mapy, apod., které server obálky knih.cz generuje dle pravidel na základě zaslaných dat knihovnou.

Architektura řešení (viz. <https://github.com/cbvk/obalkyknih/wiki/>):



Sklizení dalších zdrojů pro autoritní bázi

Autoritní báze projektu byla obohacena o další zdroje obrázků a informací, které knihovny používají pro vylepšení svých katalogů. Byla sklizena databáze Wikimedia a testovány další zdroje. Náhledy jsou dostupné ve vícerych rozlišeních. Stávající stav při zobrazování fotografií a informací o autorech v knihovních systémech je řešen individuálně každou knihovnou. Sklizením těchto individuálních zdrojů se data zpřístupní jakékoliv knihovně a jejím čtenářům.

Aktuálně databáze obsahuje přes 47 tisíc fotografií autorit. Díky automatizovanému sklizení a implementaci funkce odesílání autorit do skenovacího klienta přibývají měsíčně stovky nových fotografií.

Vylepšení skenovacího klienta

Cílem bylo upravit skenovacího klienta - program sloužící pro přispívání do projektu - pro zrychlení a zefektivnění práce s klientem při odesílání obálek a obsahů na server. Podněty pochází z reálného provozu od uživatelů pravidelně používajících aplikaci skenovacího klienta (za rok 2016 knihovny přes klienta přispěli přes 155 tisíc stran).

Konkrétní řešené úkoly:

- podpora skenování dvojstran, resp. menších formátů monografií
- podpora skenování na pozadí (uživatel nemusí čekat na kompletní odeslání dat na server a pokračuje ve skenování dalšího titulu)

- možnost doplnit sken do již existující skupiny skenů obsahu
- podpora více typů skenerů a větší kompatibilita se staršími skenery
- změna workflow, pro podporu vyšší výtěžnosti skenování OCR obsahů děl
- možnost ukládat naskenované obrázky do vlastní složky
- podpora režimu proxy
- skenování a odesílání hlavní autority titulu na server

Další úkoly řešené v roce 2016 mimo projekt:

- upgrade klíčových softwarových systému projektu - Node.js, Mongo.DB pro zvýšení výkonu a stability běhu služby.
- implementace větších náhledů o velikosti 510px
- vyšší rozlišení PDF souborů s obsahy pro kvalitnější zobrazení
- import cca. 55 tisíc hodnocení z katalogu Městské knihovny v Praze
- import cca. 9800 obálek zahraničních dokumentů z fondu Národní technické knihovny
- nastavení sklizení portálu zbozi.cz
- spolupráce s vývojáři knihovnických systémů na implementaci API projektu
- technická podpora pro zapojené knihovny

V roce 2016 byl projekt prezentován na knihovnických akcích:

Konference Knihovny současnosti 2016

Setkání pracovní skupiny pro seriály NK Praha

Seminář účastníků SK ČR, konference Archivy, knihovny, muzea v digitálním světě 2016

Setkání uživatelů knihovního systému ARL

V roce 2016 Jihočeská vědecká knihovna v Českých Budějovicích spolu s Moravskou zemskou knihovnou v Brně získala za rozvoj projektu ObalkyKnih.cz cenu **KNIHOVNA ROKU 2016**.

Odkazy na projekt, technická dokumentace:

<http://www.obalkyknih.cz>, <https://github.com/cbvk/obalkyknih/wiki>

V Českých Budějovicích 7.1.2017

Ing. Jiří Nechvátal
Jihočeská vědecká knihovna v Českých Budějovicích