

Systematické selekční jazyky

Předmět: Selekční jazyky

9. 10. 2009

Přednášející: Mgr. Silvie Kořínková Presová

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

Varianty termínu

- **používané české ekvivalenty** - klasifikace, klasifikační systém, třídící systém, taxonomie, kategorizace
- **anglické ekv.** - systems for categorization, systematic retrieval language, classification, bibliographic classification, classification scheme

Výklad termínu

Systematický selekční jazyk (výklad z TDKIV)

- Věcný selekční jazyk založený na klasifikačním systému.
- Slovník systematického selekčního jazyka je tvořen klasifikačními znaky, jeho struktura je prezentována nejčastěji hierarchickými vztahy mezi třídami, a to zejména vztahy generickými a partitivními.

Klasifikační systém (výklad z TDKIV)

- Logické uspořádání množiny klasifikovaných pojmů na základě vztahů a závislostí mezi charakteristickými znaky těchto pojmů... (více viz výklad TDKIV).

Výklad termínu

- **Systematický selekční jazyk** je SJ používaný při realizaci procesu systematického pořádání informací jako jeho výrazový a vyjadřovací prostředek. (Kovář, 1984, s. 15)
- **Systematické pořádání informací** - Proces, při kterém jsou dokumenty, záznamy o nich nebo záznamy o jednotlivých informacích vřazovány na příslušné místo v rámci systematicky uspořádaného systému lidského poznání, kde se jejich postavení řídí v podstatě rododruhovými vztahy a slovní formulace obsahu dokumentu bývá většinou nahrazena znaky umělého jazyka (notacemi). (volně podle Kovář, 1984, s. 13)
ekv. třídění, klasifikace, systematické zpracování, systematická katalogizace

Výklad termínu

- **Classification scheme** - Specification of a systematic organization of knowledge. (Taylor, 2004, s. 359)
- **Clasisification scheme** - A tool for systematic organization of information resources. A class. scheme helps a classifier to represent the subject content of every document in a collection by appropriate notations. (Chowdhury, 2007, s. XVIII)

Klasifikace

- SSJ umožňují věcný přístup k dokumentům, jenž je založený na **klasifikaci**.
- **Klasifikace** (výklad z TDKIV) - 1. Uspořádání pojmů nebo jiných entit do tříd a dalších skupin (podtříd apod.) za účelem vyjádření jejich sémantických vztahů. Třídy a další skupiny jsou obvykle vyjádřeny notací. - 2. Přidělování notací (znaků tříd) klasifikačního systému za účelem vyjádření obsahu dokumentu.

Terminologická poznámka

- V TDKIV jsou uváděny jako samostatné termíny:
klasifikace a systematické pořádání informací.
- V zahraniční literatuře jeden termín a dva významy.
→ **classification** - 1. Umisťování předmětu do kategorií. 2. V oblasti organizace informací je klasifikace proces určování, kam inf. objekt spadá v dané hierarchii a často přiřazení notace.
(Taylor, 2004, s. 359)

Klasifikace

- **Klasifikace je v podstatě kategorizací.**
- V průběhu 20. stol. se propojení s kategorizací vytrácelo a klasifikace je chápána jako přiřazování notace dokumentům.
- Tradiční SSJ jsou založeny na **klasické teorii kategorií** - vychází z teorie Aristotela, že kategorie jsou tvořeny entitami mající společné vlastnosti.
- **Teorie prototypů (kognitivní psychologie)** - Teorie, že kategorie mají prototypy. Více viz Sternberg, 2002, s. 289-291.

Princip dělení/klasifikační kritérium

(Pinkas, 2002, s. 81, Kovář, 1984)

- **Množinu prvků** lze rozdělit na **podmnožiny** (podtřídy) pomocí zvoleného **principu dělení/principium divisionis**.
 - Principem dělení může být nějaká vlastnost nebo vztah.
 - **Inkluze množin** - Množina může být podmnožinou dané množiny jedině tehdy, jestliže všechny její prvky jsou také prvky dané množiny.
- Je-li podmnožin více, pak každá z nich musí mít alespoň jednu vlastnost odlišnou od dalších podmnožin.
- Vzniklé podmnožiny prvků lze dále dělit podle dalšího klasifikačního kritéria.

Princip dělení/klasifikační kritérium

- V klasifikaci se množinám říká **třídy**, podmnožinám **podtřídy**.
 - Podtřída je ve vztahu podřazenosti ke své třídě.
 - Třída je ve vztahu nadřazenosti ke své podtřídě.
 - Podtřídy stejné třídy jsou ve vztahu souřadném.
- Jestliže jsou v systému tříd definovány vztahy nadřazenosti, podřazenosti a souřadnosti, pak je v systému určena **hierarchie**.
 - **Hierarchický třídící systém** vzniká na množině prvků tehdy, je-li v něm definována hierarchie.

Příklad praktické klasifikace - Klasifikace ICT služeb

- **Klasifikace podle předmětu služby**
 - Informační služby
 - Aplikační služby
 - Infrastrukturní služby
 - Vývojové služby
 - Podpůrné služby
- **Klasifikace podle způsobu spotřeby ICT služby**
 - Služby s jednorázovou spotřebou
 - Služby s kontinuální spotřebou
 - Služby s diskrétní spotřebou
- **Klasifikace podle typu příjemce**

(Jelínek, 2007)

Účel SSJ

- Původní smysl SSJ - pořádání a získávání informačních objektů, dnes též pořádání bibliogr. záznamů (Taylor, 2004, s. 301)
- Funkce SSJ
 - využití pro tvorbu signatury
 - sdružuje obsahově stejné nebo podobné dokumenty (uživatelé mohou najít dokumenty o daném tématu na jednom místě)
 - spojuje dokumenty - z bibliogr. záznamu dokumentu navede klasif. znak na místo uložení dokumentu, pokud je využit pro vytvoření sign.
 - prohlížení - fyzické prohlížení fondu, prohlížení v online prostředí (Chowdhury, 2007, s. 72)

Důležité komponenty SJ

- **Třídy a podtřídy - třída** - Skupina klasifikovaných pojmů vzniklá na základě shodných charakteristik za účelem definování sémantických vztahů mezi nimi, která je označena notací. Tvoří základní prvek klasifikačního systému označující základní kategorii. (výklad z TDKIV)
- **Notace** - Symbol vyjádřený posloupností znaků, který se používá k označení tříd klasifikačního schématu. (výklad z TDKIV)
- **Klasifikační tabulky** - Klasifikační systém znázorněný formou tabulek, tj. systematické seskupení pojmů do tříd, podtříd, oddílů a pododdílů označených notacemi a opatřených slovním vyjádřením (slovními ekvivalenty). Klasifikační tabulky představují slovník SSJ. (klasifikačního systému). (více viz výklad z TDKIV)

Úkazka schématu MDT - Základní třída 3 Společenské vědy

30 Teorie, metodologie a metody společenských věd všeobecně.
Sociografie

31 Demografie. Sociologie. Statistika

32 Politika

321 Formy politické organizace. Státy jako politické síly

322 Vztahy mezi církví a státem. Politika a náboženství.
Církevní politika

323 Vnitřní politika. Vnitřní záležitosti

323.1 Národnostní, lidová, etnická hnutí a problémy.
Národnostní a etnické menšiny

323.2 Vztahy mezi obyvateli a státem. Vnitropolitické
akce

notace

slovní vyjádření

Společné prvky SSJ

- **Verbální vyjádření** klasifikovaných pojmů (témat reprezentovaných ve schématu).
- **Uspořádání těchto slovních vyjádření do tříd** nebo logickým způsobem tak, aby bylo zajištěno smysluplné pořádní témat.
- **Notace** (objevuje se vedle slovního vyjádření), které reprezentují slovní vyjádření a skrze které jsou explicitně vyjádřeny vztahy mezi třídami. Notace se slovním vyjádřením tvoří hlavní část klasif. tabulek.
- **Odkazy** uvnitř tabulek (poukazují na různé aspekty tématu nebo na jiné vztahující se témata).
 - Příklad Notace MDT: 025.4 <http://aip.nkp.cz/mdt/>

Společné prvky SSJ

- **Abecední seznam slovních vyjádření a synonym** (odkazy na notace).
- **Pravidla pro užití SSJ**. Většinou všeobecná pravidla na začátku tabulek a uvnitř tabulek specifická pravidla pro jednotlivé části. Příklad:
 - specifická pravidla Tabulka 1d. Všeobecné pomocné znaky formy <http://aip.nkp.cz/mdt/>
 - Obecná pravidla DDC <http://www.oclc.org/dewey/versions/abridgededition14/intro.pdf>
- **Institucionální zázemí** - udržování SSJ, revize, nové vydání. Příklad
 - MDT - UDC Consortium <http://www.udcc.org/>
 - DDC - OCLC <http://www.oclc.org/dewey/>

Základní typy SSJ

- **Enumerativní SSJ** - Vyjmenování všech možných tříd a podtříd klasifikačního schématu (klasifikátor přiděluje klasif. znak vyjadřující obsah dokumentu z předem vytvořeného schématu), příklad Library of Congress Classification (LCC), Dewey Decimal Classification (DDC) (dnes již má syntetizující prvky).
 - **prekoordinované SSJ**
- **Analyticko-syntetické SSJ** - při jejich aplikaci dochází k rozkladu složených témat na jednotlivé komponenty, pro jednotlivé komponenty jsou v tabulkách SSJ nalezeny notace, které jsou na základě daných pravidel spojeny v klasifikační znak, příklad MDT.
 - **míra prekoordinovanosti je stále významná, při indexaci lze uplatnit princip postkoordinace**

Příklady MDT versus DDC

- vyjádření složeného tématu **stupně místní, oblastní a ústřední správy**

→ klasifikační znak MDT **352/354** DDC **351**

notace z hlavních tabulek

symbol rozšíření

- vyjádření složeného tématu **Americká poezie**

• MDT **821.111(73)-1**

DDC **811**

notace
z hlavních tabulek

notace z pomocných
tabulek

Základní typy SSJ

- **Fasetové SSJ** - organizace pojmů do faset, klasifikační znak označující téma je vytvářen klasifikátorem s pomocí specifických pravidel pro syntézu, částečná hierarchie
příklad Colon Classification (CC)
 - **Faseta** (výklad z TDKIV) - Kategorie entit (množina podtříd) vytvořená uplatněním jedné klasifikační charakteristiky (principium divisionis), která je pro danou kategorii (třidu) podstatná, strukturální. Fasety vyjadřují vlastnosti použité pro seskupování pojmů podle jejich podstaty.
- **S. R. Ranganathan** (Indický knihovník a matematik) - uvedl fasetovou analýzu do knihovnictví, autor CC.

Fasetové SSJ pokračování

- CC je založeno na principu **fasetové analýzy**, podle které každý předmět může být analyzován na základě **základních kategorií**:
 - osobnost – hmota – energie – prostor – čas.
- Sled základních kategorií tvoří tzv. **fasetovou formuli, PMEST**, která určuje obecné pořadí faset při analýze a syntéze.

Příklad klasifikační znak **L 45:421:4** - (léčení tuberkulózy plic)

označení hlavní třídy lékařství

faseta orgány těla (osobnost)

faseta problémy

medicíny (hmota-vlastnost)

(energie)

faseta ošetřování a léčení

Příklady faset pro třídu nápoje

- [druh] [faseta]
 - pivo
 - víno
 - voda
 - mléko
- [obsah alkoholu]
 - alkoholické nápoje
 - nealkoholické nápoje
- [původ nápoje]
- [původ nápoje - pivo]
 - Prazdroj
 - Velkopopovické
- [původ nápoje - víno]
 - Velkopavlovická podoblast
 - Mělnická podoblast
- [typ nápoje]
- [typ nápoje - pivo]
 - ležák
 - kvasnicové
- [typ nápoje - víno]
 - bílá vína
 - Müller-Thurgau
 - Tramín červený
 - Sauvignon
 - červená vína
 - Frankovka
 - Portugal
 - Denis d'Or

- 1 [663.4](#) Pivo. Pivovarnictví. Sladování.
- 2 [663.41](#) Typy piva.
- 3 [663.42](#) Suroviny pro výrobu piva.
- 4 [663.421](#) Obiloviny pro výrobu piva. Sladovnický ječmen. Klíčivost.
- 5 [663.422](#) Náhražky ječmene.
- 6 [663.423](#) Chmel. Lupulin.
- 7 [663.424](#) Náhražky chmele.
- 8 [663.43](#) Příprava sladu. Sladování.
- 9 [663.432](#) Praní a máčení ječmene pro přípravu sladu. Úprava máčecí vody.
- 10 [663.433](#) Klíčení ječmene. Příprava zeleného sladu.
- 11 [663.433.1](#) Sladovací postupy a zařízení.

Ukázka z podřídny MDT
663.4 Pivo. Pivovarnictví.
Sladování.

Sledujte klasifikační
kritéria/charakteristiky
využité pro dělení na
další podtřídy.

<http://aip.nkp.cz/mdt/>

Samostatný úkol

- Identifikujte fasety z oblasti literární vědy a literatury pomocí témat charakterizující jednotlivá díla. U jednotlivých faset vyjmenujte pojmy z příkladu.
- **Charakteristika děl:**
 - Román ve francouzštině od belgického spisovatele.
 - Básně ve francouzštině od kanadského spisovatele.
 - Literárněvědná studie o českém spisovateli.
 - Zvuková nahrávka románu Zámek od F. Kafky.
 - Výbor z díla Milana Kundery.

Určené fasety

- **Faseta jazyka** (francouzština, čeština, němčina)
- **Faseta literárního žánru** (román, poezie)
- **Faseta národnosti/etnické skupiny/státní občanství** (Češi, Belgičané, Francouzi, Kanadčané, Židé)
- **Faseta formy dokumentu** (patří i typ nosiče) (studie, zvuková nahrávka, výběry)
- **Faseta osob** (spisovatel)

- Pomocí vytvořené fasetové klasifikace klasifikujte následující dílo:
 - Román ve francouzštině od belgického spisovatele.
- fasetová klasifikace viz příložený **soubor pdf**
EnumerativníSSJ_vers_fasetovéSSJ

Pomocí vytvořené fasetové klasifikace
klasifikujte následující dílo:

Román ve francouzštině od belgického
spisovatele.

- fasetová klasifikace viz příložený **soubor pdf**
EnumerativníSSJ_vers_fasetovéSSJ

→ klasifikační znak: A1 B1.1 C2

PRINCIP POSTKOORDINACE

- Klasifikace tématu pomocí MDT
Román ve francouzštině od belgického spisovatele.
- slovník MDT online <http://aip.nkp.cz/mdt/>
- vyhledání notace pro
 - francouzskou literaturu 821.133.1
 - Belgii (493)
 - román 82-31
- výsledný klasifikační znak 821.133.1(493)-31

Výhody a nevýhody jednotlivých typů SSJ

- **Enumerativní SSJ**
 - **výhody** - Struktura schématu se odráží v notaci, uživatel lehce rozpozná vztahy mezi pojmy (Podmínkou je hierarchická notace.).
 - **nevýhody** - Nové pojmy - pomalá aktualizace, nutné pravidelné revize. Stejný pojem se objevuje na více místech v systému (záleží za hledisku) - tabulky jsou příliš objemné.
 - **klasifikace** - Klasifikátor nemusí vytvářet klasifikační znak, na druhé straně jejich nevýhodou je, že jsou velmi neflexibilní a klasifikátor je omezen na třídy, které jsou vyjmenovány.
- (Chowdhury, 2007, s. 76 - 77)

Výhody a nevýhody jednotlivých typů SSJ

- **Analyticko-syntetické SSJ**
 - **výhody** - oproti enumerativním - Více druhů tabulek, schéma je menší - nemusí být vyjmenovány všechny třídy. Flexibilita v tvorbě specifických znaků (klasifikátor není tak omezen při vyjádření specifických témat).
 - **nevýhody** - Tvorba klasif. znaku je pro klasifikátora náročnější, různé způsoby aplikace - různé výsledné znaky pro stejné dokumenty v různých institucích.
- (Chowdhury, 2007, s. 77)

Výhody a nevýhody jednotlivých typů SSJ

- **Fasetové SSJ**
- **výhody** - Podporují vícehlediskovou klasifikaci. Lze vyjádřit specifické pojmy. Aktualizace tabulek vyvolaná změnami v teorii a praxi oboru je poměrně snadná.
- **nevýhody** - V praxi se fasetové klasifikace používají pro speciální třídění, hlavně v technice. Jejich většímu rozšíření brání poměrná složitost a náročnost tvorby klasif. znaku (Pinkas, 2002, s. 106). Z hlediska tvorby signatury je v mnoha případech nevýhodou dlouhá notace.

Charakteristika SSJ z hlediska lingvistických pojmů

Pořadací znak:

- **klasifikační znak**
 - notace
 - slovní vyjádření

Klasifikační znak (výklad z TDKIV) - Lexikální jednotka (pořadací znak) systematického selekčního jazyka, která vyjadřuje tematiku klasifikovaného dokumentu a slouží jako selekční údaj při jeho zpětném vyhledávání. Rozlišují se znaky hlavní a pomocné, základní, rozvinuté a složené.

Morfologie

Morfologie (pravidla pro tvorbu pořádacích znaků) se u SSJ týká výstavby notačního systému a tvorby notací.

Typy notace podle druhu použitých znaků

- alfabetská notace (KK, Odee)
- numerická notace (III.V, 4.6)
 - desetinná notace - pořadí notací: 1, 11, 12, 121, 2 (před první číslicí si představíme „0.“)
 - celočíselná notace - pořadí notací: 1, 2, 11, 12, 121.
- alfanumerická notace (smíšená notace) (A65, ЖIII.X)

Typy notace podle struktury

- expanzivní notace (pohostinnost notace)
 - lineární notace (A11, A12, A13...)
 - hierarchická notace (A11, A11.1, A11.1.01...)
- **mnemotechničnost notace** - např. A - Arts
- Výhodou hierarchické notace je, že reflektuje hierarchii SSJ.
- Nevýhodou hierarchické notace je větší délka notace ve srovnání s lineární. V systémech třídění, které nepoužívají hierarchické notace, je pružnost notace obvykle větší.

Charakteristika slovníku SSJ

- Založen na **klasifikaci**.
- Obvykle strukturován pomocí sémantických vztahů.
- **Slovník se nazývá - KLASIFIKAČNÍ TABULKY**
 - hlavní část uspořádána systematicky (klasifikační schéma)
 - pomocná část uspořádána abecedně
- **Struktura slovníku SJ (vztahy mezi PZ)**
 - silná hierarchie - enumerické a analyt.-synt. SSJ
 - částečná hierarchie - fasetové SSJ
 - asociace (viz též)

Syntax

- Pravidla pro užívání slovníku SJ, pravidla pro klasifikaci.
- Každý SSJ má své vlastní pravidla - Většinou všeobecná pravidla na začátku tabulek a uvnitř tabulek specifická pravidla pro jednotlivé části.

např.

Notace MDT: (1/9)

**Slovní vyjádření: Tabulka 1e. VŠEOBECNÉ POMOCNÉ ZNAKY
MÍSTA**

Použití pomocných znaků: Citační pořadí: tyto pomocné znaky se normálně používají až za notačním vyjádřením předmětu. Pořadí, které je založeno na vyjádření místa, může být tudíž uskutečněno tak, že se na začátek dostane pomocný znak jako první, například

339.5.053(73) Obchodní bilance - USA

Klasifikace na internetu

- **Kategorizační schémata na internetu** - umožňují setřídít informační zdroje podle obsahových charakteristik do tematických skupin – kategorií, které jsou většinou uspořádány do více hierarchických úrovní.
 - Služby umožňující vyhledávání EIZ prohlížením kategorizačních schémat - **předmětové adresáře** - komerční (dnes spíše portály), oborové brány/portály - www.seznam.cz, <http://dir.yahoo.com/>
- 2 základní přístupy ke kategorizaci zdrojů:
 - **aplikace tradičních SSJ** (většinou uspořádání dle notací+ slovní vyjádření)
 - **vytvoření vlastního schéma** - (většinou abecední uspořádání kategorií stejné hierarchické úrovně)

Tradiční SSJ na internetu

- Využití pro organizaci zdrojů zejména v digitálních knihovnách a oborových informačních branách (x komerční předmetové adresáře)
- Rozsáhlý výčet služeb (ne vždy se jedná o oborové informační brány) se schémata pro organizaci webových zdrojů založenými na klasifikačních systémech je dostupný z <http://www.iastate.edu/~CYBERSTACKS/CTW.htm> .
- Příklad - brána BUBL <http://bubl.ac.uk/> - využití DDC, schéma uspořádáno dle notací

Výhody využití enumer. SSJ

- **Prohlížení** - Usnadňují prohlížení v online prostředí
- **Rozšíření a zúžení rešerše** - Hierarchický princip klasifikačních schémat umožňuje potřebné rozšíření dotazu nebo postupnou specifikaci prohledávané tematické oblasti.
- **Kontext** - Výraz užitý pro kategorii nabývá smyslu v kontextu okolních kategorií.
- **Vícejazyčný přístup** - Notace jsou nezávislé na přirozeném jazyce a slovní vyjádření je dostupné ve více jazycích.
- **Interoperabilita** - Užití jednoho dohodnutého SSJ mezi jednotlivými službami může usnadnit paralelní prohlížení a vyhledávání.
- **Doménové rozdělení** - např. při tematickém rozšíření sbírky, lze fond rozčlenit na nové menší logické části.

(více viz Presová, 2005, 70-71)

Nevýhody využití tradičních univerzálních SSJ

- Vycházejí z klasifikace poznatků v 19. stol.
- Pomalá aktualizace.
- Nejsou dostatečně specifické pro konkrétní obory.
- Nedůslednost v hierarchických strukturách, např. ve schématu DDC potraty 363.46, homosexualita 363.49, pornografie 363.47 seskupeny jako problémy vztahující se k veřejné morálce v nadřazené třídě 363.4 Controversies related to public morals and customs (Co když je téma v dokumentu v odlišném vztahu, než do kterého je uveden ve schématu?)
- Roztříštěnost některých témat do různých podtříd. Např. v třídě pro literaturu nejsou v jedné skupině poetická díla, ale jsou rozdělena do jednotlivých podtříd.

Použitá a povinná literatura

- Pinkas, Otakar. 2002. *Zpracování informačních fondů*. Vyd. 1. V Praze : Vysoká škola ekonomická, 2002. Kap. 6 **Referování a indexování**, s. 79-88, Kap. 7 **Selekční jazyky**, s. 89-114.
- Presová, Silvie. 2005. *Metoda Konspektu a její vztah k selekčním jazykům : současný stav a trendy se zaměřením na situaci v České republice*. 1. vyd. V Brně : Masarykova univerzita v Brně, 2005. Kap. 4 **Možnosti a omezení využití konspektového schématu při zpřístupnění elektronických informačních zdrojů se zaměřením na oborové informační brány**, s. 67-72. ISBN 8021036850.
- Taylor, Arlene G. 2004. *The organization of information*. Westport : Libraries Unlimited , 2004. **Kap. 11 Systems for categorization**, 297-307 s. ISBN 1-56308-976-9

Použitá a doporučená literatura

- Chowdhury, G. G. ; Chowdhury, S. 2007. *Organizing information : from the shelf to the Web*. London : Facet, 2007. 230 s.
- Jelínek, Petr. 2007. *Kategorizace a architektury infromatických služeb* [online]. 2007 říjen [cit. 2009-10-04]. Dostupné z WWW: <http://nb.vse.cz/~vorisek/texty.htm>
- *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>
- Sternberg, Robert J. 2002. *Kognitivní psychologie*. Vyd. 1. Praha : Portál, 2002. 636 s. ISBN 8071783765.