



KNIHOVNY.CZ

Vyhledávání na portálu Knihovny.cz

Petr Žabička, Václav Rosecký, Petra Žabičková

Moravská zemská knihovna v Brně

Obsah

- Co indexuje portál Knihovny.cz
- Rozšíření o centrální index
- Hledání v portálu
- Řazení výsledků

Obsah portálu

- Katalogy velkých a některých menších knihoven
- Souborný katalog ČR (bez přímo zapojených knihoven)
- Bibliografické databáze (ANL, BMČ,...)
- Open Library (volně dostupné)
- Normy (ČSN)
- Digitální knihovny (Manuscriptorium, Kramerius MZK)
- Patentové informace (příhlášky, patenty, průmyslové vzory)
- Zákony (pro lidi)
- Autoritní databáze (jmenné authority, Osobnosti regionu)
- Adresář knihoven - prohledávaný samostatně

Klíčové změny od spuštění

Mimo přidávání dalších zdrojů:

- **vylepšení relevance,**
- úpravy v hledání,
- vylepšení zobrazování výsledků dotazů,
- vylepšení deduplikace,
- zprovozněna možnost stahování záznamů přes Z39.50 (sdílená katalogizace)
 - můžeme vytvářet profily jak z vlastního indexu, tak ze Z39.50 zdrojů

Připravujeme

- posílení hardwaru - zkrácení odezvy
- zapojení dalších knihoven
 - vítáme všechny zájemce
 - nabízíme konzultace
- indexace článkových databází
- zapojení centrálního indexu

A co Anopress ?

- Zatím nezapojen (žádáno o prostředky na zakoupení roku 2017, nepřiděleno z důvodu nedostatku prostředků ve VISK8B)
- Nabídka na zakoupení 31 mil. záznamů (jen metadata: 1996 - 2017)
- Každý další rok dokoupení 1,5 - 2 mil. záznamů
- Odkazy na plné texty, ale ne fulltextový index

Chceme Anopress v portálu?

Centrální index

- Běží výběrové řízení na dodavatele indexu
 - 2. pol. 2017 + opce na rok 2018
 - Udržitelnost financování??
- Výsledky vyhledávání v lokálním a centrálním indexu budou zobrazovány odděleně
- Využití SFX JIB

Pozn. EDS indexuje Newton Media, Summon ne - knihovny ale kupují Anopress. Neexistuje převodník odkazů z NM do Anopressu. NM zpřístupňuje nejen text, ale i původní podobu novin.

Otázky k integraci CI

- Jak nabídnout uživateli možnost hledání v CI?
 - zatržítka u vyhledávacího okna
 - druhá záložka u seznamu výsledků
 - nechceme zobrazovat dva seznamy výsledků zároveň (rozdílné fasety)
- Národní index vrací na prázdný dotaz všechno
 - Summon také všechno
 - EDS prázdný seznam

Hledání v portálu

- co je to Solr
- základní vyhledávání
- pokročilé vyhledávání
- fasety, filtry
- úryvky
- relevance
- vychytávky
- limity

Co je to Solr

- prohledávač plných textů
- podporuje fasetovou navigaci
- zvýrazňuje vyhledané termíny v úryvcích textu
- řazení dle relevance
- geografické vyhledávání
- našeptávač
- řada možností pro rozložení zátěže na více serverů
- možnost pokládání komplexních vyhledávacích dotazů

Základní vyhledávání

- Jedno vyhledávací okno
- pokročilé vyhledávací operátory
 - AND, OR, NOT
 - ?,* (náhrada jednoho nebo více znaků)
 - ~ (podobnost - Mácha~0,8)
 - ~ (vzdálenost slov - "Mácha Máj"~3)
 - ^ (váha slova - Mácha Máj^8 básně^6)
 - rozsahy: {Ma TO Mc} nebo [2001 TO 2003]
 - používání závorek
- Našeptávač - rozdělení na název, autor, předmět (např. Babička)

Pokročilé vyhledávání

- Umožňuje sestavení komplexních dotazů bez nutnosti znát syntaxi
- Dvě úrovně - skupiny + výrazy
 - skupiny lze spojit pomocí AND nebo OR
 - výrazy pomocí AND, OR, NOT

? Základní vyhledávání [Uložit hledání](#)

Všechny skupiny

Všechny výrazy

Ve všech polích	Jan Hus	×
Ve všech polích	Karel IV	×

[Přidat vyhledávací řádek](#)

Neobsahuje výrazy

Ve všech polích	Jan Žižka	×
-----------------	-----------	---

[Přidat vyhledávací řádek](#)

[Přidat vyhledávací skupinu](#)

Specifika

- Diakritika - ignorujeme háčky a čárky
 - uživatelé je vždy nepíše (na mobilu apod.)
 - ale zebra vs. žebra apod.
 - změna po posílení HW - nejprve to, co uživatel napsal, pak varianta s diakritikou
- Lemmatizace je vypnutá, bude analyzováno po posílení HW (násobné zvětšení indexu)
- Podobnost (fuzzy search) je při standardním hledání vypnutá, vracela příliš mnoho irelevantních výsledků - bude ale využita v případě, kdy přesný dotaz nic nenajde

Specifika

- Obohacování záznamu při indexaci využití řízených slovníků:
 - alternativní jména autorů (hledání Vladimíra Vaška vrátí i díla Petra Bezruče, pokud byl daný záznam titulu napojen na Národní autority)
 - synonyma u předmětových hesel (národní autority, MESH)
- Úryvky - zobrazí se první výskyt hledaného termínu (ale limitováno pamětí serveru)

Řazení výsledků dle relevance

- Vyhodnocení řady kritérií:
 - jak často se hledaný termín v dokumentu vyskytuje?
 - jak často se termín vyskytuje celkově?
 - jak je nalezený dokument dlouhý?
 - vyskytuje se termín v názvu, autorských nebo jiných údajích nebo v textu dokumentu?
 - jak je starý?
 - je ve fondu mé knihovny nebo knihoven?
- Přidělení různých vah jednotlivým kritériím
- Vyhodnocení celkového pořadí výsledků

Limity

- Pochopitelnost pro uživatele (např. využití faset nebo pokročilého vyhledávání)
- Identifikace jazyka (dotazu, metadat i dokumentu) a aplikace jazykově specifických funkcí (např. lemmatizace)
- Velikost jednotlivých fulltextů (zejm. tituly periodik)
- Požadavky na pokročilá vylepšení vyhledávání versus rostoucí velikost indexu a času zpracování dotazu

Děkuji za pozornost!

Otázky?

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

