

Příloha č. 4

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. API služby využívá aktuálně přes 200 registrovaných knihoven a stovky neregistrovaných uživatelů.

V současnosti naše databáze nabízí:

- **Obálky knih a periodik** - uložit obálku je však možno i u speciálních dokumentů - map, hudebnin, CD, DVD, ...
- **Obsahy knih a periodik** - naskenované obsahy zpřístupněné v podobě PDF souborů
- **Fulltexty obsahů** - převedené obsahy na text pomocí technologie OCR a zpřístupněné pro indexaci knihovním systémem
- **Anotace** - autorské, nakladatelské, ale i uživatelské popisy dokumentů pro snadnější výběr čtenáři
- **Komentáře a hodnocení** - možnost stažení dostupných komentářů a zároveň i možnost exportu komentářů z jiných systémů na obalkyknih.cz a tím zpřístupnění ostatním knihovnám

Statistická data projektu:

- databáze aktuálně obsahuje **cca. 1 200 000 obálek a 170 000 obsahů** českých a zahraničních publikací
- hlavní servery jsou provozovány v Jihočeské vědecké knihovně v Českých Budějovicích, záložní server je umístěn v Moravské zemské knihovně v Brně (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 55 miliónů požadavků, cca. 2 milióny denně, průměrně 20 dotazů za vteřinu, ve špičkách (9-15:00) odbavují servery 40-80 požadavků za vteřinu
- denně do databáze je nově nahráno nebo je upraveno průměrně 250 dokumentů
- denní přírůstek dat činí 3 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 služby

V roce 2015 bylo v rámci projektu VISK3 řešeno několik nových funkcí:

Rozšíření projektu obálky knih o propojení s databází autorit:

smyslem úlohy bylo poskytnout knihovním katalogům obrazové a jiné materiály napojené na ID autority. Například fotografie osob, GPS souřadnice míst, www odkazy,

Na serveru obalkyknih.cz byla vytvořena autoritní báze dat, která čerpá data z databáze autorit NK Praha přes OAI-PMH. Autoritní báze je pravidelně aktualizována. Knižní katalogy mají možnost stahovat data pomocí nově vytvořeného API autoritní báze a obohatit své katalogy. Přes webové stránky projektu lze autority vkládat a vyhledávat. Přihlášený uživatel může nahrávat nové foto i odkazy odkazující na další info. o autoritě. Zároveň se odkazy a informace zobrazují na webu obálek knih.

Rozšíření projektu obálky knih o podporu obálek zvukových dokumentů a hudebnin:

na základě dohody se spolupracujícími knihovnami byl dohodnut identifikátor pro jedinečnou identifikaci zvukových dokumentů a hudebnin - urn:okcz:typ_dokumentu:vydavatelske_cislo.

Pro typ dokumentu byl vytvořen číselník (MC, CD, LP, HUD).

Součástí úlohy byl i import cca. 15 000 obálek zvukových dokumentů získaných od vydavatelství Supraphon. Import bude obohacovat záznamy nejen o obálky, ale i o seznamy skladeb/stop, těles, případně jiných typů informací, které připravené XML obsahuje. Vyhledání záznamů probíhá pomocí API a parametru nbn. Obohacení doplňkovými informacemi je v podobě fulltextu v odpovědi API. Knižní systémy jsou touto cestou schopni vyhledávat tituly pomocí názvů skladeb/stop, těles, apod. s využitím dnes už pro knižní systémy známé formy.

Automatizované sklizení obsahů a obálek ze systému Kramerius:

v důsledku krácení rozpočtu byl tento úkol řešen jen jako metodický postup stahování záznamů z Kraméria do báze obálekknih.cz. V roce 2016 bylo požádáno o pokračování projektu a možnost dokončení úlohy automatizovaného sklizení.

Deduplikace záznamů podle identifikátorů:

úkol řešil historický problém s ukládáním separátních záznamů stejného dokumentu pod různými identifikátory - ISBN, číslo ČNB, OCLC. Vzniklo v důsledku vkládání přes různé zdroje - webové rozhraní obalkyknih.cz, sklízecí proces nad daty poskytovanými vydavateli a knihovnami tzv. crawler a harvestery, kdy k záznamu nebyly přiřazeny všechny dostupné identifikátory. Problém nastával hlavně s obsahy, komentáři, fulltexty ... , kdy je nutné jejich uložení pod správný - jedinečný - dokument.

Pro deduplikaci byl využit jako zdroj dat báze Národní knihovny Praha a jako deduplikační klíč jsou brány identifikátory titulu. Tituly, které budou spadat pod proces deduplikace budou slučovány do jediného záznamu tak aby informační hodnota byla zachována, nebo zvýšena.

Proces deduplikace byl spuštěn nad celou databází projektu obalkyknih.cz a dále byl vytvořen automatizovaný proces, který bude deduplikaci pravidelně provádět.

Převzetí komentářů z jiných systémů, nebo sklizení dalších zdrojů dat :

Bylo upraveno importní XML rozhraní projektu Obálky knih.cz, tzv. crawler tak, aby bylo možné pravidelně sklízet níže uvedené zdroje dat. Upravený proces je možné dále používat pro další partnerské servery/zdroje dat, které se později zapojí do projektu a poskytnou své XML importní soubory knižních titulů.

Sklízené zdroje dat:

- portál Můj antikvariát - <http://muj-antikvariat.cz/> - bylo staženo cca. 46 tisíc nových obálek
- portál Srovnáme.cz - <http://www.srovname.cz/> - bylo staženo cca. 103 tisíc obálek a anotací
- web Svaz českých knihkupců a nakladatelů: SČKN - <http://www.sckn.cz/> - opraven nefunkční harvester, bylo staženo cca 40 tisíc nových obálek a anotací
- portál Ereading.cz - <http://www.ereading.cz/> - bylo staženo cca. 10 tisíc titulů a anotací
- Národní technická knihovna - <https://www.techlib.cz/> - importováno cca. 1700 obálek zahraničních periodik
- Městská knihovna v Praze - <https://www.mlp.cz/> - importováno cca. 100 tisíc anotací dokumentů, cca. 56 000 hodnocení
- Souborný katalog ČR - <http://www.caslin.cz/> - importováno cca. 50 tisíc nových anotací
- Souborný katalog uživatelů Lania - <http://www.skat.cz/> - importováno cca. 14 tisíc komentářů a cca. 7 tisíc nových obálek dokumentů

Bylo jednáno s dalšími portály - [zboží.cz](http://zbozi.cz), heurka.cz, [databázeknih.cz](http://databazeknih.cz), ... - o možné spolupráci. Bohužel z jejich strany nebyl zájem o spolupráci případně se stále vedou jednání ([zboží.cz](http://zbozi.cz)).

Mimo projekt byl proveden upgrade klíčových softwarových systému projektu - Node.js, Mongo.DB a nasazení TypeScript - pro zvýšení výkonu a stability běhu služby.

V roce 2015 byl projekt prezentován na 7 celostátních setkáních odborné knihovnické obce (setkání uživatelů knihovního systému Aleph, konference Knihovny současnosti 2015, setkání pracovní skupiny pro seriály NK Praha, seminář účastníků SK ČR, konference Archivy, knihovny, muzea v digitálním světě 2015, setkání uživatelů knihovního systému ARL, přednáška o zapojení knihoven do projektu Obálkyknih.cz na MUNI Brno) a napsány články pro knihovnická periodika - Biblio, Čtenář.

V příštích letech bychom rádi do projektu zapracovali poskytování citací dle normy ISO 690, případně jiných norem a také dokončili automatizované sklízení obsahů a obálek ze systému Kramerius včetně obohacování datového modelu obalkyknih.cz o URL na fulltext titulu v systému Kramerius navázaný na siglu knihovny (pro možný výběr zdroje). Funkce umožní vytěžování dokumentu pořízených v rámci projektu NDK a projektů krajských digitalizací.

Odkazy na projekt, technická dokumentace:

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

07.01.2016