

# Rešeršní činnost

Mgr. Petr Šmejkal

43262@mail.muni.cz

# Rešeršní strategie

- možnosti, jak postupovat při vyhledávání
- výzkum práce referenčních knihovníků a rešeršérů

**1. strategie stavebních kamenů**

**2. vyhledávání pomocí  
nejspecifičtější fasety**

**3. strategie rostoucí perly**

**4. osekávání**

# Strategie stavebních kamenů

- samostatné dílčí dotazy vyjadřující ústřední pojmy původního rešeršního požadavku
- identifikace klíčových/významných pojmů
- množina výrazů vztahující se k pojmu: synonyma, kvazisynonyma, pravopisné formy, nadřazené, podřízené výrazy
  - **OR**, truncation (krácení podle slov. kořenů), stemming, wild cards (zástupné znaky)
- spojení dílčích formulací ve finální soubor
  - **AND**
- vhodné použít, když usilujeme o úplnost u úzce specifikovaných témat
- lze předpokládat, že budeme dotaz zužovat

# Strategie stavebních kamenů

- Př.: Léčba atypických očních vad neinvazními metodami

Léčba – léčení,  
medicína, uzdravovací  
metody, léčebné  
procesy, ...

**AND**

Atypický –  
nestandardní,  
neobvyklý, málo  
častý, ...

**AND**

Oční vady – poruchy,  
nemoci, zákal, ...

**AND  
NOT**

Invazní metody –  
operace, ...

# Strategie stavebních kamenů

## POJEM 1

- vyhledávací služby OR vyhledávací nástroje

## POJEM 2

- klasifikační schémata OR klasifikace OR  
kategorizační schémata OR klasifikační  
systém

## DOTAZ:

(vyhledávací služby OR vyhledávací nástroje) AND  
(klasifikační schémata OR klasifikace OR  
kategorizační schémata OR klasifikační systém)

# Příklad v LLIS

řešeršní požadavek:

vzdělávání dospělých v knihovnách se zřetelem na zlepšení jejich informační gramotnosti

# Příklad

- adult education OR lifelong learning
- information literacy OR information skills
- Library

## Taktiky:

různé taktiky pro zúžení, např.: omezit na deskriptory, proximitní vyhledávání, chronolog. zpřesnění, výběr pole vyhledávání (v titulu, abstraktu apod.)

# Vyhledávání pomocí nejspecifičtější fasety

- vztahuje se k vyhledávání složených témat – více aspektů
- uživatel musí znát všechny dílčí témata a musí být schopen určit, které téma je nejspecifičtější
- Vyhledávání
  - podle nejužšího pojmu z řešeršního požadavku
- pokud je výsledek uspokojivý, nemusí být do řešerše zahrnuta další dílčí hlediska



# Příklad

- rešeršní požadavek: „předmětová kategorizace IZ v oborových informačních branách“
  - categorization, classification, scheme
  - quality-controlled subject gateways, subject gateways, portal, gateways
  - zahájení vyhledávání na základě nejužšího pojmu „quality-controlled subject gateways“
- rešeršní dotaz ("quality-controlled subject gateways") <in> ALL

# Příklad

- příklady taktik
  - vyhledání pod autorem
  - rozšíření na vyhledávání v keywords, tj.vyhledávání v Subject, Title, Abstract, Author, Journal name, ...

# Strategie rostoucí perly

- Dotaz je postupně modifikován dle výsledků rešerše – záznamy jsou postupně procházeny a zjišťovány relevantní termíny (řízené termíny, slova z názvů apod.), které jsou použity k revidování dotazu.
- Cílem je alespoň jeden záznam
  - zjištění použitelných selekčních termínů
  - úprava formulace rešeršního dotaz

# Strategie osekávání

- první formulace dotazu - širší formulace, tj. pomocí obecného pojmu – cílem je vyhledání více záznamů (hitů)
- postupná specifikace dotazu
- uplatnění taktik pro zúžení záběru (AND, NOT, proxim. oper., field searching, formální omezení)
- formulace širší kategorie (obor, vědní disciplína), klasifikace
- náročnější na čas

# Příklad

- Vytvořte dotaz v db LLIS
  - řešeršní požadavek: Fenomén Web 2.0 a jeho vliv na knihovny
- řešeršní dotaz:
  - Web 2.0 and Library
  - zužte dotaz pomocí nabízených deskriptorů
  - omezte na plnotextové dokumenty
  - vyzkoušejte v jiných db – LISA, ProQuest

# Rešeršní strategie – praktické rady

## **Bud'te flexibilní**

- berte připravené kroky strategie orientačně
- přizpůsobujte další taktiky výsledkům rešerše
- nulový výsledek – hledání příčiny

## **Využívejte řízených slovníků**

- využívejte souvisejících pojmů ke konkrétnímu řízenému termínu (nadřazené, podřazené pojmy)
- nikdy nespojujte termíny s malou frekvencí výskytu (zjistitelné v katalogu) operátorem AND

# Rešeršní strategie – praktické rady

## Vytvářejte množiny termínů

- je velmi důležité k jednotlivým klíčovým slovům vytvářet množiny souvisejících termínů
- termíny v množině se spojují pomocí logického součtu – OR

## Využívejte klasifikací

- v českých knihovnách zejména klasifikování podle znaků MDT
- pomocí klasifikací vyhledáte většinou mnoho záznamů, proto se hodí jejich využití při strategii osekávání

# Rešeršní strategie – nejčastější chyby

## Logické chyby

- chybné kombinace výrazů pomocí booleovských operátorů (AND místo OR a naopak)

## Při psaní slov

- záměna O x 0, písmeno versus číslice



# Rešeršní strategie – nejčastější chyby

## Neuvědomění si odlišností dotazovacích jazyků jednotlivých databází

- rozdílné dotazovací jazyky a s tím spojená odlišná formulace rešeršního dotazu

## Ignorování rozdílů mezi různými vyhledávacími systémy

- je nutné si uvědomit, že databáze, elektronické katalogy a vyhledávací nástroje internetu jsou odlišné vyhledávací systémy a liší se charakter vyhledávání pomocí nich a typ vyhledaných dokumentů

# Věcné vyhledávání

# Věcné vyhledávání / subject searching

-tj. vyhledávání, kdy uživatel/rešeršér usiluje o nalezení dokumentů k určitému tématu (X uživatel ví, jaký dokument hledá, zná např. autora, část titulu apod.)

Jeden z klíčových problémů při vyhledávání v rešeršních systémech:

- Jaké vyhledávací výrazy by měly být vybrány pro formulaci dotazu?
- ---> Odkud by měly být termíny vybrány?

# Výběr termínu pro formulaci dotazu

Rešeršér – dva základní okruhy zdrojů termínů:

- během interakce s uživatelem před a během vyhledávání
- během interakce s rešeršním systémem

# Věcné vyhledávání/subject searching

Dva způsoby:

- pomocí pořadacích znaků/prvků věcných sj – deskriptorů, předmětových hesel, klasifikačních znaků
- pomocí přirozeného jazyka

V praxi se doporučuje kombinovat vyhledávání pomocí přirozeného jazyka i pomocí věcného SJ – obojí v konkrétních případech přispívá ke zlepšení přesnosti a úplnosti

# Důležité termíny

- **věcný SJ** – umělý jazyk, „jazyk používaný pro zpracování dokumentů pomocí věcných údajů s cílem umožnit vyhledávání dokumentů podle obsahu“
- **přirozený jazyk v IR**– jazyk, kterým lidé mluví a píší, není pro potřeby IR limitován a definován (týče se slovníku, syntaxe, sémantiky, vztahů)
  - jazyk užívaný pro formulaci dotazu bez „konzultace“ řízeného slovníku

# Formulace dotazu a ladění rešerše

Jde o základní okruhy využití přirozeného a selekčního jazyka.

- Formulace dotazu viz přednáška č. 2 – formulace rešeršního dotaz
- Ladění rešerše – query expansion (Shiri, 2002)
  - manuální – uživatel se rozhodne, jak může být výsledek rešerše využit pro další úpravu dotazu
  - interaktivní – uživatelé vybírají systémem navržené vyhl. výrazy (např. LLIS, ProQuest)
  - automatické – vyhledané dokumenty, které označil uživatel jako relevantní jsou systémem vyhodnoceny (určení sady vyhl. výrazů pro nové hledání) a je provedeno nové vyhledávání

# Efektivní věcné vyhledávání

Efektivní věcné vyhledávání vyžaduje následující druhy znalostí:

- znalost polí, které mohou být pro vyhledávání využity a jejich charakteristiky
- znalost věcného SJ, který systém využívá
- znalost strategií, kde a jak je aplikovat
- znalost vyhledávacích možností systému a jak je použít
- znalost tématu
- znalost toho, jak převést informační potřebu na informační dotaz



# Příklad

Vyzkoušejte vyhledávání v katalogu NK ČR  
– nejprve pomocí předmětu  
postmodernismus (zvolte vhodné pole)

dále dle – postmodernismus literatura

# Selekční jazyky

Usnadňují vyhledávání tím, že:

- umožňuje kontrolovat synonyma a kvazisynonyma (zvyšuje úplnost)
- umožňuje rozlišit homonyma, kvalifikátor v závorce (zlepšuje přesnost)
- poskytuje vysvětlující poznámky
- zobrazuje vztahy – hierarchické, asociace, ekvivalence – využití při specifikaci či zobecnění dotazu
- vyjadřuje termíny, které nejsou obsaženy v záznamu

# Selekční jazyky

Usnadňují vyhledávání tím, že:

- odstraňuje problémy se syntaxí

Dokument je reprezentován těmito slovy  
v přirozeném jazyku:

- automobily, export, Spojené státy americké, Japonsko

Možné významy

- export japonských automobilů do USA
- export amerických automobilů do Japonska

Řešení v tezaurech – využití rolí

Řešení pomocí PH – dán kontext, hledání pomocí fráze

# Selekční jazyky

Při vyhodnocování relevantnosti výsledků vyhledávání (řazení vyhledaných záznamů) mají selekční jazyky větší váhu než slova přirozeného jazyka

## PROČ?

- Termín SJ byl přiřazen dokumentu na základě obsahové analýzy, z toho plyne indexace/postižení významného tématu, a to je pro vyhodnocení dotazu relevantnější

# Selekční jazyky – využití

Využití při taktikách:

- Zúžení dotazu:
  - klíčová slova se kombinují s věcným selekčním jazykem
  - kombinace množiny deskriptorů/hesel s podřazenými klíčovými slovy
- Rozšíření dotazu:
  - dodatečné uvedení širších jednotek věcného SJ, tj. těch, které jsou nadřazeny použitým termínům (deskriptorům, předmětovým heslům) – ty naleznete v příslušných řízených slovnících
  - uvedení jednotek věcného SJ jako klíčových slov (např. vyhledávání ve všech polích)

# SJ- slabé stránky

- nedostatek specifičnosti - např. „víceslovné předložky“
- není okamžitá aktualizace – časová prodleva než je termín zahrnut
- některá témata mohou být při indexování opomenuta
- slova autora mohou být nesprávně interpretovaná – nepochopení látky

# SJ- slabé stránky

- chyby v indexaci zapříčiňují ztráty
- řešeršéri se musí učit selekční jazyk
- nekompatibilita – znesnadnění paralel. vyhledávání, bariéra snadné výměny
- časové ztráty související s tvorbou, údržbou a osvojením si SJ

# Indexátor vs. řešeršér

- Uživatel popisuje něco, co nezná X indexátor má dokument v ruce, „všechno je před ním“
- Indexátor by měl zkoušet předvídat:
  - podle jakých termínů budou vyhledávat uživatelé
  - jakou informaci jim daný dokument poskytne, že povede k uspokojení jejich informační potřeby?
- porozumění tématu, chápání významu slov



# Indexátor vs. řešeršér

- Indexátoři neindexují dokumenty takovým způsobem, aby zachytili nekonečně mnoho rozmanitých dotazů.
- Většinou jsou indexována hlavní a dílčí témata, tj. what is in the record.
- Nekonečně mnoho dotazů může být uspokojeno dokumentem.
- Jde o úhel pohledu - document-oriented approach x user-centered indexing

# Příklady

- požadavek: články týkající se vztahu knihoven a Webu 2.0
  - formulace dotazu: řešerši uskutečněte pomocí předmětového hesla/hesla z hesláře - (tj. v Subject) db LLIS
- Jakými jinými tematickými autoritami byste nahradili chybný termín organizace poznání/pořádání informací
- Jakými jinými tematickými autoritami byste nahradili chybný termín systém správy obsahu/redakční systém
- Naleznete v katalogu MU dokumenty pojednávající o postavení žen v české společnosti (pomoci SVA)
- Naleznete v katalogu MU dokumenty vztahující se k odívání, módě

# Přirozený jazyk – výhody

- vysoká specifičnost ovlivňuje pozitivně přesnost - např. vlastní jména (osob, institucí apod.)
- schopnost vyčerpávajícím způsobem pokrýt téma, zvyšuje úplnost - neplatí u neanotovaných záznamů, zejména tam, kde je zahrnut abstrakt a plný text
- aktualizace – nové termíny jsou okamžitě dostupné
- slova užitá autorem – nemůže dojít k dezinterpretaci indexátorem
- snadnější výměna materiálu mezi databázemi – jazyková neslučitelnost odstraněna
- není třeba se jazyku učit (rodilý mluvčí)

# PJ- slabé stránky

- intelektuální úsilí řešeršéra – problém souvisící se synonymy (formulace dílčích dotazů) a homonymy (nutnost uvedení do kontextu)
- problémy se syntaxí – nesprávné spojení termínů, asociace – řešení pomocí proximitních operátorů
- schopnost vyčerpávajícím způsobem pokrýt téma může vést ke ztrátě přesnosti
- odlišná terminologie u jednotlivých autorů

# Povinná literatura na příště

Houdek, Petr; Schwarz, Josef; Snášel, Václav. *Moderní metody vyhledávání dokumentů v rozsáhlých plnotextových databázích : příklad vektorového modelu.* [www.ikaros.cz/node/1276](http://www.ikaros.cz/node/1276)

Harald Reiterer, Thomas M. Mann, Gabriela Mußler - *Visual Information Retrieval for the WWW.* – studijní materiály na ISu

Panagiotis Petratos. *Information Retrieval Systems: A Human Centered Approach.* – studijní materiály na ISu