



Aneb jak se dělá discovery systém

Jana Kurfürstová

Katalog beletrie a odborné literatury

Vypůjčte si knihy v knihovně nebo najděte informace online

Zadejte hledaný dotaz, např. název knihy

 **HLEDAT**



[Pokročilé vyhledávání](#)



ČESKÉ KNIHOVNY

Vyhledávání v českých knihovnách a databázích



ZAHRANIČNÍ ZDROJE

Odborné a vědecké licencované publikace ze zahraničí



E-KNIHY

Volně dostupné elektronické knihy ke stažení

Online knihy do karantény



Lázně

Martin Reiner, 1964-



Bílá velryba

Herman Melville, 1819-1891



Román pro muže

Michal Viewegh, 1962-

E-knihy: Detektivky



Lvíče



Hlavní přelíčení



Ruce vzhůru!

Informujeme

Rozcestník e-zdrojů

Připravili jsme rozcestník e-zdrojů podle věkových skupin. Podívejte se, co mohou online využívat děti a mládež, studenti VŠ, dospělí nebo senioři.

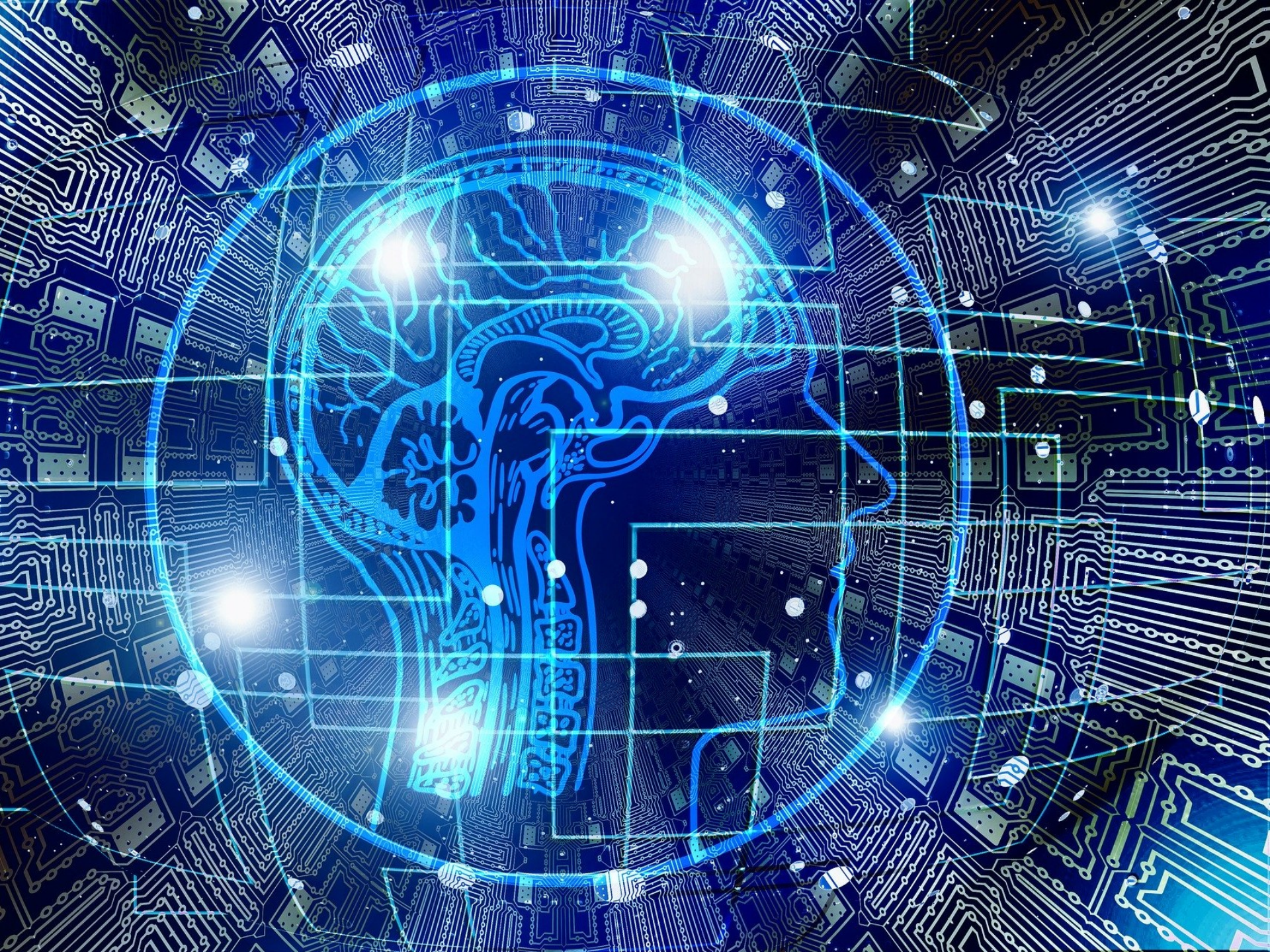


O čem to bude?

- Jak portál funguje, jak je technicky řešený
- Co znamená zapojení pro knihovnu
- Tj. **PROČ JE PORTÁL TAKOVÝ JAKÝ JE?**

A pokud možno taky
demonstrace toho,

- **PROČ JE KNIHOVNICTVÍ ATRAKTIVNÍ OBOR Z
HLEDISKA DATOVÉHO A INFORMAČNÍHO
MANAGEMENTU?**



O čem to nebude?

- **Co všechno portál umí**
- **Představení zapojených zdrojů**
- **Jak se liší portál od Souborného katalogu**
- **Jak se portál používá, tipy a triky**
- **Proč je portál dobrá věc (filozofie centralizovaných služeb atd.)**
- **Tj. ne CO JE PORTÁL VLASTNĚ ZAČ
A K ČEMU JE DOBRÝ?**

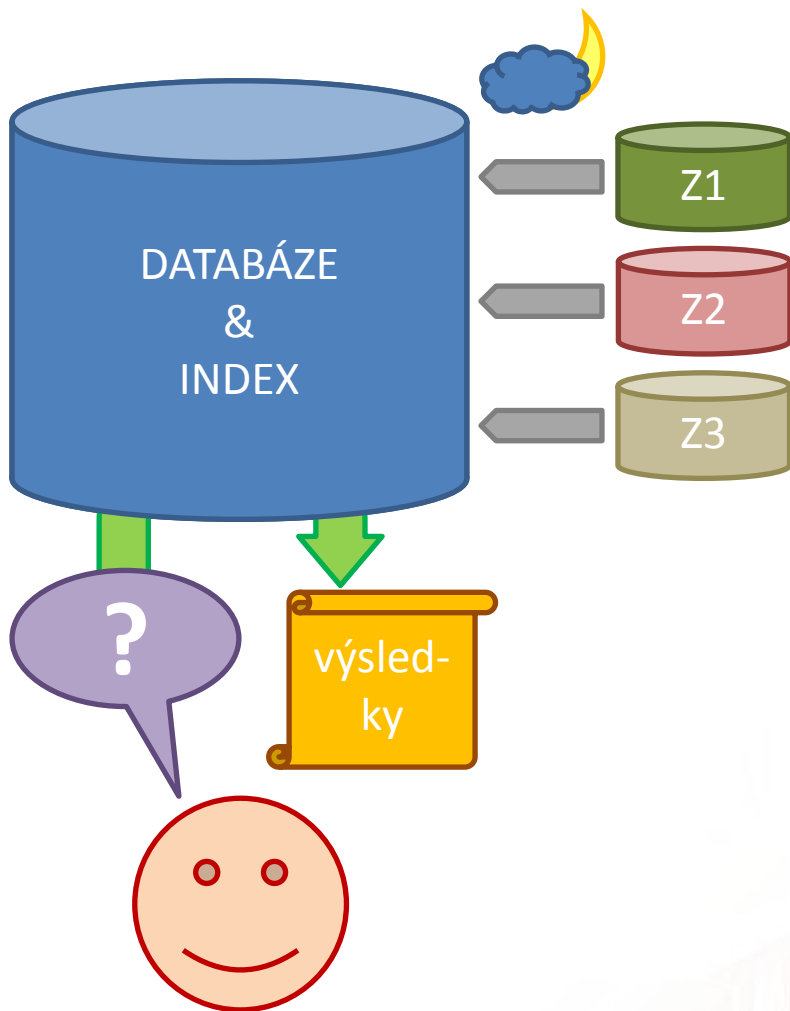
VELMI stručné shrnutí funkcí portálu

- Prohledávání záznamů ze [zapojených knihoven](#), bibliografií a dalších zdrojů (+ EDS)
 - Pokročilé vyhledávání, fasety, našeptávač, [řezy](#)...
 - Co nejpřehlednější prezentace výsledků vyhledávání (deduplikace, relevance)
- Zobrazení informací o dokumentech
 - Info z metadat, aktuální dostupnost, citace...
- Funkce pro přihlášené uživatele
 - Čtenářské konto, požadavky, propojování identit...
- Další funkce:
 - Seznamy oblíbených, uložená vyhledávání, inspirační seznamy, adresář knihoven...
- Služby knihovnám:
 - Z39.50, federativní vyhledávání pro MVS, linkserver

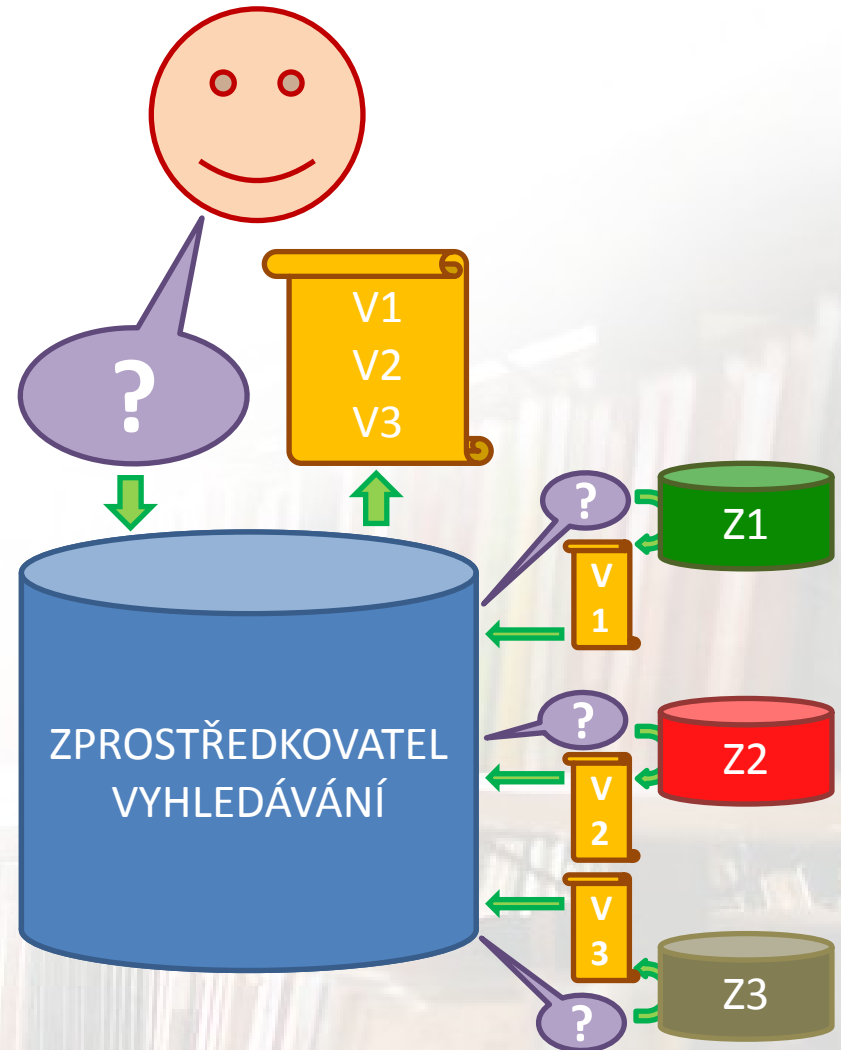
Discovery systémy

- Trend současného knihovnictví.
- Komplexní nástroje, které vycházejí z principů intuitivně použitelných internetových vyhledávačů a dalších běžně používaných webových služeb.
- Specifika:
 - Prohledávání **společného indexu** budovaného z průběžně sklízených a deduplikovaných **záznamů z vícera informačních zdrojů**.

Discovery systém



Federativní vyhledávač



Discovery systémy

- Trend současného knihovnictví.
- **Komplexní nástroje**, které vycházejí z **principů intuitivně použitelných internetových vyhledávačů** a dalších běžně používaných webových služeb.
- **Specifika:**
 - Prohledávání **společného indexu** budovaného z průběžně sklízených a deduplikovaných **záznamů z vícera informačních zdrojů**.
 - Řazení výsledků vyhledávání dle relevance a minimalizace případů nulového výsledku.
 - Obohacování záznamů pomocí tezaurů, obálek, obsahů a odkazů na tematicky návazné záznamy.
 - **Odkazování na plné texty, kdykoliv je to možné.**
 - Ukládání záznamů i výsledků vyhledávání, exporty záznamů pro citační manažery, možnost vložit hodnocení nebo komentář.

Záznamy v knihovním systému

- Bibliografické záznamy dle mezinárodních standardů
 - Metadatový formát [MARC 21](#)
 - Katalogizační pravidla **RDA** (dříve AACR2)
 - Syntaxe ISBD (*pro Knihovny.cz irelevantní*)

LEADER	03034nam a2200613 i 4500
001	nkc20172885048
003	CZ PrNK
005	20170427120736.0
007	ta
008	170310s2017 xr ac f p 001 0 cze
015	a cnb002885048
020	a 978-80-88123-10-1 q (Flow ; q brožováno)
035	a (OCoLC)987024055
040	a BOA001 b cze d ABA001 e rda
041	1 a cze h eng
072	7 a 316.77 x Informační věda 2 Konspekt 9 12
072	7 a 37.016 x Učební osnovy. Vyučovací předměty. Učebnice 2 Konspekt 9 22
080	a 316.77-027.21 2 MRF
080	a (048.8:082) 2 MRF
080	a (07) 2 MRF
100	1 a Bawden, David, d 1952- 7 js20080511002 4 aut
240	1 0 a Introduction to information science. Česky
245	1 0 a Úvod do informační vědy / c David Bawden, Lyn Robinson ; z anglického originálu Introduction to information science přeložili Michal Lorenz, Karel Mikulášek, Dana Věvodová
250	a 1. vydání
264	1 a Doubravnik : b Flow, c 2017
300	a 451 stran : b ilustrace, portréty : c 24 cm
336	a text b txt 2 rdacontent
337	a bez média b n 2 rdamedia
338	a svazek b nc 2 rdacarrier
500	a "Kniha vychází ve spolupráci s Kabinetem informačních studií a knihovnictví FF MU a Městskou knihovnou Polička"--Rub titulní stránky
500	a 1000 výtisků
504	a Obsahuje bibliografie a rejstřík
650	0 7 a informační věda 7 ph121108 2 czenas
650	0 9 a information science 2 eczenas
655	7 a kolektivní monografie 7 fd501537 2 czenas
655	7 a učební texty 7 fd133599 2 czenas
655	9 a collective monographs 2 eczenas
655	9 a educational texts 2 eczenas
700	1 a Robinson, Lyn 7 js20110609003 4 aut
700	1 a Lorenz, Michal, d 1978- 7 js20090717002 4 trl

Záznam ve formátu MARC 21

- Pole
 - Indikátory
 - Podpole
- Údaje v přirozeném jazyce
- Kódované údaje
- Pro vazby na autority a kontrolované slovníky

Záznamy v knihovním systému

- Knihovníci nemusí tvořit záznamy ručně
 - Záznamy lze **přebírat** ze SK ČR, NK ČR či jiné knihovny
 - Přebírání vhodné vždy, když se nekatalogizuje nějaký lokální unikát
 - Ruční tvorbou záznamů se do metadat vnáší chyby
 - **KREATIVITA PŘI KATALOGIZACI ŠKODÍ!!!**
- Údaje o knihovních jednotkách (exemplářích) vedeny zvlášť (na portálu řešeno přidáním zvláštním MARCovým polem 996)

Jak dostat záznamy do portálu?



Jak dostat záznamy do portálu?

- Protokol [OAI-PMH](#)

- Velmi jednoduchý (6 instrukcí)
- **HTTP** (GET, POST), **XML** (MarcXML)
- Každý záznam má **hlavičku** (ID, datum poslední změny, status, příslušnost do setů) a **tělo**
- Knihovní systém má **OAI data provider**
 - Vystavuje záznamy v setu v požadované podobě
- Portál má **OAI harvester**
 - Každý den sklízí nové a aktualizované záznamy
 - Při zapojování knihovny a cca 2x do roka sklízí všechno znovu
- **Někdy to vážne:** data provider neumí preferované instrukce, neaktualizuje datestampy, negeneruje do záznamu info o jednotkách, korektně nemaže záznamy

Zpracování záznamů v databázi (PostgreSQL)

- **Sklizené záznamy jdou do databáze, kde:**
 - **Uloží se celé, jak jsou** (s případným ošetřením výjimek)
 - **Vygenerují se z nich deduplikační klíče**
 - **Provede se deduplikace**, tj. na základě porovnávání klíčů se vytvoří hierarchie sloučených a lokálních záznamů
 - **DEDUPLIKUJEME NA ÚROVNI VYDÁNÍ** (provedení)



Deduplikační klíče

- **Tvořeny z údajů, jejichž kombinace identifikuje dokument:**
 - **Typ dokumentu** (nezbytné znát pro další zpracování, ale velmi složitě rozpoznatelné)
 - **ISBN, ISSN, ISMN, čČNB, OCN (OCLC), EAN, UUID, nakladatelské číslo** (nebývají ve starších záznamech, v ISBN bývají chyby)
 - **Autor, názvové údaje, rok vydání, vydavatel, jazyk, edice, počet stran, zdrojový dokument** (u článků)
 - **Údaje o sloučení v Souborném katalogu, cluster NKP/MZK/VKOL** (nejvyšší možná úroveň sdílené katalogizace)



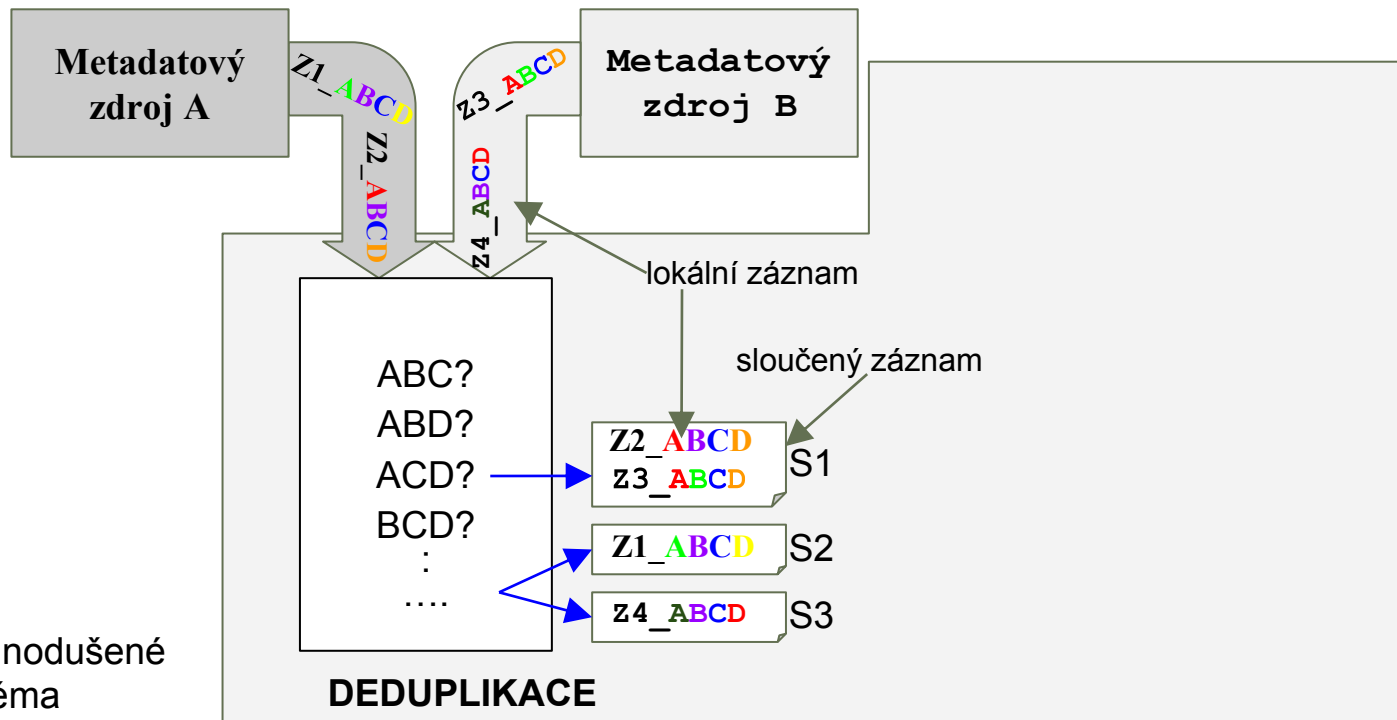
J. ENDELMANN AG

J. ENDELMANN AG

Deduplikační kroky

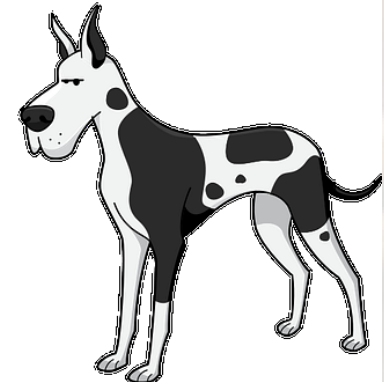
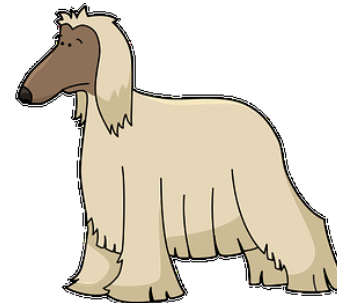
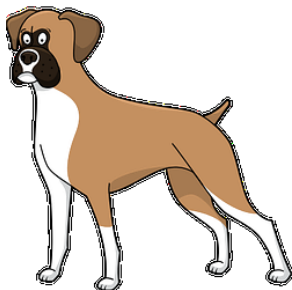
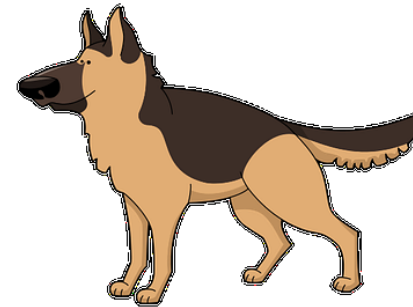
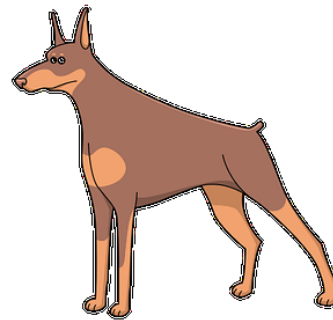
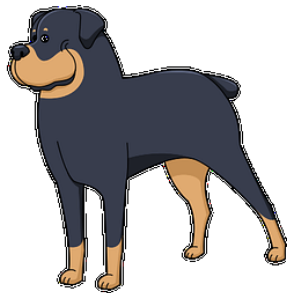
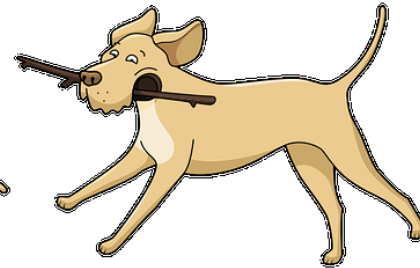
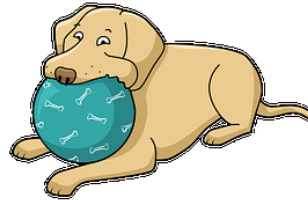
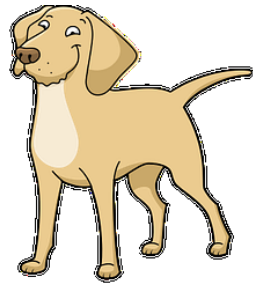
- **Proces porovnávání záznamů pomocí různých kombinací klíčů**
- **V případě shody (výjimečně i podobnosti) klíčů v kombinaci usoudíme, že záznam popisuje stejný dokument**
- **Výstupem deduplikace jsou sloučené záznamy, které propojují lokální záznamy popisující stejný dokument (ve sloučeném záznamu může být i jen jeden záznam, když zůstane sám)**

Deduplikační kroky



Zjednodušené
schéma

Problémy s deduplikací





Labyrint pohybu

Renata Červenková, 1950- 
2018

 Knihy

Nalezeno v 1 instituci

Jihočeská vědecká knihovna v
Českých Budějovicích



Labyrint pohybu

Renata Červenková, 1950- 
2018

Knížní rozhovor s renomovaným fyzioterapeutem prof. Kolářem. „Hýbeme se buď málo, nebo špatně a pohyb se vytrácí nejen ze života, ale i z medicíny,“ tvrdí profesor Pavel Kolář, renomovaný...

Nalezeno v plném textu:

“...Renata Červenková / Pavel Kolář **LABYRINT POHYBU** i I Vyšehrad ROZHOVORY, / g'a ;ť í **LABYRINT ...**”

 Knihy


Nalezeno v 44 institucích

Studijní a vědecká knihovna
Plzeňského kraje

Vědecká knihovna v Olomouci
[Zobrazit další instituce](#) ▾



Labyrint pohybu : rozhovory

Pavel Kolář, 1963- 
2018

Knížní rozhovor s renomovaným fyzioterapeutem prof. Kolářem. „Hýbeme se buď málo, nebo špatně a pohyb se vytrácí nejen ze života, ale i z medicíny,“ tvrdí profesor Pavel Kolář, renomovaný...

 Knihy

Nalezeno v 1 instituci

Městská knihovna Písek

Problémy s deduplikací

- Každý si to napsal po svém, někdo s překlepy, někdo schválně proti pravidlům (protože prý kvůli čtenářům), něco je z retrokonverze, něco je z dob, kdy na správnost knihovna moc nehleděla...
- Klíče děláme bez diakritiky, malými písmeny, pouze alfanumerické znaky, arabská čísla v titulech převádíme na římská (kvůli sjednocení, opačným směrem by to nešlo), ale ani to nestačí, když se někomu např. nechtělo psát celý název, protože byl moc dlouhý.
- **Nejčastější problémy:**
 - Kusé záznamy periodik, knižní série a vícedílné monografie, komiksy a zákony, rozkopírovaná ISBN, „tvůrčí“ nakládání s názvovými údaji, AV média, nerespektování typů dokumentů (svázaná periodika, mapy...)
- **Naše nesplnitelné přání:**

Nejlepší by bylo, kdyby knihovna zahodila, všechno, co už existuje v Souborném katalogu, a stáhla to z něho.

Indexace

- Databáze je dobrá na manipulaci se záznamy, ale při obrovském množství záznamů není vhodná jako vyhledávač
- K obstarání velkého množství dotazů s potřebou rychlé odezvy slouží **index**

Databáze

ID	Autor	Titul	Rok
01	Čapek, Karel	R.U.R.	1990
02	Čechov, A. P.	Tři sestry	1928
03	Kafka, Franz	Proces	1990
04	Čapek, Karel	Bílá nemoc	1965

Index

Autor

Čapek, Karel	01, 04
Čechov, A. P.	02
Kafka, Franz	03

Titul

Bílá nemoc	04
Proces	03
R.U.R.	01
Proces	02

Rok

1928	02
1965	04
1990	01, 03

```

field name="version" type="long" indexed="true" stored="true"/>
<!-- Block join support -->
<field name="_root_" type="string" indexed="true" stored="true" multiValued="false" required="false"/>
<!-- Core Fields -->
<field name="id" type="string" indexed="true" stored="true"/>
<field name="fullrecord" type="string" indexed="false" stored="true"/>
<field name="marc_error" type="string" indexed="false" stored="true" multiValued="true"/>
<field name="allfields" type="text" indexed="true" stored="false" multiValued="true"/>
<field name="allfields_unstemmed" type="textProper" indexed="true" stored="false" multiValued="true"/>
<field name="fulltext" type="text" indexed="true" stored="true"/>
<field name="fulltext_unstemmed" type="textProper" indexed="true" stored="false"/>
<field name="spelling" type="textSpell" indexed="true" stored="true" multiValued="true"/>
<field name="spellingShingle" type="textSpellShingle" indexed="true" stored="true" multiValued="true"/>
<!-- Institutional Fields -->
<field name="institution" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="region_institution" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="collection" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="building" type="string" indexed="true" stored="true" multiValued="true"/>
<!-- Generic Fields -->
<field name="language" type="string" indexed="false" stored="true" multiValued="true" docValues="true"/>
<field name="format" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="author" type="textProper" indexed="true" stored="true" termVectors="true" omitNorms="true"/>
<field name="author-letter" type="string" indexed="true" stored="true"/>
<field name="authorStr" type="textFacet" indexed="true" stored="false"/>
<field name="title" type="text" indexed="true" stored="true"/>
<field name="title_sort" type="string" indexed="true" stored="true"/>
<field name="title_sub" type="text" indexed="true" stored="true"/>
<field name="title_short" type="text" indexed="true" stored="true"/>
<field name="title_full" type="text" indexed="true" stored="true"/>
<field name="title_full_unstemmed" type="textProper" indexed="true" stored="true"/>
<field name="title_fullStr" type="string" indexed="true" stored="true"/>
<field name="title_auth" type="text" indexed="true" stored="true"/>
<field name="physical" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="publisher" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="publisherStr" type="string" indexed="true" stored="false" multiValued="true"/>
<field name="publishDate" type="string" indexed="false" stored="true" multiValued="true" docValues="true"/>
<field name="publishDateSort" type="string" indexed="true" stored="false" sortMissingLast="true" omitNorms="true"/>
<field name="edition" type="string" indexed="true" stored="true"/>
<field name="description" type="text" indexed="true" stored="true"/>
<field name="contents" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="url" type="string" indexed="false" stored="true" multiValued="true"/>
<field name="thumbnail" type="string" indexed="false" stored="true"/>
<!-- Catalog Specific Fields -->
<field name="lccn" type="string" indexed="true" stored="true"/>
<field name="ctrlnum" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="isbn" type="isn" indexed="true" stored="true" multiValued="true"/>
<field name="issn" type="isn" indexed="true" stored="true" multiValued="true"/>
<field name="oclc_num" type="string" indexed="true" stored="true" multiValued="true" />
<field name="callnumber" type="string" indexed="true" stored="true"/>
<field name="callnumber-a" type="string" indexed="true" stored="true"/>
<field name="callnumber-first" type="string" indexed="true" stored="true"/>
<field name="callnumber-first-code" type="string" indexed="true" stored="true"/>
<field name="callnumber-subject" type="string" indexed="true" stored="true"/>
<field name="callnumber-subject-code" type="string" indexed="true" stored="true"/>
<field name="callnumber-label" type="string" indexed="true" stored="true"/>
<field name="dewey-hundreds" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="dewey-tens" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="dewey-ones" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="dewey-full" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="dewey-sort" type="string" indexed="true" stored="true" />
<field name="dewey-sort-browse" type="string" indexed="true" stored="true" multiValued="true" />
<field name="dewey-raw" type="string" indexed="true" stored="true" multiValued="true" />
<field name="author2" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="author2Str" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="author2-role" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="author_fuller" type="textProper" indexed="true" stored="true" />
<field name="author_additional" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="author_additionalStr" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="title_alt" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="title_old" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="title_new" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="dateSpan" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="series" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="series2" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="topic" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="topic_unstemmed" type="textProper" indexed="true" stored="false" multiValued="true"/>
<field name="topic_facet" type="string" indexed="false" stored="true" multiValued="true" docValues="true"/>
<field name="topic_browse" type="string" indexed="true" stored="false" multiValued="true"/>
<field name="author_browse" type="string" indexed="true" stored="false" multiValued="true"/>
<field name="genre" type="text" indexed="true" stored="true" multiValued="true"/>

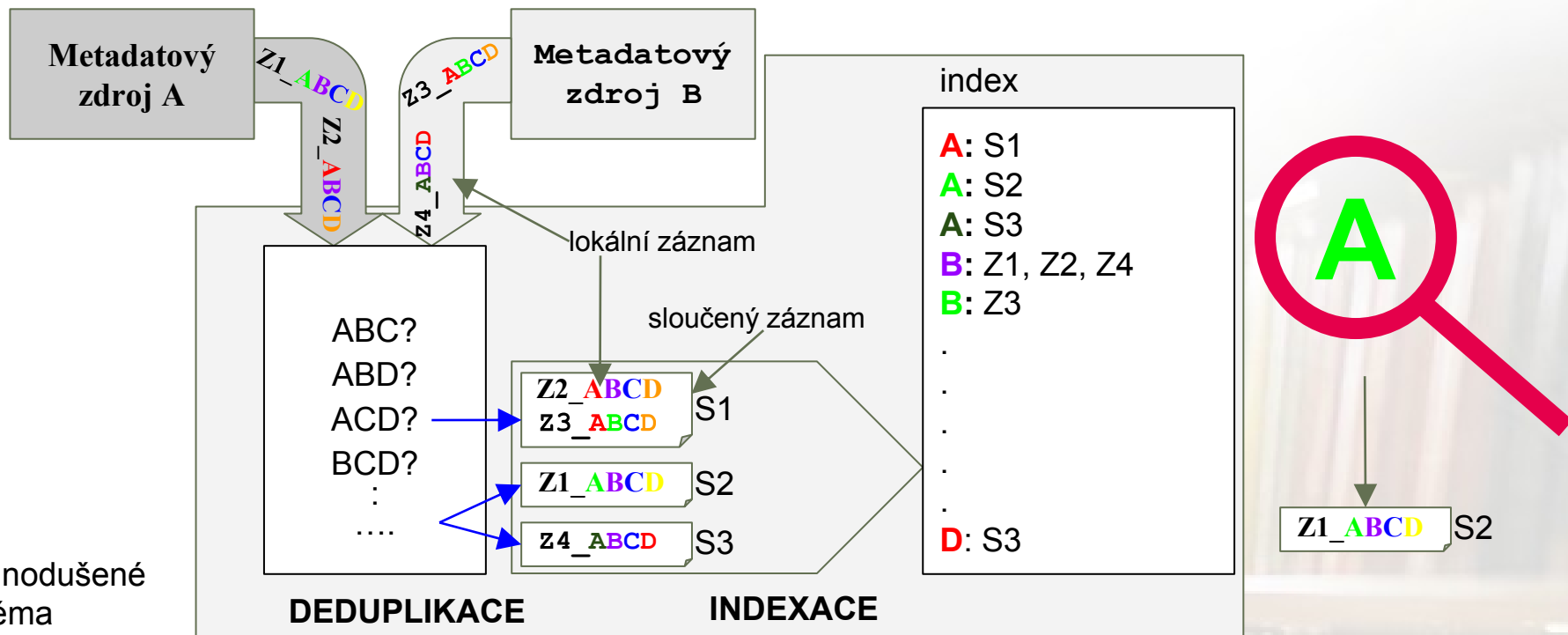
```

```

<field name="genre_facet" type="textFacet" indexed="true" stored="true" multiValued="true"/>
<field name="geographic" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="geographic_facet" type="textFacet" indexed="true" stored="true" multiValued="true"/>
<field name="era" type="text" indexed="true" stored="true" multiValued="true"/>
<field name="era_facet" type="textFacet" indexed="true" stored="true" multiValued="true"/>
<field name="illustrated" type="string" indexed="true" stored="true" multiValued="false"/>
<field name="long_lat" type="textFacet" indexed="true" stored="true" multiValued="false"/>
<!-- Container fields (i.e. for describing journal containing an article) -->
<field name="container_title" type="text" indexed="true" stored="true"/>
<field name="container_volume" type="text" indexed="true" stored="true"/>
<field name="container_issue" type="text" indexed="true" stored="true"/>
<field name="container_start_page" type="text" indexed="true" stored="true"/>
<field name="container_reference" type="text" indexed="true" stored="true"/>
<!-- Hierarchy Fields -->
<field name="hierarchytype" type="string" indexed="true" stored="true" multiValued="false"/>
<field name="hierarchy_top_id" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="hierarchy_top_title" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="hierarchy_parent_id" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="hierarchy_parent_title" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="hierarchy_sequence" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="is_hierarchy_id" type="string" indexed="true" stored="true" multiValued="false"/>
<field name="is_hierarchy_title" type="string" indexed="true" stored="true" multiValued="false"/>
<field name="hierarchy_browse" type="textFacetRaw" indexed="true" stored="false" multiValued="true"/>
<!-- Used for loading correct record driver -->
<field name="recordtype" type="string" indexed="true" stored="true"/>
<!-- Tracking fields to keep track of oldest and most recent index times -->
<field name="first_indexed" type="date" indexed="true" stored="true"/>
<field name="last_indexed" type="date" indexed="true" stored="true"/>
<!-- Custom fields -->
<field name="statuses" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="title_display" type="string" indexed="false" stored="true" multiValued="false"/>
<field name="publishDate_display" type="string" indexed="false" stored="true" multiValued="true"/>
<field name="nbn" type="string" indexed="false" stored="true" multiValued="false"/>
<field name="author_search" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="barcodes" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="sfx_links" type="string" indexed="false" stored="true" multiValued="true"/>
<field name="citation_record_type" type="string" indexed="true" stored="true" multiValued="false"/>
<field name="id001_search" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="author_find" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="inspiration" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="author_viz" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="corporation_viz" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="subject_viz" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="genre_viz" type="textProper" indexed="true" stored="true" multiValued="true"/>
<field name="upv_ipc_search" type="upv" indexed="true" stored="false" multiValued="true"/>
<!-- Authority fields -->
<field name="heading" type="string" indexed="true" stored="true"/>
<field name="use_for" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="see_also" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="source" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="scope_note" type="string" indexed="false" stored="true" multiValued="true"/>
<field name="id_authority" type="string" indexed="true" stored="true"/>
<field name="short_note_cs" type="string" indexed="true" stored="true"/>
<field name="short_note_en" type="string" indexed="true" stored="true"/>
<!-- Doesn't need to be stored after successful testing -->
<field name="loanRelevance" type="integer" indexed="true" stored="true"/>
<!--Indexed fields -->
<field name="region_district_town_str_mv" type="string" indexed="true" stored="true" multiValued="true" docValues="false"/>
<!-- Dynamic fields for customization without schema modification -->
<dynamicField name="*_date" type="date" indexed="true" stored="true"/>
<dynamicField name="*_date_mv" type="date" indexed="true" stored="true" multiValued="true"/>
<dynamicField name="*_isn" type="isn" indexed="true" stored="true"/>
<dynamicField name="*_isn_mv" type="isn" indexed="true" stored="true" multiValued="true"/>
<dynamicField name="*_str" type="string" indexed="false" stored="true" docValues="true"/>
<dynamicField name="*_str_mv" type="string" indexed="false" stored="true" multiValued="true" docValues="true"/>
<dynamicField name="*_txt" type="text" indexed="true" stored="true" docValues="false"/>
<dynamicField name="*_txt_mv" type="text" indexed="true" stored="true" multiValued="true" docValues="false"/>
<dynamicField name="*_txtF" type="textFacet" indexed="true" stored="true"/>
<dynamicField name="*_txtF_mv" type="textFacet" indexed="true" stored="true" multiValued="true"/>
<dynamicField name="*_txtP" type="textProper" indexed="true" stored="true"/>
<dynamicField name="*_txtP_mv" type="textProper" indexed="true" stored="true" multiValued="true"/>
<dynamicField name="*_int" type="integer" indexed="true" stored="true"/>
<dynamicField name="*_int_mv" type="integer" indexed="true" stored="true" multiValued="true"/>
<dynamicField name="*_boolean" type="boolean" indexed="true" stored="true" required="false" />
<dynamicField name="*_autocomplete" type="text_autocomplete" indexed="true" stored="true" multiValued="true" omitNorms="true" omitTermFreqAndPositions="true"/>
<dynamicField name="*_geo" type="geo" indexed="true" stored="true" multiValued="true" />
<dynamicField name="*_display" type="string" indexed="false" stored="true" multiValued="false" />
<dynamicField name="*_display_mv" type="string" indexed="false" stored="true" multiValued="true" />
<dynamicField name="*_exact" type="exact" indexed="true" stored="true" multiValued="false" />
<dynamicField name="*_exact_mv" type="exact" indexed="true" stored="true" multiValued="true" />

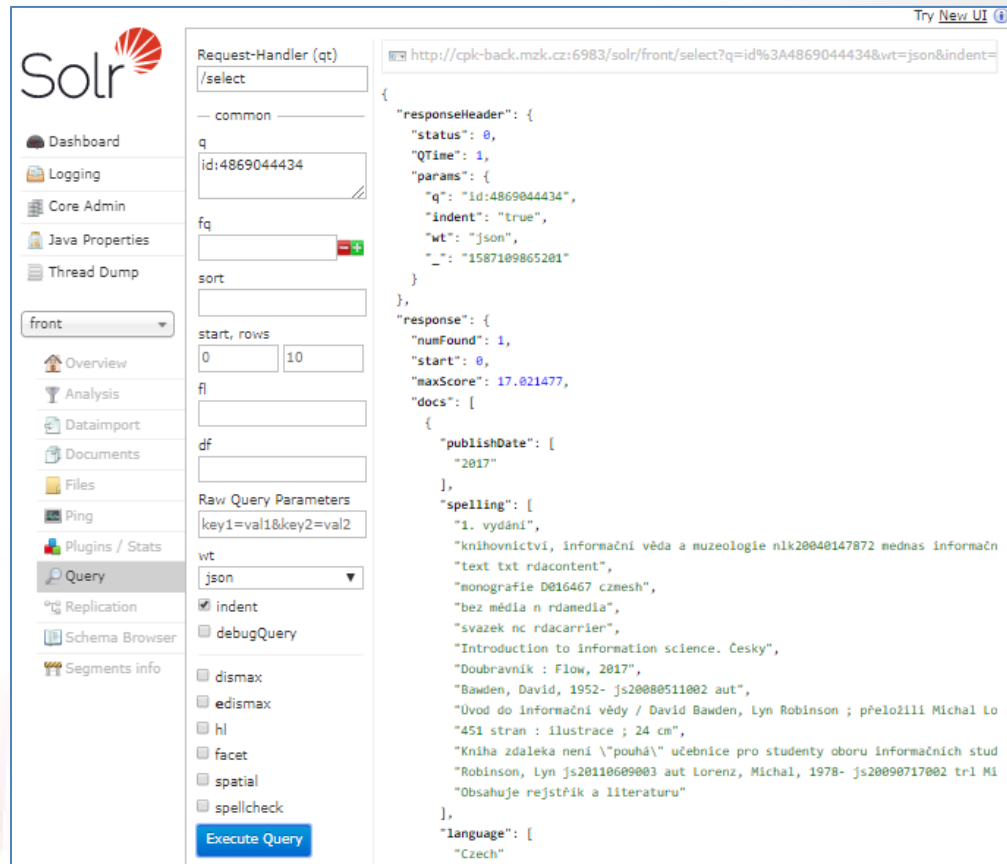
```

Deduplikační kroky



Indexace

- Index neslouží jen ke klasickému **vyhledávání**. Jsou v něm zvláštní typy polí pro **fasety, řazení či zobrazování** (v těch ani vyhledávat nejde).
- Používáme platformu **Solr** (založeno na knihovně Apache Lucene, stejně jako ElasticSearch)
- *Solr používá např. i DuckDuckGo nebo Netflix*



The screenshot displays the Solr Admin interface. On the left is a navigation sidebar with options like Dashboard, Logging, Core Admin, Java Properties, Thread Dump, Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (selected), Replication, Schema Browser, and Segments info. The main area shows the 'Request-Handler (qt)' configuration for '/select'. The 'q' field contains 'id:4869044434'. The 'wt' dropdown is set to 'json'. The 'indent' checkbox is checked. Below the configuration, the 'Execute Query' button is visible. On the right, the JSON response is shown, including 'responseHeader' with status, QTime, and params, and a 'response' object with 'numFound', 'start', 'maxScore', and a list of 'docs' containing document details like 'publishDate', 'spelling', and 'language'.

Správce zdrojů (Record Manager 2)

- **Celý dosavadní proces**
(tj. sklizeň, deduplikaci a indexaci)
dělá [Správce zdrojů](#)
(Java, inspirovaný Record Managerem z Finna.fi)

VuFind (frontend)

- **Open source online katalog vylepšený o vyhledávání ve více zdrojích najednou**
- Vyvíjený na Villanova University (ale podílet se mohou i jiní – např. i MZK)
- PHP (+ Apache + MySQL) a využívání Solru
- Více zde: <https://vufind.org/vufind/>
- Pro potřeby portálu byly nutné zásadní modifikace (*práce s uživ. účty, sloučenými a lokálními záznamy, AND a OR hierarchickými fasetami, komunikace s různými knihovními systémy, oborové brány...*)

Řazení výsledků vyhledávání dle relevance

- Konfigurace pro Solr ve VuFindu
- Nastavuje, ve kterých indexových polích má Solr hledat zadaný výraz
- Důležitost jednotlivých polí upřesňuje ^boost
- Lze specifikovat i nakládání s vyhledávacím výrazem (přesná shoda má vyšší relevanci než substring, substring má vyšší relevanci než výskyt samotných nesousedících slov)

Příklad konfigurace pro vyhledávání dle názvu

```
adv_search_title_series:
  DismaxFields:
    - title_short^500
    - title_full^400
    - title^300
    - title_alt^200
    - title_new^100
    - title_old
    - titleSeries_search_txt_mv
  DismaxParams:
    - [bq, "inspiration:top_results^1"]
    - [bf, "if(exists(merged_records_boolean),10000,0)" ]
    - [bq, "author_exact:\\"@query\\""^20000"]
    - [bq, "title_exact:\\"@query\\""^10000"]
    - [bq, "term_exact:\\"@query\\""^30000"]
    - [bf, ord(publishDateSort)^1]
  DismaxHandler: edismax
  QueryFields:
    titleSeries_search_txt_mv:
      - [onephrase, 100]
      - [and, 50]
  FilterQuery: "NOT recordtype:library AND NOT recordtype:sfx"
```

Komunikace s knihovními systémy

- **Dotahování aktuálních statusů a dostupností knihovních jednotek**
 - párováno podle identifikátorů, které už známe ze sklizených záznamů
- **Funkce čtenářského konta**
 - pamatujeme si jen ID uživatelů
 - vše ostatní dotahujeme jen pro potřebu zobrazení
- Nutné **zabezpečení**: nejen SSL, ale i zamítání požadavků, které nejsou autentizované či z povolených IP adres

Komunikace s knihovními systémy

- **Aleph a Koha**
 - Vlastní API
 - Request: `https://url.cz/api?q1=v1&q2=v2&...`
 - Response: JSON nebo XML
- **Ostatní knihovní systémy**
 - [NCIP](#) = NISO Circulation Interchange Protocol
 - Request i response: XML
- **Na naší straně nutný driver pro každý takový kanál (spousty výjimek pro každý systém)**

NCIP

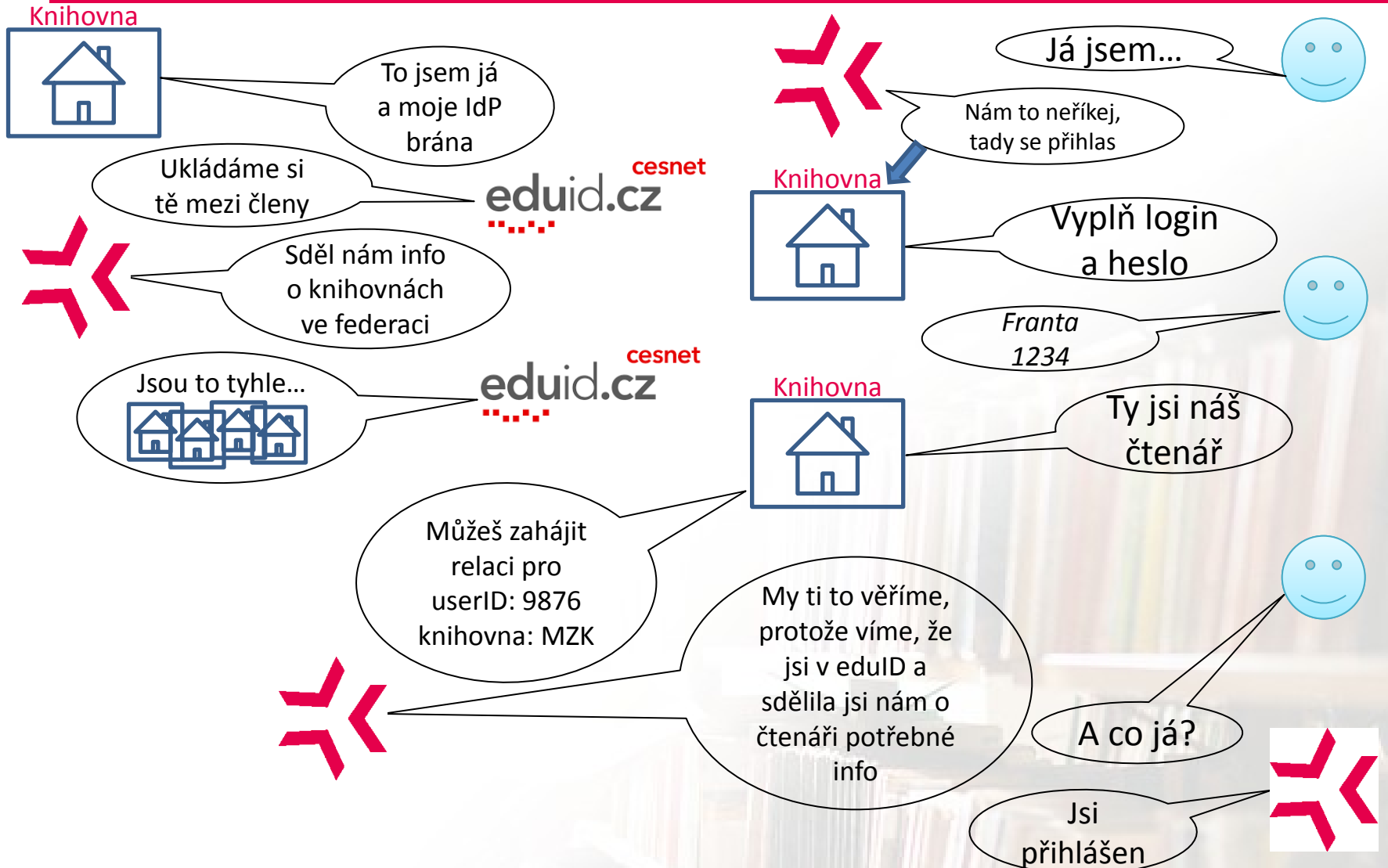
```
<?xml version="1.0" encoding="UTF-8"?>
<ns1:NCIPMessage
  xmlns:ns1="http://www.niso.org/2008/ncip" ns1:version="http://www.niso.org/schemas/ncip/v2_02/ncip_v2_02.xsd">
  <ns1:LookupAgency>
    <ns1:InitiationHeader>
<ns1:FromAgencyId>
  <ns1:AgencyId ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyidtype/agencyidtype.scm">
    CPK</ns1:AgencyId>
</ns1:FromAgencyId>
<ns1:ToAgencyId>
  <ns1:AgencyId ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyidtype/agencyidtype.scm">
    CHG001</ns1:AgencyId>
</ns1:ToAgencyId>
    </ns1:InitiationHeader>
    <ns1:AgencyId ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyidtype/agencyidtype.scm">
      CHG001</ns1:AgencyId>
    <ns1:AgencyElementType ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyelementtype/
      agencyelementtype.scm">Agency Address Information</ns1:AgencyElementType>
    <ns1:AgencyElementType ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyelementtype/
      agencyelementtype.scm">Agency User Privilege Type</ns1:AgencyElementType>
    <ns1:AgencyElementType ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyelementtype/
      agencyelementtype.scm">Application Profile Supported Type</ns1:AgencyElementType>
    <ns1:AgencyElementType ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyelementtype/
      agencyelementtype.scm">Authentication Prompt</ns1:AgencyElementType>
    <ns1:AgencyElementType ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyelementtype/
      agencyelementtype.scm">Consortium Agreement</ns1:AgencyElementType>
    <ns1:AgencyElementType ns1:Scheme="http://www.niso.org/ncip/v1_0/schemas/agencyelementtype/
      agencyelementtype.scm">Organization Name Information</ns1:AgencyElementType>
  </ns1:LookupAgency>
</ns1:NCIPMessae>
```

Přihlašování uživatelů

- Na portálu si nelze založit účet, je třeba mít registraci v knihovně (ale stačí i Google, FB apod.)
- Knihovna musí být členem federace poskytovatelů identit eduID.cz
 - Knihovna pak má přihlašovací IdP bránu
 - Služby jako Knihovny.cz tam nechávají přihlásit uživatele, aniž by se dozvěděly jejich login a heslo
 - Knihovna pak službě uvolní takové údaje o uživateli, jaké služba potřebuje znát (tzv. atributy)



Přihlašování uživatelů



Přihlašování uživatelů

- Některé knihovny byly v eduID.cz už před zapojením do portálu kvůli EIZ pro čtenáře
- Pro většinu knihoven je ale zprovoznění vlastní IdP problém
 - Pověří zaměstnance MZK a my jim vše zařídíme
 - *(Tzn. velkou část konverzace na předchozím slidu vedeme sami se sebou)*

Provoz a vývoj

- Produkční a testovací servery
- Maximální využití možností virtualizace (Docker)
- Git (na každou issue vlastní větev a kontejer)
- Neustále vyvíjíme:
 - přechod na Vufind 6 (z v2) – cíl je držet se upstreamu
 - zrychlování odezvy optimalizací indexu a AJAXu
 - úpravy deduplikace a indexace
 - nové funkce a reflektování podnětů knihoven (aktivní je třeba KFBZ, MKP nebo SVKPK)

Děkuji za pozornost. Nějaký dotaz?

www.knihovny.cz

Možnosti vyhledávání

[Historie vyhledávání](#)

O portálu knihovny.cz

[O portálu](#)
[Zapojené knihovny a zdroje](#)
[Historie portálu Knihovny.cz](#)
[Jak se zapojit pro knihovny](#)

O knihovnách

[Projekty knihoven](#)
[Služby knihoven](#)
[Významné akce knihoven](#)
[Ptejte se knihovny](#)

Hledáte pomoc?

[Zpětná vazba](#)
[Nápověda](#)
[Možnosti přihlášení a registrace](#)

Centrální portál knihoven

Portál Knihovny.cz je výsledkem několikaletého vývoje směřujícího k vytvoření centrálního místa, kde můžete současně prohledávat katalogy knihoven z celé České republiky. Cílem bylo najít řešení, díky němuž snadno získáte knihu či informace, aniž byste danou knihovnu museli navštívit nebo složitě vyhledávat v databázích každé jednotlivé knihovny.

Vyvíjí



Sledujte nás

