

Dokumentace

POKYNY PRO VYPRACOVÁNÍ A ROZVÍJENÍ JEDNOJAZYČNÝCH TEZAUŘŮ

ČSN 01 0193

Documentation - Guidelines for the establishment and development of monolingual thesauri
Documentation - Principes directeurs pour rétablissement et le développement de thesaurus monolingues
Richtlinien für die Erstellung und Weiterentwicklung einsprachiger Thesauri

Y této normě jsou zapracovány údaje z ISO 2788: 1986 Dokumentace. Pokyny pro vypracování a rozvíjení jednojazyčných tezaurů.

Údaje souhlasné s ISO 2788: 1986 jsou označeny postranní čarou na levém okraji.

Předmluva

Citované normy

ISO 5963 dosud nezavedena

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

ČSN 01 0185 Transliterácia cyriliky

ČSN 01 0188 Tvorba předmětových hesel

ČSN 01 0192 Rejstříky publikací

ČSN 97 6030 Abecední řazení

Další souvisící normy

ČSN ISO 5127-6 Dokumentace a informace - Slovník - Část 6: Selekční jazyky (01 0163)

Obdobné mezinárodní, regionální a zahraniční normy

ANSI Z39. 19-1980 Guidelines for thesaurus structure, construction, and use (Pravidla struktury, vypracování a použití tezauru)

DIN 1463-1972 Richtlinien für die Erstellung und Weiterentwicklung deutschsprachiger Thesauri (Pravidla pro vypracování tezauru v němčině)

NC 39-14-1983 Elaboracion del tesauro unilingüe de búsqueda informativa (Vypracování jednojazyčného tezauru)

NF Z 47-103-1980 Thésaurus monolingues et multilingues. Symbolisation des relations (Jedno- a vícejazyčné tezaury. Symboly vztahů)

ÖNORM DIN 1463 Richtlinien für die Erstellung und Weiterentwicklung von Theasauri (Pravidla pro vypracování tezauru)

Porovnání s ISO 2788: 1986

Tato norma je v textových ustanoveních a v části příkladů s ISO 2788: 1986 plně kompatibilní. Pouze tam, kde to bylo nutné z hlediska názornosti zásad pro vypracovávání jednojazyčných tezurů v češtině, byly uplatněny tradiční české zkratky pro vztahy mezi lexikálními jednotkami a byly zpracovány nové příklady.

© Český normalizační institut, 1995

18419

ČSN 01 0193

Nahrazení předchozích norem

Touto normou se se nahrazuje ČSN 01 0193 z 27. 3. 1987.

Vypracování normy

Zpracovatel: Národní informační středisko České republiky, IČO 00001571, Dr. K. Pech Pracovník Českého normalizačního institutu: Jindřiška Bouřilová

Předmluva

ISO (Mezinárodní organizace pro normalizaci) je celosvětovou federací národních normalizačních organizací (členů ISO). Na mezinárodních normách obvykle pracují technické komise ISO. Každý člen ISO, který se zajímá o předmět, pro který byla vytvořena technická komise, má právo být zastoupen v této technické komisi. Práce se zúčastňují i mezinárodní organizace, vládní i nevládní, s nimiž ISO navázalo pracovní styk. ISO úzce spolupracuje s Mezinárodní elektrotechnickou komisí (IEC) ve všech záležitostech normalizace v elektrotechnice.

Návrhy technických norem přijaté technickými komisemi se rozesílají členům ISO k hlasování. Vydání mezinárodní normy vyžadují souhlas alespoň 75 % hlasujících členů.

Mezinárodní norma ISO 2788 byla připravena technickou komisí ISO/TC 46 Dokumentace.

Toto drahé vydání ruší a nahrazuje první vydání (ISO 2788-1974), ze kterého vycházela technická revize.

Všechny mezinárodní normy jsou čas od času revidovány. Každý odkaz na některou jinou mezinárodní normu v následujícím textu se proto vždy vztahuje k jejímu poslednímu vydání, pokud není uvedeno jinak.

Obsah

Strana

0 Úvod..... 3

1	Předmět normy.....	4
2	Odkazy.....	4
3	Termíny a definice.....	4
4	Zkratky a symboly.....	5
5	Kontrola slovníku.....	7
6	Lexikální jednotka.....	8
6.1	Všeobecně.....	8
6.2	Tvar lexikální jednotky.....	9
6.3	Singulár nebo plurál.....	10
6.4	Homografy a polysémy.....	12
6.5	Výběr lexikální jednotky.....	12
6.6	Vysvětlující poznámky a definice.....	14
7	Sousloví.....	15
7.1	Všeobecně.....	15
7.2	Sousloví ponechaná beze změny.....	16
7.3	Lexikální jednotky rozložené na syntaktické složky.....	16
7.4	Pořadí slov v sousloví.....	19
8	Základní vztahy v tezauru.....	19
8.1	Všeobecně.....	19
8.2	Vztah ekvivalence.....	19
8.3	Vztah hierarchie.....	21
8.4	Vztah asociace.....	24
9	Uspořádání lexikálních jednotek a vyjádření vztahů.....	27
9.1	Všeobecně.....	27
9.2	Abecední uspořádání.....	28
9.3	Systematické uspořádání.....	28
9.4	Grafické uspořádání.....	37
10	Organizační aspekty vypracování tezauru.....	45

10. 1 Metody shromažďování lexikálních jednotek.....	45
--	----

2

ČSN 01 0193

10. 2 Záznamy lexikálních jednotek.....	45
103 Verifikace lexikálních jednotek.....	45
10. 4 Specifičnost lexikálních jednotek.....	45
10. 5 Doplnění a vyřazování lexikálních jednotek.....	46
10. 6 Využití výpočetní techniky.....	46
10. 7 Formální úprava tezauru.....	46
10. 8 Jiné ediční aspekty.....	47
Příloha - Symboly pro vyjádření vztahů u tezaurů.....	48

0 Úvod

Efektivnost předmětového pořádání jako prostředku identifikace a vyhledávání dokumentů závisí na dobře zpracovaném selekčním jazyce. To platí pro každý systém, v němž výběr lexikálních jednotek jako pořadačích prvků vyžaduje intelektuální rozhodnutí člověka, tj. i pro ty systémy, v nichž je pro uložení lexikálních jednotek a manipulace s nimi nebo pro identifikaci dokumentů označených určitými lexikálními jednotkami, resp. jejich kombinacemi, využíváno počítače.

Při předmětovém pořádání vystávají tři hlavní úkoly

- a) vymezit téma dokumentu;
- b) vybrat termíny, které v souhrnu vyjadřují toto téma;
- c) vyznačit vztahy mezi pojmy, jež jsou vyjádřeny těmito termíny *).

První z uvedených úkolů je předmětem samostatné normy ISO 5963. Zbývající dva úkoly se netýkají jen indexátora, ale také uživatele systému. Tato norma se zabývá některými aspekty výběru lexikálních jednotek a obsahuje doporučené postupy při kontrole slovníku. Zejména je však věnována způsobům stanovení a vyjádření určitých typů vztahů mezi lexikálními jednotkami.

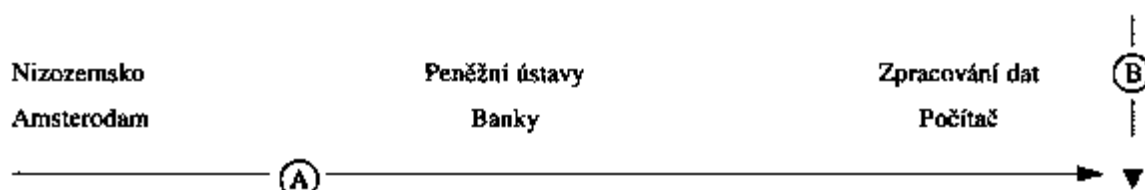
Lze rozlišit dva druhy těchto vztahů:

- a) syntaktické čili aposteriorní vztahy mezi lexikálními jednotkami, které v souhrnu vyjadřují téma dokumentu; např. indexátor zpracovávající dokument "Počítače v amsterodamských bankách" může použít tři lexikální jednotky: "Banky", "Počítače" a "Amsterdam"; v postkoordinovaném systému vztah mezi těmito lexikálními jednotkami není explicitně vyjádřen a dokument bude vyhledán, jestliže některá z nich nebo všechny budou obsaženy v dotazu; v prekoordinovaném systému mohou být vztahy mezi lexikálními jednotkami vyznačeny různými způsoby, např. pomocí symbolů vyjadřujících specifické vztahy, umístěním lexikálních jednotek v selekčním obrazu dokumentu, grafickou úpravou

lexikálních jednotek a/nebo interpunkcí; lexikální jednotky použité v uvedeném příkladu nejsou běžně asociovány v určitém obecném referenčním rámci a vztahy mezi nimi lze chápat tedy jako závislé na dokumentu;

b) apriorní čili tezaurové vztahy mezi lexikálními jednotkami, jejichž příbuznost je implicitní a vyplývá z určitého společného a obecně sdíleného referenčního rámce; v uvedeném příkladu "Banky" navozují nadřazenou lexikální jednotku "Peněžní ústavy", "Počítače" jsou myšlenkově asociovány se "Zpracování dat" a, Amsterdam" implikuje širší lokalizaci "Nizozemsko"; každá z uvedených myšlenkově asociovaných lexikálních jednotek může sloužit jako uživatelský vstup do systému; tyto vztahy jsou nezávislé na dokumentu, protože jsou uznávány obecně a mohou být ověřeny ve standardních příručkách, např. slovnících a encyklopediích.

Rozdíl mezi oběma uvedenými druhy vztahů může být znázorněn takto:



(A) = aposteriorní (syntagmatické) vztahy mezi lexikálními jednotkami přiřazenými dokumentu;

(B) = apriorní (paradigmatické) vztahy vyjádřené v tezauru

Tato norma je věnována apriorním vztahům, které mohou být vyjádřeny v tezauru. Tím získává určitý selekční jazyk, jak bylo výše doloženo, v jistém smyslu druhý rozměr.

*) Z praktických důvodů jsou "lexikální jednotka" a "pojem" občas zaměňovány.

3

ČSN OI 0193

1 Předmět normy

1.1 Smyslem doporučení obsažených v této normě je zajistit jednotnou praxi v rámci určitého systému indexování nebo několika různých systémů (např. propojených ve formě sítě). Nelze je však chápat jako striktně závazné instrukce. V řadě případů jsou zde popisovány postupy v různých variantách, např. pro vyjádření vztahů mezi lexikálními jednotkami, aniž se přitom některé z uvedených variant explicitně dává přednost. Volba určité z nich se totiž může lišit systém od systému v závislosti na organizačních rozhodnutích, která se vymykají z rámce této normy. Techniky popsané v normě jsou pokud možno založeny na obecných principech, které lze použít v jakékoli tematické oblasti. Přihlíží se však i k tomu, že indexátor v určité vymezené oblasti může někdy zdůvodněně překročit tato obecná doporučení, tyto případy jsou v dalším textu zpravidla výslovně vyznačeny.

1.2 Techniky popsané v této normě se podle možnosti neomezují na určitou konkrétní metodu indexování, např. postkoordinovanou nebo prekoordinovanou. Nicméně při jejím využívání je nutno respektovat následující podmínky:

a) norma se vztahuje na vyjádření a uspořádání lexikálních jednotek, které tvoří kontrolovanou podmnožinu přirozeného jazyka; neobsahuje postupy pro pořádání a vyjádření matematických a chemických vzorců;

b) norma se opírá o preferované termíny (viz 3. 5);

c) norma je určena pro ty systémy, v nichž lidé, indexátoři, provádějí analýzu dokumentů a vyjadřují jejich témata lexikálními jednotkami řízeného selekčního jazyka; není určena pro systémy, v nichž jsou zavedeny techniky plně automatického indexování a v nichž termíny vyskytující se v textu jsou seskupovány podle hledisek stanovených počítačem, např. podle frekvence jejich výskytu a/nebo podle jejich příležitosti v textu; dobře vypracovaný jednojazyčný tezaurus však může být cennou pomůckou i při vyhledávání v těchto systémech;

d) norma se zabývá postupy spojenými s indexováním dokumentových souborů zaznamenaných v katalogích nebo bibliografiích; nevztahuje se tedy na vypracování rejstříků k publikacím, i když mnohé z postupů zde doporučených mohou být vhodné i pro tento účel.

POZNÁMKY

1 Pravidla pro tvorbu předmětových hesel uplatňovaných při vytváření předmětových katalogů a kartoték stanoví ČSN 01 0188.

2 Pravidla pro tvorbu předmětových rejstříků stanoví ČSN 01 0192.

1. 3 Doporučení obsažená v této normě se vztahují k jednojazyčným tezaurům a nepřihlížejí ke speciálním požadavkům na vícejazyčné tezaury, tj. takové tezaury, v nichž pojmové ekvivalenty jsou vyjádřeny lexikálními jednotkami z více nežli jednoho přirozeného jazyka. Vypracování a vedení vícejazyčného tezauru je předmětem samostatné ISO 5964. Protože principy, na nichž je založena tato norma, lze pokládat za nezávislé na kontextu určitého přirozeného jazyka nebo určité konkrétní kultury, byly přijaty také za základ zmíněné vícejazyčné normy. Z toho vyplývá, že obecné principy a postupy platné pro oba typy tezaurů jsou v úplnosti obsaženy pouze v této mezinárodní normě, tj. nejsou opakovány v ISO 5964.