



VIKBA11

Selekční jazyky

9. 11. 2018: Sémantická redukce
dokumentů I (P7)


FF MU, podzim 2018

Mgr. Josef Schwarz
126172@mail.muni.cz

Zvolená témata tezaurů

- astrofyzika
- genetika
- právní teorie
- středoškolská matematika
- počítačová věda a softwarové inženýrství
- počítačové sítě
- informační bezpečnost (2x)
- informační gramotnost
- informační vzdělávání
- ropný průmysl
- pěstitelství
- letadla
- savci
- kočky
- psi
- filmová teorie a tvorba
- filmový průmysl
- fotografie (fotografování)
- fotografování
- umělecká malba
- rocková hudba
- zpěv
- hry
- biatlon
- jezdeckví
- LARP (live action role playing)
- britská kuchyně
- vietnamská kuchyně
- baristika
- vinařství
- cukrářské výrobky
-
- zdravý životní styl
- oděvy (2x)
- alergie
- domácnost
- nábytek
- pokojové rostliny
- vojenská technika (2. světová válka)
- vznik Československa
-

Úkol č. 4 – zadání A

- 
- V následující množině výrazů stanovte preferovanou lexikální jednotku. Při volbě preferované jednotky se řiďte normou ČSN 01 0193, kap. 6.5 (viz příloha úkolu). Zdůvodněte výběr preferované lexikální jednotky a stanovte typy nepreferovaných lexikálních jednotek. Které z uvedených výrazů jsou synonymní, které kvazisynonymní a které slangové?

-
- drogová závislost
- fetování
- kouření
- morfinismus
- narkomanie
- smažba
- tabakismus
- toxikomanie
- užívání drog
- workoholismus
-

Úkol č. 4 – řešení A



- **drogová závislost**

- (S)
- toxikomanie
- narkomanie
- užívání drog

- (K)
- kouření
- Morfinismus
- tabakismus
- workoholismus


- (A)
- fetování
- smažba

(S) = synonymum

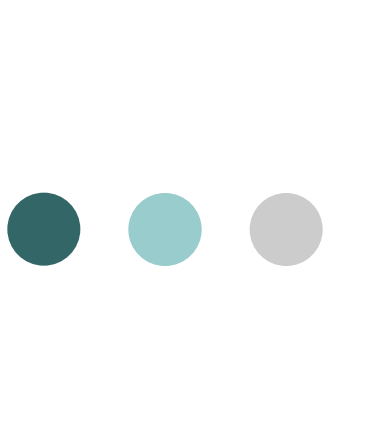
(K) = kvazisyonymum

(A) = slang

Úkol č. 4 – zadání B

- 
- V následující množině výrazů stanovte, která sousloví a na jaké jednotky budou rozložena při tvorbě tezauru (a která naopak zůstanou v původní podobě). Řiďte se normou ČSN 01 0193, kap. 7, a metodickým materiálem AAT Compound Term Rules (oboje viz příloha úkolu). U každého sousloví stanovte typ podle normy (např. 7.3.2.1a) i podle uvedené metodiky AAT (např. 1.2).
 -
 - barokní kostel
 - Maxwellův démon
 - brněnská univerzitní knihovna
 - česání chmele
 - čokoládové vejce
 - francouzská zahrada
 - klíč od bytu
 - krkonošské vodopády
 - sadba raných brambor
 - tepelné čerpadlo
 - výroba karlovarského porcelánu
 - vzdělávání dospělých

Úkol č. 4 – řešení B

- 
- PREKOORDINACE
 - Maxwellův démon (7.2.2a; 1.3)
 - čokoládové vejce (7.2.2e; 2.1)
 - francouzská zahrada (7.2.1b; 2.2)
 - tepelné čerpadlo (7.2.1b; 2.4)
 - POSTKOORDINACE
 - barokní kostel (7.3.2.1a; 1) = baroko + kostel
 - brněnská univerzitní knihovna (7.3.2.1a; 4.1) = Brno + univerzitní knihovna (nebo: univerzita + knihovna)
 - česání chmele (7.3.2.2a; 2.3) = česání + chmel
 - klíč od bytu (7.3.2.1a; 4.1) = klíč + byt
 - krkonošské vodopády (7.3.2.1a; 4.1) = Krkonoše + vodopády
 - sadba raných brambor (7.3.2.3a; 2.3) = sadba+rané brambory
 - výroba karlovarského porcelánu (7.3.2.2a+7.3.2.1a; 2.3+1) = výroba + Karlovy Vary + porcelán (možno i: karlovarský porcelán)
 - vzdělávání dospělých (7.3.2.3a; 4.3) = vzdělávání + dospělí (ve specializovaném pedagogickém tezauru by se pravděpodobně nerozkládalo, jednak z důvodu velkého počtu dokumentů k danému tématu, jednak z důvodu možných vazeb k dalším pojmům (např. andragogika jako obor zabývající se vzděláváním dospělých))

Úkol č. 4 – zadání a řešení C

- Uvedte alespoň jedno sousloví, které obsahuje synkategorematické substantivum. Vyloučeny jsou příklady uvedené v prezentaci (tj. *čokoládová vejce*, *světelný rok*, *umělé květiny*).

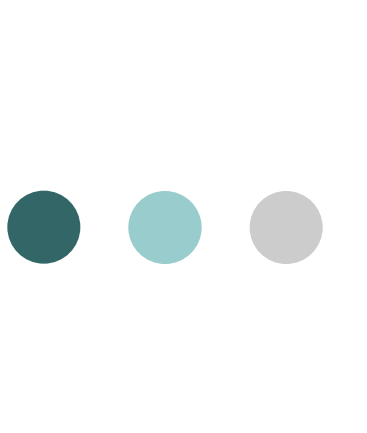
Synkategorematickými substantivy jsou:

- čokoládový zajíc
- hroší kůže
- papírová růže
- tekuté řetězy

Synkategorematickými substantivy nejsou:

- hvězdná brána (metaforické označení)
- kuřecí bábovka (i to je bábovka, byť není sladká)
- lávová lampa (dekorační předmět, jedna z funkcí je ovšem svícení)
- modrý portugal (označení vína)
- nealkoholické pivo (i toto pivo je nápojem)
- obilná káva (i tato káva je nápojem)
- studená fronta (polysém – fronta = řada / fronta = bojová linie / fronta = meteorologický jev)
- umělý kanál (i umělý kanál je kanálem)
- vést válku (není vůbec substantivem)


Úkol č. 5 – zadání

- 
- dopravní pravidla
 - dopravní provoz
 - kruhový objezd
 - křížení silnice – železnice
 - křižovatka
 - křižovatka typu kruhového objezdu
 - městská hromadná doprava
 - MHD
 - mimoúrovňová křižovatka
 - návěstidlo
 - neštěstí na železnici
 - osobní železniční doprava
 - semafor
 - signalizační zařízení
 - Tatra T3
 - tramvaj
 - tramvajová doprava
 - trolejová doprava
 - znamení „stůj“
 - železniční doprava
 - železniční nehoda
 - železniční přejezd

Úkol č. 5 – řešení - fazety

- TYPY DOPRAVY
- městská hromadná doprava
- MHD
- osobní železniční doprava
- tramvajová doprava
- trolejová doprava
- železniční doprava
- DOPRAVNÍ INFRASTRUKTURA
- kruhový objezd
- křížení silnice – železnice
- křižovatka
- křižovatka typu kruhového objezdu
- mimoúrovňová křižovatka
- návěstidlo
- semafor
- signalizační zařízení
- železniční přejezd
- DOPRAVNÍ PROSTŘEDKY
- Tatra T3
- tramvaj
- DOPRAVNÍ PŘEDPISY
- dopravní pravidla
- znamení „stůj“
- DOPRAVNÍ SITUACE
- dopravní provoz
- neštěstí na železnici
- železniční nehoda

Úkol č. 5 – řešení - struktura

- 
- 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]

Odražené výrazy = podřazené výrazy

Výrazy v kulaté závorce = ekvivalenty

Výrazy v hranaté závorce = příbuzné výrazy

Sémantická redukce dokumentů



○ účel

- reprezentace plného textu dokumentu
- možnost vyhledávání podle obsahu

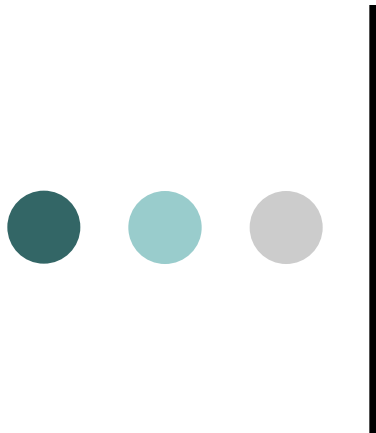
○ procesy

- obsahová analýza
- indexace, klasifikace
- abstrahování/referování

○ nástroje

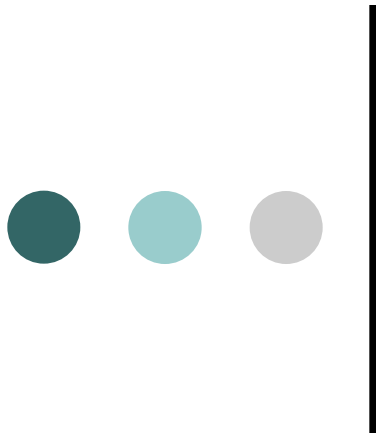
- intelekt/algoritmus
- SJ/PJ

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)

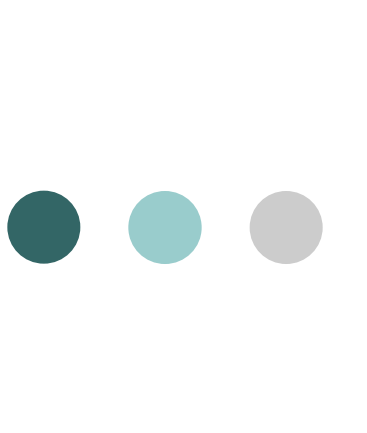
- autorský záměr

- 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