



VIKBA11

## **Selekční jazyky**

3. 11. 2017: Sémantická redukce  
dokumentů I (P7+K3)

FF MU, podzim 2017

Mgr. Josef Schwarz  
[126172@mail.muni.cz](mailto:126172@mail.muni.cz)

# Úkol č. 5

- dopravní pravidla [dopravní provoz]
  - znamení „stůj“ [signalizační zařízení]
- dopravní provoz [dopravní pravidla]
  - železniční nehoda (neštěstí na železnici) [železniční doprava] [železniční přejezd]
- městská hromadná doprava (MHD) [tramvajová doprava] [osobní železniční doprava]
- křižovatka [semafor]
  - kruhový objezd (křižovatka typu kruhového objezdu)
  - mimoúrovňová křižovatka
  - železniční přejezd (křížení silnice – železnice) [železniční nehoda]
- signalizační zařízení [znamení „stůj“]
  - semafor [křižovatka]
  - návěstidlo [železniční doprava]
- tramvaj [tramvajová doprava]
  - Tatra T3
- trolejová doprava [železniční doprava]
  - tramvajová doprava [tramvaj] [městská hromadná doprava]
- železniční doprava [železniční nehoda] [návěstidlo] [trolejová doprava]
  - osobní železniční doprava [městská hromadná doprava]
- Pozn.:
  - odrážka = podřazený pojem (např. kruhový objezd je podřazen křižovatce)
  - kulaté závorky = ekvivalent
  - hranaté závorky = asociovaný pojem

# Sémantická redukce dokumentů



## ○ účel

- reprezentace plného textu dokumentu
- možnost vyhledávání podle obsahu

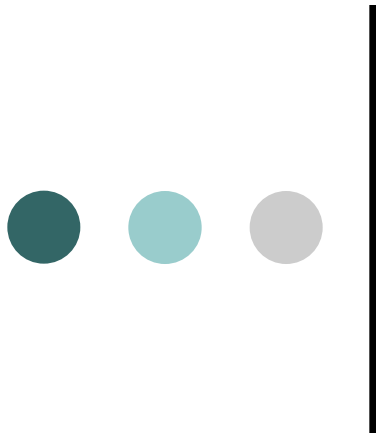
## ○ procesy

- obsahová analýza
- indexace, klasifikace
- abstrahování/referování

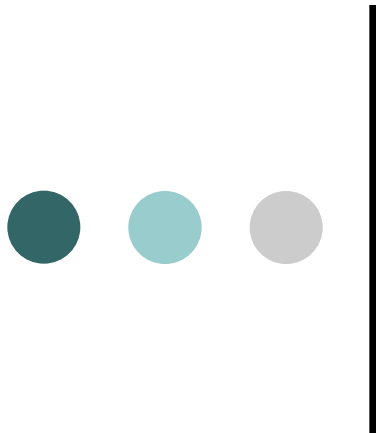
## ○ nástroje

- intelekt/algoritmus
- SJ/PJ

# Obecný proces věcného zprac.

- 
- Obsahová analýza
  - Identifikace pojmů
  - Výběr znaků SJ a tvorba selekčního obrazu dokumentu (SOD)
  
  - *ČSN ISO 5963. Dokumentace. Metody analýzy dokumentů, určování jejich obsahu a výběru lexikálních jednotek selekčního jazyka. Úč. 1.4.1996. 12 s.*
  - pravidla (indexační pravidla)

# Obsahová analýza

- 
- *content/subject analysis*
  - určení obsahu dokumentu na základě:
    - plného textu
      - (orientační, kurzorické/diagonální, selektivní, statorické, racionální čtení – významy viz TDKIV)
    - redukovaného textu
      - abstrakt, resumé, výtah, anotace
    - dalších adekvátních částí
      - název, obsah, rejstřík, úvod, závěr, zvýrazněné části textu apod.
  - zásadní význam pro kvalitu SOD
    - porozumění textu
      - jazykové
      - odborné
    - pečlivost

# Identifikace pojmů



- východiska

- zohlednění funkce IS a potřeb uživatele
- zohlednění významu a rozsahu dokumentu
- zohlednění všech adekvátních obsahových hledisek
- stanovení důležitosti pojmů/témat
  - hlavní pojmy/témata
  - vedlejší pojmy/témata

- vyjádření

- klíčová slova
- přirozený jazyk

# Výběr znaků SJ

- indexace/klasifikace (PSJ/SSJ)
- předpoklad: znalost konkrétního SJ
- úplnost indexace/klasifikace
  - rozsah zachycení hlavních a vedlejších témat
  - počet znaků SJ
- specifičnost indexace/klasifikace
  - výběr nejspecifičtějšího znaku SJ
- indexační hlediska
  - např. předmět/téma, čas, místo, entita, proces/činnost, vztah
- prekoordinace/postkoordinace
- subjektivní aspekty
- indexační pravidla ([příklad: ČTT/UK-ETF](#))

# Specifičnost indexace (příklad)



parlament

PD1 složení parlamentu

PD2 komory parlamentu

PD3 jednokomorový systém

PD3 dvoukomorový systém

PD4 horní komora

PD4 dolní komora

PD4 federální sněmovna



# Indexace krásné literatury



- důvody indexace

- velká část fondu a výpůjček (veřejné knihovny)
- část fondu odborných knihoven

- možné uživatelské dotazy:

- beletrizovaný životopis van Gogha
- sci-fi odehrávající se na Měsíci
- román situovaný do italského kláštera odehrávající se ve 14. století

- příklady indexace

- Eco: Jméno růže
  - NK, KJM, MZK: italské romány \* detektivní romány \* historické romány
  - KBBB: historie - františkáni - Vilém z Baskervillu (františkáni - Vilém z Baskervillu - František z Assisi – benediktini)
  - KMO: italská tvorba, 14.stol., benediktini, kláštery, detektivky, historické příběhy
  - MK Tábor: detektivky - historické romány - středověk - teologie - 14. stol. – Itálie (historické romány)

# Indexace krásné literatury



- aspekty indexace

- předmět

- akce, události, témata
- psychologický vývoj postav, motivy
- sociální vztahy

- rámec

- doba (minulost, současnost, budoucnost)
- místo (geografie, sociální prostředí, profese)

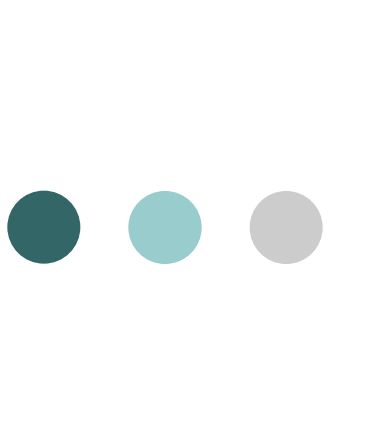
- autorský záměr

- emocionální působení
- poznání

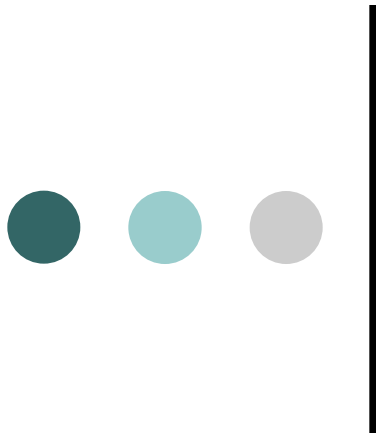
- přístupnost

- čtivost
- fyzické charakteristiky (velké písmo apod.)
- literární forma (žánr)

# Indexace krásné literatury

- 
- ALA: *Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc. (1990)*
  - Britská národní bibliografie
    - indexace krásné literatury od r. 1997 dle pravidel ALA

# Indexace krásné literatury

- 
- problémy
    - obtížnější indexace než u odborné literatury – problém identifikace základních obsahových charakteristik a pojmů
    - nepřítomnost pomocného aparátu (obsah, rejstřík)
    - vyšší subjektivita
    - nízká konzistence
    - nedostatek specializovaných řízených slovníků