



VIKMA06

Rešeršní a studijně rozborová činnost

16. 3. 2012: Přednáška P4+K2: Metodika vyhledávání, modely vyhledávání

FF MU, jaro 2012

Mgr. Josef Schwarz
126172@mail.muni.cz



Metodika vyhledávání

- typy vyhledávání
- nástroje vyhledávání
- analýza informačního požadavku a formulace řešeršního dotazu

Typy vyhledávání

○ Podle hledané informace:

● identifikační vyhledávání

- (známe některé údaje o hledaném dokumentu nebo položce)
- vyhledávací výrazy: formální údaje - osobní jméno, název, nakladatel, rok, místo vydání, název časopisu, ISBN, ISSN, datum (konání konference, vydání, narození aj.) apod.
- příklad: [NTK](#), [telefonní seznam](#), [Obchodní rejstřík](#)

● věcné vyhledávání

- (neznáme požadovaný dokument, hledáme určité téma)
- vyhledávací výrazy: věcné údaje - klíčová slova z názvu, předmětová hesla, klíčová slova, deskriptory tezauru, klíčová slova z textu dokumentu (redukovaného nebo plného), klasifikace ([MDT](#), [OKEČ](#), [NAICS](#)) apod.

● faktografické

- (chceme zjistit konkrétní informaci)
- vyhledávací výrazy: údaje podle obsahu a struktury zdroje



Typy vyhledávání

- hledání (seeking)
- vyhledávání (searching)
- prohlížení (browsing)
- filtrace (filtering)
- data mining



Typy vyhledávání

- nestrukturované (freetextové)
 - celý záznam dokumentu
- strukturované
 - metadata
 - selekční obraz dokumentu
 - redukovaný text
 - vazby dokumentů
 - citační vazby
 - formální vazby (FRBR)
- plnotextové



Typy vyhledávání

- nestrukturované vyhledávání
 - základní, jednoduché vyhledávání
 - [KNAV](#), [PubMed](#), [Google](#), [Scirus](#)
- strukturované vyhledávání
 - pokročilé, podrobné vyhledávání
 - [KNAV](#), [PubMed](#), [Google](#), [Scirus](#)
 - řízený slovník (tezaurus, seznam předmětových hesel nebo klíčových slov apod.)
 - není dostupný: [KNAV](#)
 - je dostupný samostatně: [NTK](#)
 - je dostupný při vyhledávání: [HAMU](#), [UPM](#)
 - je plně integrovaný do vyhledávání: [PSP](#)
- plnotextové (fulltextové) vyhledávání
 - invertovaný rejstřík
 - sekvenční vyhledávání



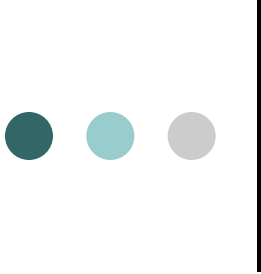
Nástroje vyhledávání

- vyhledávací (dotazovací) jazyky
 - složky
 - standardizace (CCL)
 - tendence ke (kvazi)přirozenému jazyku
- selekční jazyky
 - věcné
 - identifikační (authority)
 - sémantické sítě
- uživatelské rozhraní
 - příkazový řádek
 - GUI
- algoritmy vyhledávání



Formulace rešeršního dotazu

1. pojmová analýza
2. synonyma a související pojmy
3. převedení na výrazy řízeného slovníku
4. aplikace (booleovských) operátorů
5. aplikace dalších vyhledávacích technik



Pojmová analýza

- identifikace klíčových pojmů
- reprezentace pojmů (substantiva a adjektiva, slovesa nahrazena operátory)



Synonyma a související pojmy

- vytvoření seznamu synonym a dalších příbuzných výrazů
- využití seznamu:
 - výběru vhodného vyhledávacího výrazu
 - převod na výraz věcného SJ
 - rozšiřování a zužování tématu



Převedení na výrazy řízeného slovníku

Varianty

1. výraz v seznamu je shodný s výrazem ŘS
2. výraz v seznamu je synonymem/ekvivalentem výrazu ŘS
3. pro výraz v seznamu existuje pouze širší výraz ŘS
4. pro výraz v seznamu existují pouze specifičtější/podřazené výrazy ŘS



Aplikace (booleovských) operátorů

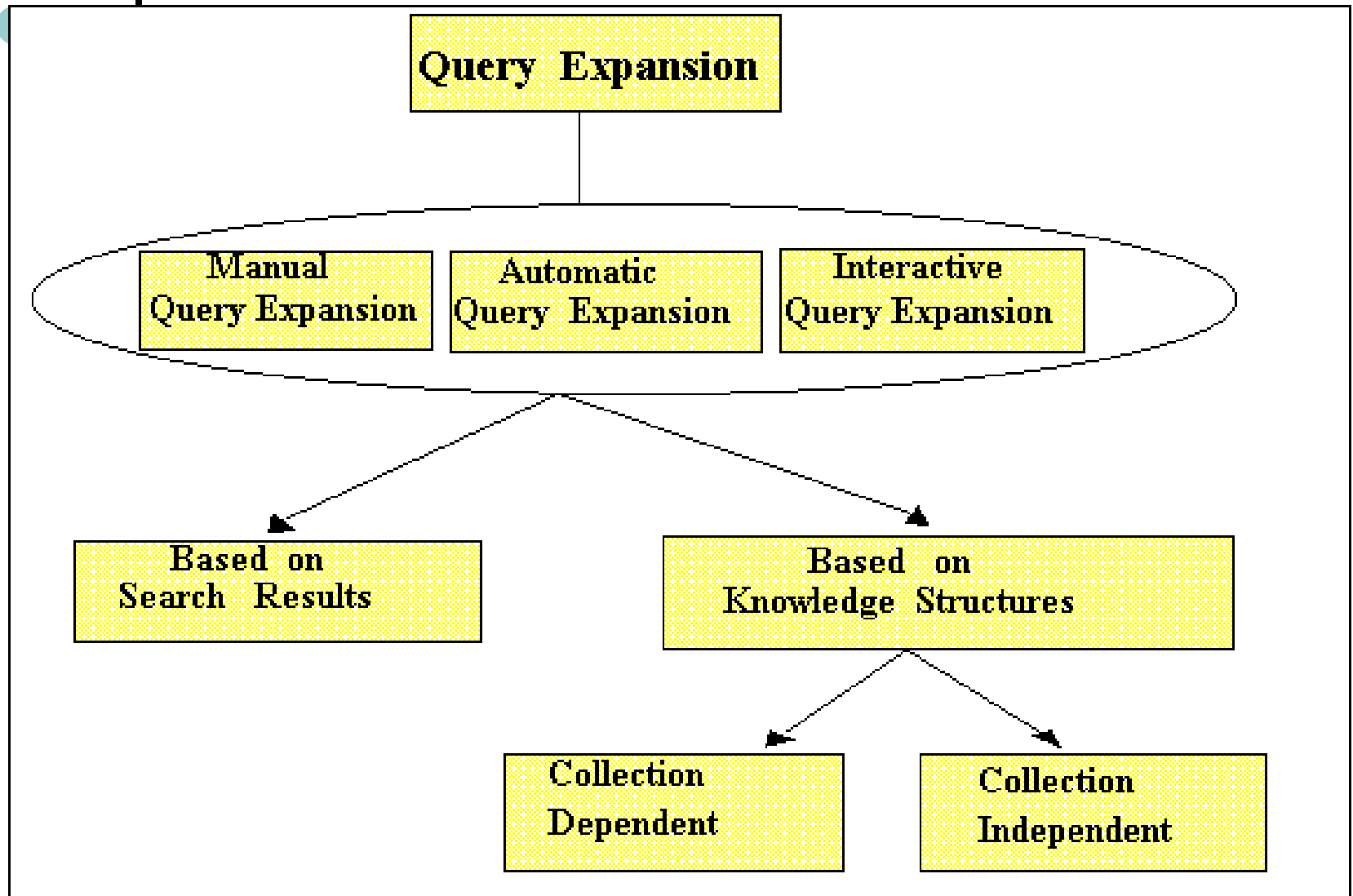
Určení vztahů mezi pojmy

- **operátor AND** – spojení významově odlišných výrazů
- **operátor OR** – spojení synonym a příbuzných výrazů
- **operátor NOT** – vyloučení nežádoucích výrazů



Aplikace dalších vyhledávacích technik

- škála možností závisející na konkrétním informačním zdroji
 - krácení, zástupné znaky
 - proximitní operátory
 - vyhledávání podle polí
 - rozšiřování a úprava dotazu (query expansion – relevance feedback)
 - vyhledávání ve více databázích (multiple database searching)





Vyhledávací techniky

obvyklé možnosti

- booleovské operátory
- fráze
- vyhledávání podle polí
- formální omezení
- krácení, zást. znaky, stemming
- ukládání rešerše a historie
- proximitní vyhledávání
- užití řízených slovníků

specifické možnosti

- prohlížení časopisů a obsahů jednotlivých titulů
- rozšiřování dotazu
- navrhování výrazů ŘS
- dotaz příkladem
- automatický překlad
- odkazy na plný text prostřednictvím jiné služby, odkazy na web, napojení na katalog
- vyhledávání pomocí notací SSJ