

Organizace informací VIII.

12. 11. 2021

PhDr. Jiří Stodola, PhD.

Předmětové selekční jazyky (PSJ)

- používané české ekvivalenty - věcné autority, 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

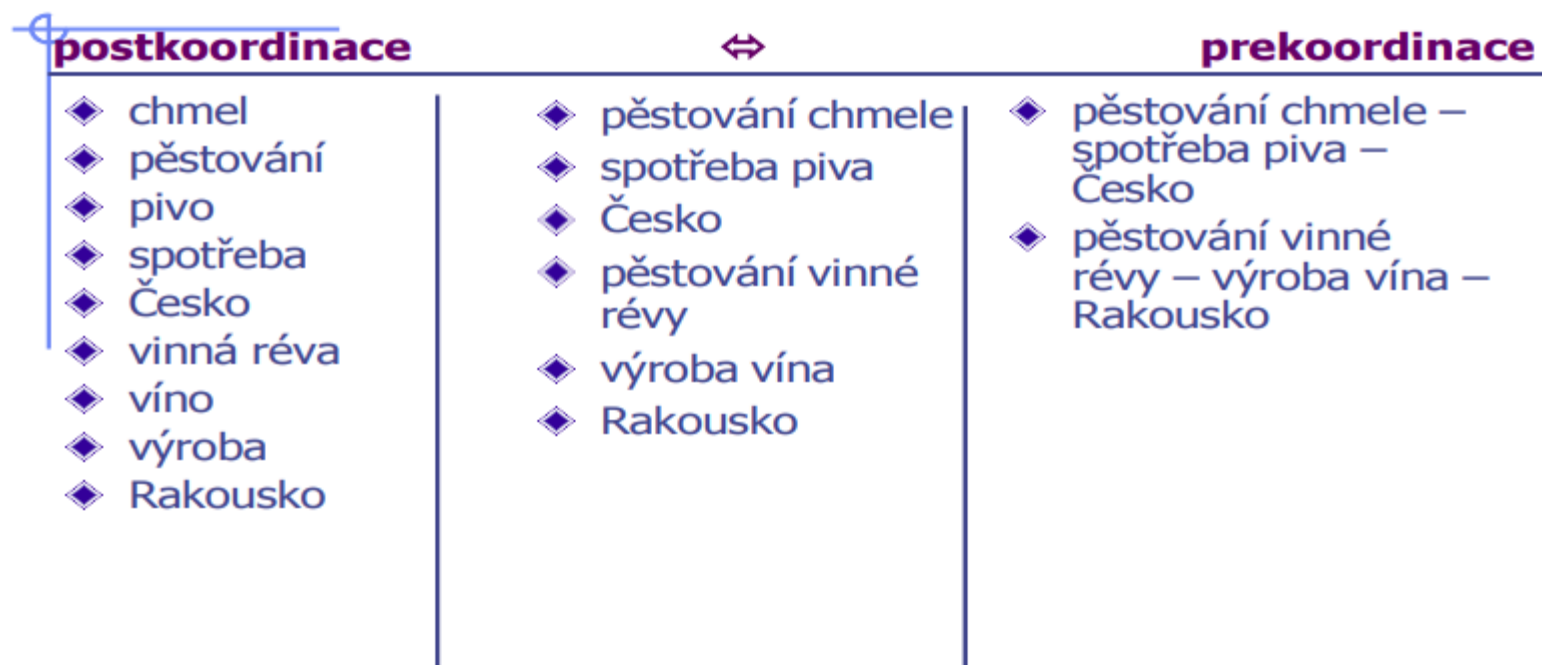
- Selekční jazyk založený na abecedně uspořádaném systému hesel, popř. lexikálních jednotek.
- **Předmětový SJ** je SJ používaný při realizaci procesu předmětového pořádání informací jako jeho výrazový a vyjadřovací prostředek. (Kovář, 1984, s. 129)
- **Předmětové pořádání informací** je proces, při 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. (volně podle Kovář, 1984, s. 13)
- **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 PSJ

- 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.
- Podle způsobu vazby znaků
 - 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 katalozích a kartotékách <http://www.lib.cas.cz/katalogy/predmetovy-katalog/>
- použití v online prostředí - modifikace pravidel

PH

- 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)
- ČSN 01 0188 Tvorba předmětových hesel

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, významově však odlišné prvky předmětového hesla.

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

- **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
 - komplikovaná údržba

Příklady

- Library of Congress Subject Headings (LCSH)
- nejrozšířenější univerzální PSJ pro využití v lístkových a online katalozích <http://authorities.loc.gov/>
- LISTA with Fulltext

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 k DSJ založenému na deskriptorech (tezaurům), jejichž forma a vztahy jsou standardizovány slovníkem se speciální strukturou - tezaurem.

Thesaurus

- **Řízený a měnitelný slovník** deskriptorového selekčního jazyka uspořádaný tak, že explicitně zachycuje apriorní (paradigmatické) vztahy mezi lexikálními jednotkami. (výklad TDKIV)
- **Thesaurus** - slovník PSJ, formálně organizovaný takovým způsobem, aby explicitně zachycoval apriorní (paradigmatické) vztahy mezi pojmy. (Aichison, 2000, s. 1)
- **paradigmatický vztah** - Vztah mezi pojmy, popř. výrazy, který existuje nezávisle na větném kontextu; v řízeném slovníku bývá vyjádřen vztahy mezi lexikálními jednotkami (např. vztah synonymie, homonymie, hierarchie, asociace apod.). (výklad TDKIV)

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 a asociované deskripty.

Kategorie pojmů

- **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

- substantivum (např. rodina)
- substantivní slovní spojení
- adjektivní např. národnostní menšina, „národnostní“ je modifikátor (adjektivum), „menšina“ je základ (substantivum)
- s předložkovou vazbou (např. kurzy pro pokročilé, školy pro nevidomé děti)
- adjektiva, příslovce, slovesa – obvykle se nezařazují, činnost – vyjádřena substantivem – výjimky - např. velmi krátkodobé financování, velmi krátké vlny (odborný termín)

Tvar

- plurál:
 - počítatelná substantiva (knihy, 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 např. křídlo (ptačí končetina), křídlo (klavír)

Základní vztahy v tezauru

- vztah ekvivalence
- vztah hierarchie
- vztah asociace

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 ekv/UF - před nepřeforovaným termínem/nedeskriptorem
viz/USE - před přeforovaným termínem/deskriptorem

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

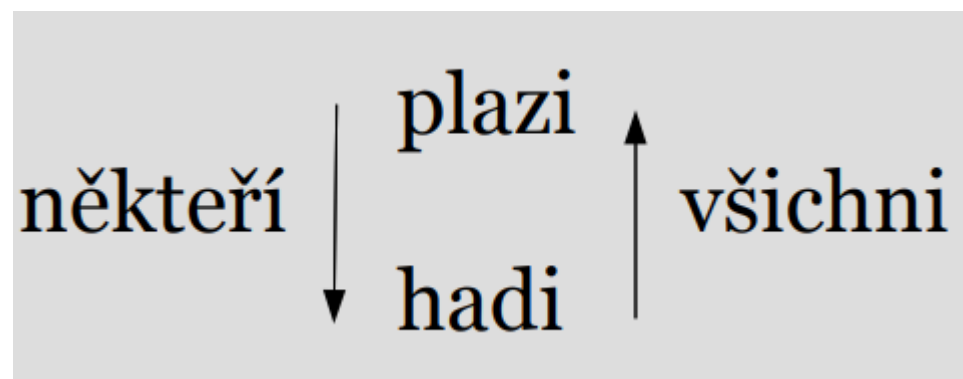
Hierarchický vztah

- 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.
NT/PD ochrana komunikací	podřazený d.

Hierarchický generický vztah

- Rodo-druhový vztah

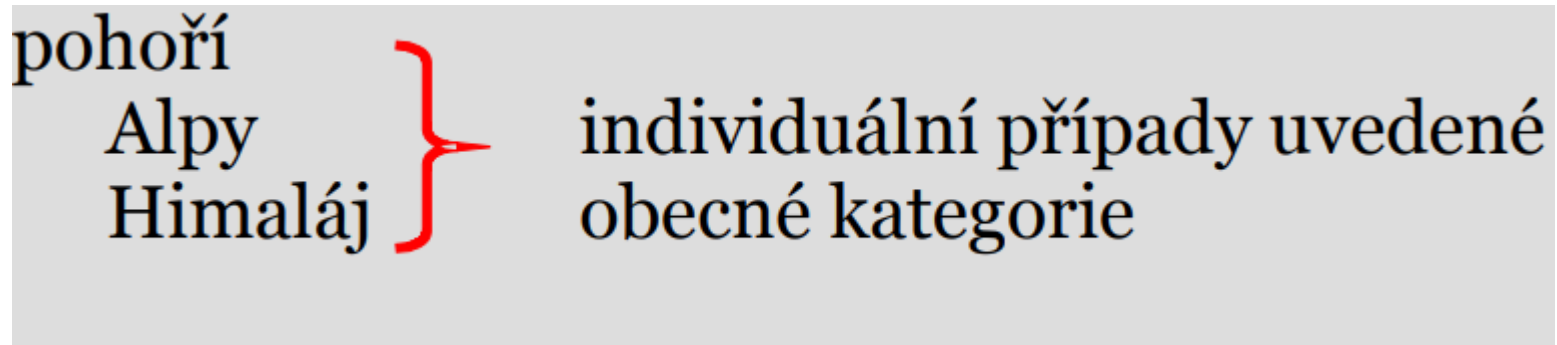


Hierarchický partitivní vztah

- 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 kauzální vztah

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



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 – vejce; umělé květiny -- květiny)

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 Tibet + 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 jazyky Substantivum, které vyjadřuje širší třídu věcí nebo jevů, jejímž členem je lexikální jednotka jako celek.
- modifikátor selekční 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.
- Příklad. Selekční jazyky, jazyky = základ, selekční = modifikátor

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, ale spíše popírají členství v dané třídě.

Sousloví rozdělená 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

ALE mražená zelenina, stěhovaví ptáci

Uspořádání lexikálních jednotek

- abecední uspořádání
- systematické uspořádání doplněné abecedním rejstříkem – fasetový přístup - např. Art & Architecture Thesaurus (záložka Browse the AAT hierarchies)
- grafické uspořádání s abecední částí

Organizační aspekty tezauru

- Excerpce LJ
 - deduktivní metoda
 - induktivní metoda
- Záznamy LJ
- Verifikace LJ
 - odborné slovníky, encyklopedie
 - existující tezaury
 - SSJ
- Specifičnost LJ
- Doplnování 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?source=mesh 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>