

Zpracování a analýza metadat

Jana Kurfürstová
Moravská zemská knihovna v Brně

OpenRefine

švýcarský nůž na práci se strukturovanými daty

- Ke stažení: <https://openrefine.org/download.html>
- Návody: <https://docs.openrefine.org/manual/grelfunctions>
- Stáhněte si soubor, se kterým budeme pracovat:
<https://github.com/JanaKurfurstova/vyukaKISK/blob/main/cvicnaData.mrk> nebo ve studijních materiálech
- Předpokládaná reálná situace:
 - Z knihovního systému vyexportujete MRC
 - V [MarcEditu](#) zkonvertujete MRC do MRK
 - S MRKem uděláte potřebné manipulace v OpenRefinu
 - Výsledný MRK zkonvertujete MarcEditem opět do MRC
 - MRC nahrajete do knihovního systému
 - Případně se přizpůsobíte situaci v jiném prostředí...

OpenRefine - možnosti

- **MarcEdit:** ok na kontrolu syntaxe, validaci ISBN, hromadné úpravy bez návaznosti na zbytek záznamu
- **OpenRefine:** umožní zkoumat širší souvislosti:
 - Pole X je syntakticky v pořádku, ale je v něm nesmysl.
 - Pole X je v rozporu s tím, co říká pole Y.
 - Záznamy s nějakým společným jmenovatelem mají chybu v poli X.
 - Vytvoření nového pole namapováním údajů z jiného souboru.
 - OpenRefinem můžete analyzovat a čistit i jiná data (exporty z Google Analytics, systémové logy, datasety vzniklé při výzkumu atd.)

Úskalí hromadných úprav

OPATRNĚ s hromadnými úpravami!

- To, že něco jde, neznamenaá, že je to dobrý nápad.
- Nevidíte pramen popisu (nemáte knihu v ruce). Co když je to tam opravdu špatně (a podle RDA to musíte opsat i s chybou)? Co když si nakladatel říká pokaždé trochu jinak?
- Chybná ISBN a další identifikátory nejsou vzácnost (takže mapování podle nich je risk).
- Authority jsou skvělá věc, ale automatické doplňování je hazard.

Úskalí hromadných úprav

Co je bezpečné?

- Nalezení podezřelých či chybějících údajů a oprava v katalogizaci s knihou v ruce.
- Oprava syntaktických chyb: posuvy a chyby na konkrétních znakových pozicích, interpunkce, špatně použitá podpole.
- Doplnování opravdu jistých RDA polí.
- Sjednocování či doplňování volně tvořených klíčových slov.

OpenRefine - příklady

- Různé způsoby otevření MRK, kontrola v řádkové podobě
- Překlopit bit.ly/3RyE2UP, seřadit bit.ly/3Rwgdhu sloupce
- Nalezení záznamu s opakovaným polem 245.
- Nalezení a oprava s posunutým jazykem.
- Nalezení autorů bez autority. Nalezení neseďících roků.
- Doplnění interpunkce do polí 26X.
- Nalezení hudebnin zapsaných jako textový dokument.
- Nalezení knih vydaných po r. 1989 bez ISBN.
- Nalezení RDA záznamů bez RDA polí. Doplnění chybějících 336-338 dle typu dokumentu.
- Cluster & merge volně tvořených klíčových slov.
- Zpět do MRK - <https://bit.ly/48ggaMp>

Prostor na dotazy

DĚKUJI ZA POZORNOST

Jana Kurfürstová
kurfurstova@mzk.cz