



ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ
FF UK V PRAZE

Jan Pokorný

Velký index
jako moderní integrační nástroj

Verze 1.0

Praha

2011

Problematika velkého indexu

1. Proč optimalizovat vyhledávacího prostředí?

V dnešní době začíná 80% uživatelů své hledání na webu. Proto se v posledních pěti letech se pracuje na nových online katalozích, tzv. next-generation OPACích - „**discovery systémech**“: došlo k posunu paradigmatu způsobu hledání, sílí snaha nabídnout ucelené vyhledávání. Velká část zahraničních knihoven přešla nedávno na nová rozhraní. Je to luxus nebo součást strategie knihoven? Nové systémy kladou důraz na koncového uživatele, jeho potřeby, jeho samostatnost a spokojenost. V akademickém prostředí tyto systémy pomáhají tomu, aby studující a pedagogové využívali elektronické informační zdroje. Velmi užitečný je portál [těchto nových systémů](#) [2], vytváří v posledních třech letech Marshall Breeding.

2. Discovery systémy

Discovery systémy

- vyhledávají rychle
- vyhledávají jednoduše
- vyhledávají intuitivně (bez zaškolení)
- podporují fasetovou navigaci
- podporují RSS
- mají relevantní obsah

Discovery systémy mají výraznou přidanou hodnotu: jedno společné rozhraní znamená *méně práce s lepšimi výsledky*. Výkonnější vyhledávání umožňují nové způsoby indexace. Vytvořením jednotného indexu informačních zdrojů knihovny umožňuje snadnou a efektivní cestu k celé kolekci knihovny (tištěnému fondu i elektronickým fulltextům) z *jediného vyhledávacího prostředí*.

Zastaralejší vyhledávací systémy nemají jednotné vyhledávacího prostředí:

- navštěvujeme více vyhledávacích míst
- musíme přecházet mezi systémy
- opakovaně se přihlašujeme
- ztrácíme kontinuitu vyhledávání
- „jdeme přes Google“
- systém nám nenabízí jednotné výsledky
- knihovna si nemůže do systému zařadit libovolný zdroj

Discovery systémy pomáhají v boji s množstvím informací:

- mají konfigurovanou relevanci (řadí výsledků: rozliší „zlato“ versus „šum“)
- personalizují výpočet relevance - při seskupování výsledků zohledňují uživatele: studenti 1. ročníku dostanou jiné řazení dokumentů než doktorandi
- seskupují výsledky (FRBR)
- využití chování jiných uživatelů - doporučení dalších titulů
- zapojení uživatelů

Cíle discovery systémů jsou:

- zodpovědět vyhledávací dotaz *všemi* zdroji knihovny
- dát každému ze zdrojů *stejnou* šanci na zodpovězení dotazu
- pro výsledky se stejnou relevancí se zobrazit 1 výsledkovou množinu
- plné texty dostupné na 1 kliknutí
- jediné přihlašování, pouze 1 hledání

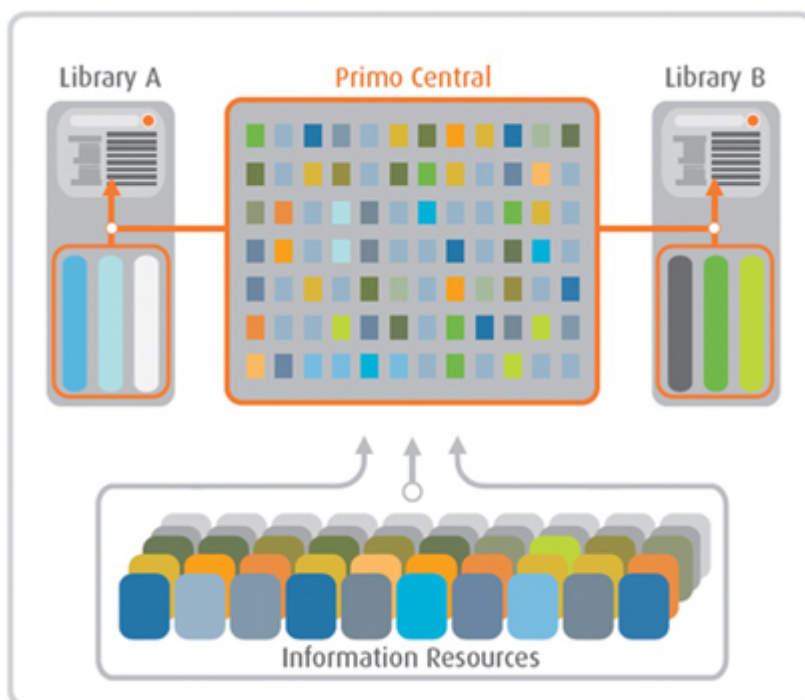
Mezi tyto systémy patří systém Primo:

K vyhledávacímu dotazu se zobrazí po 1 kliknutí vyhledané informace na stejném místě (včetně vlastnictví nosičů) a zprostředkuje se dodání (pokud jsou dostupné plné texty). Celý proces je oproti současným možnostem rychlejší a ve vyhledávání důkladnější - rychle získáváme více dokumentů. V souvislosti s tím, že danou službu poskytuje knihovna, je samozřejmé, že vyhledané informace jsou uspořádané a výsledky jsou seřazené podle relevance; způsoby řazení výsledků lze nastavit. Zde je možné nahlédnout do [zkušební verze](#) ^[3] systému Primo.

4. Co je to velký index?

Systém vyhledává dokumenty pomocí discovery indexu. Discovery systém si index buduje sám. Discovery systém se stará totálně o lokální zdroje a na dálku (nemá je u sebe) o elektronické informační zdroje (EIZ).

Všechny zdroje knihovny jsou indexovány a prohledávány, největší část zdrojů tvoří elektronické zdroje, pouze tato největší část je nazývána „**velký index**“.



Příkladem velkého indexu je indexační služba Primo Central (jedná se pouze o část systému Primo), viz [YouTube](#) [41]. Indexační služba Primo Central není lokalizována v knihovně, běží na vzdálených serverech, připojují se k ní různé knihovny.

5. Velký index versus metavyhledávání

Obvyklá struktura uživatelských dotazů je:

- ½ dotazů je na 1 konkrétní dokument - discovery systémy by měli preferovat tento typ dotazu (dokument měl by být vždy nabízený v první desítce výsledků)
- tematické dotazy - dotaz, který má uživatele dovést dál - zde má uživatel získat širokou nabídku výsledků

Při metavyhledávání (například Jednotná informační brána (JIB) nebo 360 Search) systém *paralelně* hledá ve sbírkách; my víme, kde hledáme a co za výsledky se nám vrací.

Velký index *vyhledává v obsahu*, a pak se přesměruje do plného textu: má oddělené vyhledávání a dodávání. Dodávání může být omezeno na jednu nebo více plnotextových databází (např. **z finančních důvodů**):

- vyhledání (např. v databázi Gale, EBSCO, ProQuest)
- indikace, jestli k plnému textu můžu
- zobrazení plného textu

Dostupnost zdrojů pro uživatele může být různá:

- zdroje dostupné všem, obsah také dostupný komukoliv; vyhledávání zdarma, dodávání zdarma (např. Guttenberg.org ^[6])
 - je povoleno vyhledávání, text dodán jen pokud je předplácen
 - není povoleno ani vyhledávání: např. Web of Knowledge - i vyhledávání je placené
- Dodávání vychází z elektronických holdingů: informace o exemplářích pochází z linkovacího serveru. Elektronické holdingy mají smysl jen pro uživatele dané knihovny. Nejpřesnější informace o EIZ (elektronických informačních zdrojích) poskytují dodavatelé. Velcí dodavatelé EIZ zatím své databáze do velkého indexu neuvolnily. **EBSCO** se v roce 2010 rozhodlo, že do projektu velkého indexu nevstoupí, protože má kvalitní databázi a rozhraní (a také vyvíjí svůj discovery systém). **ProQuest** proklamuje, že se chce dohodnout, ale jednání již trvají 3 roky - zatím zvolil vyčkávací taktiku. Tyto firmy mají svá aktivní rozhraní, která pro jejich zdroje nejlépe fungují a jsou nejlépe optimalizována.

Ovšem paralelní vyhledávání je vždy o jednu třídu horší než narativní rozhraní:

pokud velký index zapojíme do metavyhledávače (např. do JIB), tak ten je potom výrazně rychlejší; nenajde všechny zdroje, ale nalezne zdroje nejfrekventovanější. U Primo Central je za poplatek cca 3000€ možné získat přístup k EIZ jiné knihovny, např. Harward Library ^[6]. Velký index může být také cílem v paralelním vyhledávání.

6. Jak vzniká velký index - plnění a vlastnosti

Firma Ex Libris buduje velký index od roku 2009. Vkládají se data od tradičních agregátků a doplňují se volná data.

Vyhledávání ve velkém indexu probíhá na úrovni článků.

Velký index má svou „cestovní mapu“: správce velkého indexu ví, co v něm má a co nemá, a o co by měl usilovat.

Petice IGELU chce přesvědčit Ebsco, aby do projektu velkého indexu vstoupilo. V budoucnosti je pravděpodobné kombinování více velkých indexů (např. lokálního aglobálního).

Pro kvalitu velkého indexu je rozhodující:

- pokrytí
- relevance
- dostupnost
- neutralita vůči knihovně

- jak zaleží na pokrytí lokálním obsahem?

7. Zapojení českého obsahu do velkého indexu

Národní knihovna ČR ukončila vytváření článkové bibliografie (databáze ANL), v ideálním případě by ji mohl velký index „nahradit“! Jaké české články by mohly být v „českém“ velkém indexu?

U článků jsou důležité tyto dva aspekty:

- obsah musí být k dispozici online (nemusí být pro všechny zdarma)
- dodávání vázáno na licenci

Potenciál na vložení do velkého indexu mají:

- Souborný katalog ČR, databáze ANL, registr digitalizace
- oborové bibliografie
- národní digitální knihovny ([Manuscriptorium](#) ^[7], [Kramerius](#) ^[8], [WebArchiv](#) ^[9])
- oborové digitální knihovny ([DML](#) ^[10], [NUŠL](#) ^[11])
- placené zdroje ([Anopress](#) ^[12], [Newton Media](#) ^[13], [Infobanka ČTK](#) ^[14], [Beck online](#) ^[15], [Wolter Kluwer](#) ^[16] -
ASPI: jejich rozhraní má ideální popisná data do velkého indexu

Nová média:

- české wiki stránky
 - archivy vysílání
 - e-knihy
 - databáze publikační činnosti
 - moodle, doporučená literatura, nástroje pro podporu vzdělávání
 - institucionální repozitáře
- = zařazení zdrojů, které byly doposud na okraji zájmů knihoven

- další volné zdroje

Národní knihovna by mohla být ideálním nositelem řízení národní strategie plnění velkého indexu. (Pro větší vstřícnost potenciálně konkurenčních subjektů je nutná participace knihovny nebo jiná nestranné organizace.)

8. Vyzkoušení velkého indexu v praxi

Při ukázce vyhledávání ve velkém indexu se otevíralo dost otázek, ale již nebylo mnoho času. Změny k lepšímu byly provázeny nečekaným „znásobením“ výsledků - systém neprovádí deduplikace. Bude

velký index reálná cesta pro přiblížení se ke Googlu?

V závěru semináře byl zmíněn článek Viléma Sklenáka ^[17] v ProInflow.

Seminář byl pořádán ve velmi příjemném prostředí.

9. Vlastní shrnutí

Před knihovnami je další výzva - jak zjednodušit a rozšířit online vyhledávání informací. Z nových systémů na vyhledávání mi nejvíce líbí holandský Aquabrowser ^[18]. Systém Primo včetně Primo Centralu má velkou šanci u nás uspět. Dalším placeným discovery systémem, který se u nás nyní prezentuje, je Ebsco discovery system (EDS). Ale budou mít naše knihovny na tyto systémy prostředky a pro co se rozhodnou?