


2005

	Počítačové aplikace v terminologii - Systém pro značkování terminologie	ČSN ISO 16642 01 0507
---	--	---------------------------------

Computer applications in terminology - Terminological markup framework

Applications informatiques en terminologie - Plate-forme pour le balisage de terminologies informatisées

Tato norma je českou verzí mezinárodní normy ISO 16642:2003. Mezinárodní norma ISO 16642:2003 má status české technické normy.

This standard is the Czech version of the International Standard 16642:2003. The International Standard ISO 16642:2003 has the status of a Czech Standard.

© Český normalizační institut,
2005

Podle zákona č. 22/1997 Sb. smějí být české technické normy rozmnožovány a rozšiřovány jen se souhlasem Českého normalizačního institutu.

71856

Strana 2

Národní předmluva

Citované normy

ISO 1087-1 zavedena v ČSN ISO 1087-1 (01 0501) Terminologická práce - Slovník - Část 1: Teorie a aplikace

ISO 1087-2 zavedena v ČSN ISO 1087-2 (01 0501) Terminologická práce - Slovník - Část 2: Počítačové aplikace

ISO 12620:1999 Computer applications in terminology - Data categories dosud nezavedena

Citované dokumenty

Extensible Markup Language (XML) 1.0, Second edition, BRAY, T., PAOLI, J., SPERBERG-McCQUEEN, C. M., and MALER, E. (eds.), W3C Recommendation 6 October 2000, dostupné na <<http://www.w3.org/TR/REC-xml>>

Dublin Core Qualifiers, 2000-07-11, dostupné na <<http://dublincore.org/documents/2000/07/11/dcmesqualifiers/>>

XHTML™ 1.0 The Extensible HyperText Markup Language, 2nd edition, dostupné na <<http://www.w3.org/TR/xhtml1/>>

Poznámky k překladu

Celý základní (normativní) text je překladem anglického originálu ISO 16642:2003.

Příklady v kapitole 5 a v přílohách jsou z důvodu zachování plného obsahu originálu a současně maximální jasnosti a přehlednosti řešeny tak, že příkladové termíny, poznámky atd. uváděné v anglickém jazyce jsou ponechány v původní anglické podobě. Příkladový materiál v jiných jazycích zůstává ve stejné podobě, jako je uveden v anglickém originálu, poznámky, které jsou v těchto případech uváděny v angličtině, jsou přeloženy do češtiny.

Vypracování normy

Zpracovatel: NLPforge, s.r.o., IČ 26225514, RNDr. Pavel Smrž, Ph.D.

Technická normalizační komise: TNK 114 Terminologie. Principy a koordinace

Pracovník Českého normalizačního institutu: Ing. Jaromír Čížek

Strana 3

MEZINÁRODNÍ NORMA
Počítačové aplikace v terminologii -
Systém pro značkování terminologie

ISO 16642
První vydání
2003-08-15

ICS 01.020; 35.240.30

Obsah

Strana

Úvod

.....

.....	5
1 Předmět normy
.....	6
2 Normativní odkazy
.....	6
3 Termíny a definice
.....	6
4 Obecné principy a princip interoperability.....	8
5 Generický model pro popis lingvistických dat a jeho použití pro terminologii.....	10
6 Definování TML
.....	22
Příloha A (normativní) XML schéma formátu GMT.....	26
Příloha B (normativní) Jazyk pro značkování terminologie (TML) MSC.....	28
Příloha C (normativní) TML Geneter.....	32
Příloha D (informativní) Soulad terminologických dat s TMF.....	45
Bibliografie
.....	50

technických komisí ISO. Každá členská instituce zájímající se o předmět, pro nějž byla ustavena technická komise, má právo být v této komisi zastoupena. Práce se zúčastňují rovněž vládní i nevládní mezinárodní organizace spolupracující s ISO. ISO úzce spolupracuje s Mezinárodní elektrotechnickou komisí (International Electrotechnical Commission, IEC) ve všech záležitostech týkajících se normalizace v elektrotechnice.

Mezinárodní normy jsou navrhovány v souladu s pravidly určenými v Části 2 Směrnic ISO/IEC.

Hlavním úkolem technických komisí je připravit mezinárodní normy. Návrhy mezinárodních norem přijaté technickými komisemi se rozesílají členům ke schválení. Pro vydání mezinárodní normy je požadován souhlas nejméně 75 % hlasujících členů.

Je třeba upozornit na možnost, že některé části tohoto dokumentu mohou být předmětem patentových práv. ISO nelze považovat za odpovědnou za zjištění jakéhokoliv nebo všech takových patentových práv.

Norma ISO 16642 byla zpracována technickou komisí ISO/TC 37 *Terminologie a jiné jazykové zdroje*, subkomise SC 3 *Počítačové aplikace v terminologii*.

Strana 5

Úvod

Terminologická data jsou shromažďována, zpracovávána a uchovávána širokou škálou systémů, obvykle v aplikacích, tj. různých druzích systémů řízení bází dat, od aplikací pro osobní počítače určené jednotlivým uživatelům až po serverové systémy terminologických databank provozované velkými společnostmi a vládními agenturami. Terminologické databáze obsahují různé soubory datových kategorií a jsou založeny na různých typech datových modelů. Terminologická data musí být často sdílena a opětovně využívána v řadě aplikací a toto sdílení je obvykle uskutečňováno použitím formátů, do nichž jsou data speciálně převedena. K usnadnění spolupráce a zamezení duplicitních prací je třeba, aby byly vyvinuty normy a směrnice pro vytváření a používání souborů terminologických dat, jakož i pro sdílení a výměnu dat.

Metamodel definovaný touto mezinárodní normou je vhodný pro integrovaný přístup využitelný při analýze stávajících a při návrhu nových souborů terminologických dat, které jsou běžně zpracovávány relačními nebo textově orientovanými databázovými systémy. Soubory terminologických dat mohou být rovněž ukládány jako strukturované dokumenty se značkováním založeným na formátech typicky definovaných pomocí jazyka SGML (Standard Generalized Markup Language), daného v ISO 8879 [12] nebo XML (eXtensible Markup Language), který vychází z SGML, ale je doplněn Konsorciem W3C (World Wide Web Consortium) pro použití na www stránkách. Integrovaný přístup usnadňuje úlohy načtení dat z prostého souboru se značkováním do databáze a exportu z databáze do strukturovaného dokumentu. Další motivací využití integrovaného přístupu je to, že narozdíl od zcela oddělených přístupů k databázím a k strukturovaným dokumentům jsou formáty založené na XML zpracovávány novým způsobem, který se podobá klasickým systémům řízení bází dat. Tak např. jsou soubory XML dotazovány a aktualizovány přímo, aniž by byla data importována do tradičních prostředí databází.

Tento integrovaný přístup k analýze a návrhu má dvě úrovně abstrakce. První (a nejabstraktnější) úroveň integrovaného přístupu je úroveň metamodelu. Tato úroveň, kterou by bylo možné rovněž nazývat úrovní abstraktního pojmového datového modelu, podporuje analýzu a návrh ve velmi

obecné rovině. Druhou úrovní je úroveň datového modelu.

Na úrovni datového modelu má ten, kdo navrhuje soubor terminologických dat, možnost různého výběru alternativ podle potřeb praxe. Přitom musí nejprve určit nejvhodnější formu reprezentace svých terminologických dat na základě následujících voleb:

- zda použít relační databázi nebo prostý soubor se značkováním;
- zda budou data primárně použita k dotazování a aktualizacím a budou reprezentována v některých databázových systémech, a pokud ano, který systém použít;
- zda budou data primárně použita ke sdílení a vzájemné výměně a budou reprezentována v prostém souboru se značkováním.

Pro účely této mezinárodní normy je předpokládáno, že veškeré prosté soubory budou používat značkování XML.

Jakmile byla provedena volba - systém řízení báze dat nebo prostý soubor se značkováním - je třeba vybrat datový model. Pro relační databázi je typickou metodou popisu datového modelu ER-diagram (entity-relationship diagram - diagram vztahu mezi entitami). Pro dokument XML je typickou metodou popisu datového modelu DTD (Document Type Definition - definice typu dokumentu). Alternativní metoda, využívající tzv. „XML schéma“, je stanovena Konsorciem W3C. V budoucnosti bude možné používat ještě abstraktnější metody popisu XML formátu.

Konkrétní zavedení metamodelu pro značkování terminologie vyjádřené v XML se nazývá jazyk pro značkování terminologie (neboli TML, Terminological Markup Language). Ten lze popsat na základě omezeného počtu charakteristik, jmenovitě:

- jak TML vyjadřuje strukturální organizaci metamodelu (tj. stromy rozvoje TML),
- specifické datové kategorie používané TML a jak se týkají metamodelu,
- způsob, kterým mohou být tyto datové kategorie vyjádřeny v XML, a tím „ukotveny“ na stromech rozvoje TML, tj. XML styl jakékoliv dané datové kategorie, a
- slovníky používané TML k vyjádření rozdílných informačních objektů, jako jsou prvky a atributy XML podle odpovídajících XML stylů.

Některé z příkladů v této mezinárodní normě jsou instancemi formátů MSC (MARTIF with Specified Constraints - MARTIF se stanovenými omezujícími podmínkami) a Geneter, jak je popsáno v příloze B, resp. v příloze C.

1 Předmět normy

Tato mezinárodní norma stanovuje systém navržený k poskytnutí instrukcí o základních principech reprezentace dat zaznamenaných v souborech terminologických dat. Tento systém zahrnuje metamodel a metody popisu specifických jazyků pro značkování terminologie (TML) vyjadřovaných v XML. Jsou definovány mechanismy pro zavedení omezujících podmínek v TML, s výjimkou tří TML definovaných v přílohách B až D však nejde o specifické omezující podmínky pro jednotlivé TML.

Tato mezinárodní norma je navržena s cílem podpořit vývoj a použití počítačových aplikací pro terminologická data a výměnu takových dat mezi jednotlivými aplikacemi. Není normou pro datové kategorie a metody pro stanovení datových struktur, které jsou dány ISO 12620 a jinými souvisejícími mezinárodními normami.

Tato mezinárodní norma rovněž definuje podmínky, které umožňují, aby data vyjádřená v jednom TML mohla být mapována na jiné TML, a stanovuje pro tento účel generický nástroj mapování (GMT - Generic Mapping Tool - viz příloha A).

Kromě toho tato norma popisuje generický model pro popis lingvistických dat.

-- Vynechaný text --