

Předmětové selekční jazyky (PSJ)

Předmět: Selekční jazyky

30. 10. 2009

Přednášející: Mgr. Silvie Kořínková Presová

<http://kisk.phil.muni.cz/mgr-silvie-korinkova-presova-dis>

Varianty termínu

používané české ekvivalenty - věcné authority, systémy heslování, předmětová pořádací soustava, verbální pořádací systémy, předmětové třídění (nevhodné označení) apod.

anglické ekv. - subject retrieval language, verbal indexing languages, controlled vocabulary, vocabulary control tool

Výklad termínu

Předmětový selekční jazyk (výklad z TDKIV)
Selekční jazyk založený na abecedně
uspořádaném systému hesel, popř.
lexikálních jednotek.

Podle charakteru lexikálních jednotek se
rozlišují předmětové selekční jazyky
založené na použití klíčových slov (např.
klíčová slova z názvu dokumentů,
rotované rejstříky), předmětová hesla
(jazyk předmětových hesel) a
deskriptorové selekční jazyky.

Výklad termínu

Předmětový SJ je SJ používaný při realizaci procesu předmětového pořádání informací jako jeho výrazový a vyjadřovací prostředek. (Kovář, 1984, s. 129)

Předmětové pořádání informací je proces, při kterém slovně vyjádřený obsah, téma či předmět dokumentu, jeho částí nebo jednotlivých v něm obsažených informací je převáděn do jednou více, jindy méně normalizovaných a formalizovaných hesel (deskriptorů), která jsou ve slovnících uspořádána abecedně. (vcelně podle Kovář)

Výklad termínu

Controlled vocabulary - Seznam nebo databáze věcných termínů, kdy všechny termíny nebo fráze reprezentující pojem jsou sdruženy dohromady. Preferované lexik. jednotky jsou navrženy pro užití v záznamech o dokumentech v rešeršních systémech. (volně podle Taylor, 2004, s. 361)

Vocabulary control tool - Organizovaný seznam termínů nebo frází, který může být použit k indexování nebo prohledávání sbírky pomocí věcných termínů či frází. (Chowdhury, 2007, s. XVIII)

Druhy předmětových SJ

prekoordinované PSJ - Vyjadřují téma dokumentu nebo dotazu souborem slov přirozeného jazyka svázaných podle pevně stanovených pravidel a dávajících smysl jen jako celek.

postkoordinované PSJ - Téma dokumentu nebo dotazu je vyjádřeno také souborem slov přirozeného jazyka, avšak tento soubor se skládá z původně vzájemně izolovaných a samostatných slov a sousloví, která byla vybrána z řízeného slovníku.

(Kovář, 1984, s. 132 - 133)

NOVÁK, Blahoslav. Projevuje se snižování hektarových výnosů chmele a spotřeby piva u nás na rozloze vinic a výrobě vína v Rakousku? Šťastný vinař, 2009, roč. 1, č. 2, s. 10-22.

Prekoordinace / postkoordinace



prekoordinované PSJ - předmětová hesla
(jazyk předmětových hesel, PSJ typu
předmětových hesel)

postkoordinované PSJ - deskriptorové SJ
(PSJ deskriptorového typu), volně
tvořená klíčová slova, PSJ založené na
použití slov z názvu dokumentů -
permutované (rotované) rejstříky (KWIC,
KWOC)

Předmětová hesla

Jazyk předmětových hesel představuje nástroj, ve kterém je téma dokumentu vyjádřeno sestavou lexikálních jednotek podle předem stanovených syntagmatických a syntaktických pravidel už v průběhu indexování.
(Balíková, 2001)

použití - předmětová hesla jako prvky řazení v předmětových katalogích a kartotékách

<http://www.lib.cas.cz/katalogy/predmetovy->

Předmětová hesla

Pořádací znak - **předmětové heslo** (PH) -
Předmětové heslo je jednoslovné nebo
víceslovné formálně upravené vyjádření
obsahu (resp. tematiky), popř. i
charakteristiky formy dokumentu.
- **předmětový prvek**

Předmětové heslo se skládá z jednoho
nebo několika předmětových prvků -
jednočlenné či vícečlenné PH (řetězec
PH)

Schéma PH

Heslo (ev. **doplněk hesla**) - **1. podheslo** (ev. doplněk podhesla) - **2. podheslo** (ev. doplněk hesla) - n. podheslo (ev. doplněk podhesla)

Letadla - **podvozky** - **příručky**

Cukrovka (řepa) - **cukernatost**

Francie - **Prusko** - **války** - r. 1870-1871

Heslo - Vyjadřuje se jím pojem, který má pro obsahovou charakteristiku dokumentu rozhodující význam - hlavní téma.

Podheslo upřesňuje heslo obsahově a podle potřeby i formálně.

Doplňky hesla n. podhesla vysvětlují stejně znějící,

Slovník

Předmětový heslář - Seznam prvků
předmětových hesel užitých při zpracování
konkrétního předmětového katalogu či
předmětové kartotéky (Pinkas, s. 95).
Nejen seznam prvků, ale též celých řetězců
předmětových hesel.

Struktura

částečná hierarchie

přidružovací odkazy (viz též) - klasifikace viz též
indexace

vylučovací odkazy (viz) - např. klasifikace viz
kategorizace

Syntaxe

Syntaktická pravidla - uspořádání prvků
předmětového hesla/kombinace prvků
předmětového hesla

Princip prekoordinace

Výhody/Nevýhody (Balíková 2001)

Výhody

detailní vyjádření předmětu dokumentu
odpovídající specifičnost
maximální informační hodnota řetězce předmětového hesla
efektivní servis pro uživatele v tradičním prostředí

Nevýhody

rozklad víceslovných lexikálních jednotek (např. výkon vazby: vazba - výkon)
délka řetězce předmětového hesla
princip prekoordinace uplatněný v syntaktické rovině, a z toho pramenící:
komplikovaná pravidla aplikační syntaxe
redundantnost informací v bibliografickém záznamu
rozsáhlost autoritního souboru

Příklady

Library of Congress Subject Headings (LCSH) - nejrozšířenější univerzální PSJ pro využití v lístkových a online katalozích

<http://authorities.loc.gov/>

Library Literature & Information Science Fulltext <http://library.muni.cz/ezdroje/>
prohlížení témat v databázi vyjádřených pomocí předmětových hesel

Deskriptorový selekční jazyk (DSJ)

DSJ (výklad z TDKIV) - **Postkoordinovaný předmětový selekční jazyk** určený pro indexaci a vyhledávání prostřednictvím řízeného slovníku označovaného jako **tezaurus**, jehož lexikum je složeno z deskriptorů a nedeskriptorů.

Historický vývoj od DSJ založeného na unitermech (unitermů) k DSJ založenému na deskriptorech (tezaurům), jejichž forma a vztahy jsou standardizovány

Tezaurus

Řízený a měnitelný slovník deskriptorového selekčního jazyka uspořádaný tak, že explicitně zachycuje apriorní (paradigmatické) vztahy mezi lexikálními jednotkami.

Podle jazykového zaměření se rozlišuje jednojazyčný a vícejazyčný tezaurus, podle šíře tematického zaměření speciální, polytematický a univerzální tezaurus, podle struktury fasetový a tematický tezaurus, podle funkce tradiční, indexační a vyhledávací tezaurus. Dále se vyčleňuje cílový a zdrojový tezaurus.

Výklad termínu

Thesaurus - slovník PSJ, formálně organizovaný takovým způsobem, aby explicitně zachycoval apriorní (paradigmatické) vztahy mezi pojmy.
(Aichison, 2000, s. 1)

Tvorba tezauru

ČSN 010193. Dokumentace : pokyny pro vypracování a rozvíjení jednojazyčných tezaurů

Techniky tvorby tezaurů vycházející z obecných principů.

Smysl - jednotná praxe v rámci IS či více spolupracujících IS.

ČSN 01 0172 (ISO 5964). Pokyny pro vypracování a rozvíjení vícejazyčných tezaurů.

Softwarové nástroje pro správu tezaurů (thesaurus management software)

Slouží pro tvorbu, aktualizaci, editování
tezaurů.

např. MultiTes - Lze stáhnout pro zkušební
účely jako Trial Software.

<http://www.multites.com/download.htm>

Základní termíny

lexikální jednotka (LJ)

deskriptor - preferovaný termín - LJ
užívaná závazně při indexování.

nedeskriptor - nepreferovaný termín -
ekvivalent nebo kvaziekvivalent
preferovaného termínu

deskriptorový odstavec (výklad z TDKIV) -

Součást hlavní části tezauru zahrnující
deskriptor a všechny relevantní informace
uvedené v poznámkovém a odkazovém
aparátu jako je definice, vysvětlující
poznámka, poznámka o použití,
nedeskripty, podřazené, nadřazené

Pojmy vyjádřené LJ patří k následujícím obecným kategoriím:

konkrétní entity

předměty, věci a jejich části (počítač, pes, lebka)
materiály (plast, dřevo, bavlna)

abstraktní entity

činnosti a události (psaní, dýchání, plavání)
abstraktní entity a vlastnosti věcí, materiálů n. činností
(síla, přesnost, viskozita)

obory n. vědní disciplíny (sociologie, psychologie,
informační věda)

měřicí jednotky (kilogram, metr, ampér)

individuální jednotky (identifikátory)

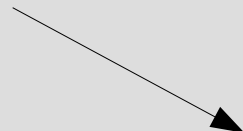
geografická jména, jména osob, jména organizací,
jména výrobků

Tvar LJ

substantivum n. substantivní slovní spojení

adjektivní

např. národnostní menšina



modifikátor (adjektivum) základ
(substantivum)

s předložkovou vazbou (např. kurzy pro
pokročilé, školy pro nevidomé děti)

adjektiva, příslovce, slovesa

obvykle se nezařazují, činnost - vyjádření
substantivy

Tvar LJ

plurál:

počítatelná substantiva (knihy, okny, ryby)

Ale části těla singulár, pokud je v těle jedna část (ústa, hlava).

singulár

nepočítatelná substantiva (mouka, písek, pára)

Ale třídy s více členy pl. (cukry, jedy, mouky)

abstraktní pojmy (osobnost, vodivost, kapitalismus)

Ale třídy s více členy pl. (přírodní vědy, chemické reakce)

Vyjádření odlišných pojmů pomocí sg./pl., např. cukr (potravinářský průmysl)/cukry (chemie)

homografy, polysémy - relátor v závorce, součástí LJ

Vysvětlující poznámky a definice

omezení významu LJ pro účely daného
systému (vyloučení dalších významů)

Další informace:

datum zařazení LJ

zdroj LJ (zejména u nových slov)

pokyny pro indexátory, např. možnost
kombinace LJ (např. instituce EU - Uvedte
ve spojení s danou institucí)

Sousloví

Obecně - LJ mají vyjadřovat jednoduché pojmy, sousloví je třeba rozložit na jednodušší prvky, pokud rozklad nezpůsobí potíže uživateli.

Sémantický rozklad - výraz, který vyjadřuje komplexní pojem, je převeden na jednodušší, tzv. definiční prvky. **Použití této techniky se v tezaurech nedoporučuje.**

teploměr

teplota + měření + přístroje

Syntaktický rozklad - tato technika se uplatňuje u sousloví, tj. u výrazů, které umožňují morfologickou analýzu na dílčí složky.

tibetský buddhismus

Sousloví

*Kritéria pro to, zda sousloví ponechat beze změny
či sousloví rozložit na syntaktické složky.*

Je nutné rozlišit jednotlivé části sousloví z hlediska
jejich rolí a vztahů:

základ neboli řídící člen

lisování vína

lisování

Substantivum, které vyjadřuje širší třídu věcí nebo
jevů, jejímž členem je lexikální jednotka jako
celek.

modifikátor

vína

Část sousloví, která odkazuje k určité
charakteristice nebo rozlišujícímu znaku.

Po připojení k danému základu zužuje jeho
významový nebo stylistický odstín a vyjadřuje
tak jednu z jeho podtříd.

Sousloví ponechaná beze změny

sousloví je běžné a užívané (informační zdroje, pořádání informací)

rozklad vede ke ztrátě významu (filozofie dějin, library science)

sousloví obsahuje vlastní jméno (Bradfordův zákon, Lotkův zákon)

modifikátor ztratil původní význam (lehký průmysl)

modifikátor bez přímého vztahu (stromová struktura)

sousloví se „středním členem“ (naftové motory)

synkategorematická substantiva (umělé květiny, čokoládová vejce, světelný rok) - Podstatná jména, která sama o sobě nevyjadřují třídu pojmů, k níž odkazují slovní spojení jako celek,

Sousloví rozložená na syntaktické složky

1.

základ - vlastnost/část (včetně materiálů),
modifikátor - celek, popř. nositel dané
vlastnosti nebo části

např. letištní hala letiště + hala
 lodní motory lodě + motory

2.

základ - činnost, **modifikátor -**
objekt/vykonavatel činnosti

např. mražení zeleniny mražení +
 zelenina
 stěhování ptáků stěhování +
 ptáci

Základní vztahy v tezauru

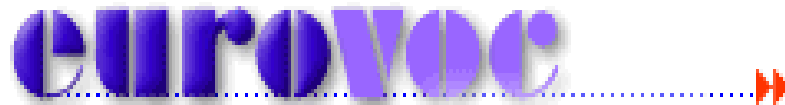
vztah ekvivalence

vztah hierarchie

vztah asociace

Vztah ekvivalence

Do vztahu ekvivalence jsou uváděny **synonymní lexikální jednotky**, tj. termíny, které se liší formou, ale jejichž obsah je identický (označují stejný denotát).



např. tezaurus

<http://europa.eu/eurovoc/>

práva žen UF ženská práva

ženská práva USE práva žen

Zkratky

viz/USE - před preferovaným
termínem/deskriptorem

ekv/UF - před nepreferovaným

Volba deskriptoru/nedeskriptoru

pravopis

používání nejrozšířenějšího popř. kodifikovaného pravopisného tvaru

využití autoritních slovníků

dublety (př.: sirup x syrob; salnitr x sanytr)

zkratky a akronymy (př.: UNICEF, OSN, ČEDOK, LIS)

preferuje se rozepsaná podoba - mezinárodní desetinné třídění X MDT [o]

výjimky v dané oblasti u široce používaných zkr. a akr.

přejatá slova (př.: software x programové vybavení; perestrojka x přestavba; gender studies x genderová/rodová studia)

neovtvary z cizího jazyka - běžně užívané lze

Volba deskriptoru/nedeskriptoru

slangové termíny a žargon (př.: hippies, íčko, koks)
lze zařadit, pokud neexistuje pro nový pojem jiná,
vhodná alternativa

jde-li o alternativní výraz k existujícímu,
ustálenému termínu, ustálený je preferovaný t.

obecné názvy a obchodní názvy (př.
polytetrafluorethylen x teflon)

preferované jsou obecné názvy jako deskriptory,
je-li obchodní název široce používán zavede se
jako nedeskriptor

běžné názvy a vědecké názvy (př.: bubo bubo x
výr velký; arthritis uratica x dna)

volba preferovaného termínu se odvíjí od uživatelů
systému

Kvazisynonyma

LJ, jejichž významy se všeobecně v běžném úzu považují za rozdílné - pro účely indexování se chápou jako synonyma.
např. tvrdost x měkkost; anotovaný záznam x neanotovaný záznam; zákonnost x nezákonnost

Hierarchizace vztahu ekvivalence

např. HORNINY

EKV břidlice

čedič

žula

Hierarchický vztah

Nastává u lexikálních jednotek téhož sémantického okruhu a vyjadřuje poměr nadřazenosti a podřazenosti.

Je vytvářen pouze u preferovaných termínů.

Podřazená jednotka musí náležet k témuž základnímu typu pojmu jako její nadřazená lex. jednotka – věc x činnost x vlastnost.

např.

ochrana soukromí

deskriptor

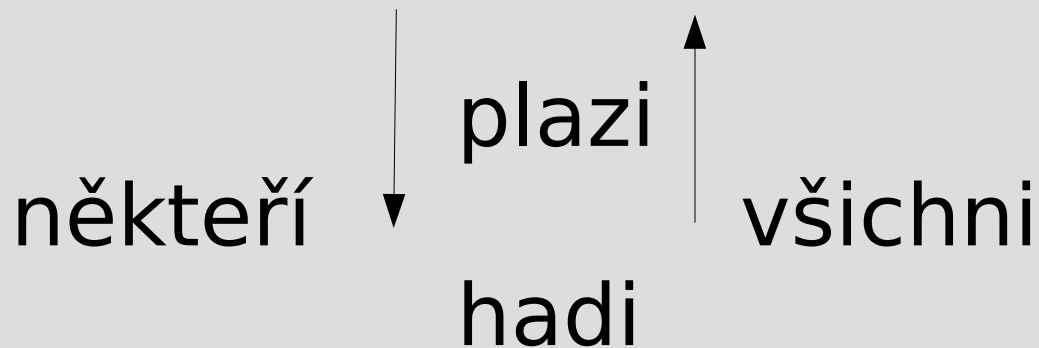
BT/ND práva jednotlivce

nadřazený

d.

Hierarchický vztah - generický vztah

Vymezuje souvislost mezi určitou třídou nebo kategorií a jejími členy nebo druhy.



rodo-druhový vztah

Hierarchický vztah - partitivní vztah

Sémantický hierarchický vztah mezi dvěma pojmy, z nichž jeden vyjadřuje celek a druhý jednu jeho část. (výklad z TDKIV)

4 kategorie

tělesné systémy a orgány (př.: kostra -> lebka -> nadočnicové oblouky)

geografické lokality (př.: Evropa -> Rakousko -> Vídeň)

disciplíny (př.: jazykověda -> obecná jazykověda -> gramatika)

hierarchické společenské struktury (př.: univerzita -> fakulta -> katedra -> kabinet)

Hierarchický vztah - kauzální vztah

Vazba mezi kategorií věcí nebo událostí vyjádřenou obecným substantivem a individuálním případem dané kategorie (jednotková třída označena vlastním jménem).

např.
pohoří

Alpy

uvedené



Himaláj

individuální případy

obecné kategorie

Vztah asociace

Vazba mezi dvěma lex. jednotkami, které nejsou ve vztahu ekvivalence nejsou spolu v hierarchickém vztahu

Použití: Lexikální jednotky jsou mentálně natolik asociovány, že je účelné vyjádřit jejich vazbu v tezauru.

Pravidlo pro stanovení vztahu: Při indexaci jednou jednotkou je vyvolána potřeba indexace druhou asoc. jednotkou.

např. ochrana soukromí

RT/ad datové právo

osobní údaje

Asociace

Disciplína/objekt (př.: informační věda -- informace)

Úkon, proces/konatel, nástroj (př.: katalogizace - katalogizační pravidla)

Činnost/výsledek (př.: katalogizace - katalogizační záznam)

Činnost/předmět (př.: indexace -- dokument)

Pojem/vlastnost (př.: informace -- pravdivost)

Pojem/původ (př.: Francouzi -- Francie)

Kauzální souvislost (př.: nemoc -- léčení)

Věc/agens působící proti ní (př.: rostliny -- herbicidy)

pojem/jednotka měření (př.: informace -- bit)

synkategorematické výrazy (př.: čokoládová vejce)

Uspořádání LJ

abecední uspořádání

systematické uspořádání doplněné
abecedním rejstříkem

grafické uspořádání s abecední částí

Organizační aspekty vypracování tezauru

Excerptce LJ

deduktivní metoda

induktivní metoda

Záznamy LJ

Verifikace LJ

odborné slovníky, encyklopedie

existující tezaury

SSJ

Specifičnost LJ

Doplňování a vyřazování LJ

Využití výpočetní techniky

více viz ČSN 010193, s. 45 - 47

Výhody tezauru

Princip postkoordinace a z toho pramenící

přehlednost selekčního obrazu dokumentu

snadná tvorba hierarchických struktur

snadná údržba

snadná manipulace

Nevýhody tezauru

Rozklad víceslovných jednotek - nerespektování kompaktnosti termínu.

Informační šum způsobený parazitním (náhodným) spojením deskriptorů a v důsledku toho: velký ohlas irelevantních dokumentů

Omezení pouze na tematickou část obsahové charakteristiky dokumentu: systém deskriptorů a nedeskriptorů zahrnuje pouze tematické termíny; tento nedostatek se odstraňuje připojením podpůrných souborů identifikátorů (personálie, jména korporací, geografické názvy, atd.).

Příklady tezaurů

vícejazyčné tezaury

Eurovoc <http://europa.eu/eurovoc/>
http://www.psp.cz/kps/knih/e_zakinf.htm

MeSH

http://www.medvik.cz/medvik/search_titles.do?

Agrovoc <http://www.knihovna.uzpi.cz/>

ČPT (EET) - Český pedagogický tezaurus

http://www.npkk.cz/npkk/zakl_tez.php

PSH - Polytematický strukturovaný heslář

<http://old.stk.cz/psh.html>

jednojazyčné české tezaury

např. **ČTT** - Český teologický tezaurus

<http://www.etf.cuni.cz/~library/infoctt.html>

Povinná a použitá literatura

Aichison, J., Gilchrist, A., Bawden, D. 2000. *Thesaurus construction and use : a practical manual*. 4th ed. Chicago : Fitzroy Dearborn Publishers, c2000. **Kap. A, B**, s. 1-13. dostupné ve fondu ÚK FF MU

Balíková, M. 2001 **Problematika věcného pořádní informací a jejich zpřístupnění**. *Národní knihovna*, 2001, roč. 12, č. 3, s. 175-186. Dostupné z WWW: <http://full.nkp.cz/nkkr/NKKR0103/0103175.html>

ČSN 01 0193. Dokumentace. Pokyny pro vypracování a rozvíjení jednojazyčných tezurů. Úč. 1.2.1996. 52 s. - dostupné v ÚK FF MU, registrační pult

Pinkas, O. 2002. *Zpracování informačních fondů*. Vyd. 1.

V Praze : Vysoká škola ekonomická, 2002. **Kap. 7 Selekční jazyky**, s. 89-114, **Kap. 8. Tezaurus**, s. 115-126 - dostupné ve fondu ÚK FF MU

Doporučená a použitá literatura

- Aichison, J., Gilchrist, A., Bawden, D. *Thesaurus construction and use : a practical manual*. 4th ed. Chicago : Fitzroy Dearborn Publishers, c2000. 218 s.
- Kovář, Blahoslav. 1984. *Věcné pořádání informací a selekční jazyky*. Praha : ÚVTEI, 1984. 251 s.
- KTD : Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha : Národní knihovna České republiky, 2003. Dostupné z WWW: <http://sigma.nkp.cz/cze/ktd>
- Schwarz, Josef. 2007. *Selekční jazyky 1 : Úvod do problematiky* [ppt]. Přednáška č. 2 (kombinované studium). 9.11.2007.