

Správa obsahu a dokumentů – databázový návrhový vzor

Helena Palovská
Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky
nám. W. Churchilla 4, 130 67 Praha 3
palovska@vse.cz

Abstrakt: Je představen databázový návrhový vzor pro správu obsahu či dokumentů. Jedná se o generický model, nabízející řešení pro jakýkoli konkrétní případ systému pro správu obsahu či dokumentů. Cesty možné implementace jsou nastíněny.

Klíčová slova: databázový návrhový vzor, správa obsahu, správa dokumentů

Abstract: A database design pattern for content or document management. This is a generic model, offering solutions for any particular case. The paths of possible implementation are outlined.

Key words: database design pattern, content management, document management

1. Úvod

V (Palovská, H. 2014) byla představena idea databázových návrhových vzorů a byly shrnuty přínosy významných autorů v této oblasti. Byly také sděleny zkušenosti z výuky databázových návrhových vzorů na Vysoké škole ekonomické v Praze. V průběhu seminářů na této škole byly postupně vyvíjeny modely a vzory, vycházející z teorie i z praxe účastníků. Ty procházely vývojem až do podoby, jež byla shledána vhodnou k publikaci. Publikace (Palovská, H. 2015), (Palovská, H. 2016a), (Palovská, H. 2016b), (Palovská, H. 2017) prezentovaly zárodečné modely "Řízení přístupových práv", „Plánování a evidence skutečné spotřeby“, „Evidence průběhu pracovních procesů“ a „Užívání vzácných zdrojů“. Tento následující článek uvádí návrhový vzor „Správa dokumentů“.

David Hay se v (Hay, D. 1996, kapitola 11) zabýval problematikou správy dokumentů, od roku 1996 však vzniklo a je používáno mnoho pokročilých systémů správy obsahu ukazujících, že rozsah dané problematiky je podstatně širší a hloubka je podstatně větší. Předložený návrhový vzor je založen na zkušenostech z široce používaných systémů správy obsahu, tyto zkušenosti integruje a rozšiřuje o obecné zkušenosti z analýzy informací a znalosti o ontologických strukturách.

2. Jak používat předložený vzor

Zopakováno z (Palovská, H. 2015):

Databázové návrhové vzory nejsou produkty k bezprostřednímu nasazení. Mohou být použity buď jako zárodky či stavební kameny pro návrh databáze pro konkrétní aplikaci. Dle terminologie Blahy (Blaha, M. 2010) je vzor předložený v tomto článku *zárodečným modelem*; zároveň je modelem generickým, jehož účelem je pokrýt

všechny případné specifické případy nasazení. Z toho plyne, že v konkrétním případě mohou být některé části nepotřebné, či může být vhodné je zjednodušit.

Před použitím návrhového vzoru je třeba tento vzor pochopit. Samotná konstrukce – datová struktura – vyžaduje být prostudována, nicméně hrozí nepochopení způsobené špatnou interpretací užitých pojmů. Proto jsou návrhové vzory doplněny *slovníkem užitých pojmů*. Pochopení použitých hesel slovníku pojmů je nezbytné pro správné pochopení návrhového vzoru. Ve vzoru jsou používána slova obecného jazyka, tedy slova, jež mají několik možných významů – na konkrétním významu užitém v daném návrhovém vzoru je proto učiněna dohoda.

Při nasazení je třeba respektovat slovník užívaný v daném aplikačním prostředí, a termíny modelu je třeba *převést*.

3. Kontext a problematika správy obsahu/dokumentů

Objekty nazývané zde *obsahem* mohou být strukturované, semi-strukturované i nestrukturované, informačně bohaté rozsáhlé datové objekty – dokumenty, obrázky, videa apod. Samy se dále mohou stát částí jiných obsahů, a to jak v navržených strukturních elementech, tak i volně v nich umístěné.

Z pohledu sděleného významu, tématu či zaměření na tyto objekty nahlížíme jako na *dokumenty o něčem*. Jakožto takové je potřebujeme uchovávat a zpřístupňovat je. K tomu je potřeba vědět, jak je nalézt, zda existují jejich další verze a jak nalézt tyto verze. Pro posouzení relevance dokumentu k danému účelu je potřeba znát jeho téma; k tomu se často používá systém štítků, značek či klíčových slov (tagů), tento systém může být všelijak sofistikovaně provázán. Ve vztahu k věcné oblasti, pro kterou má správa dokumentů sloužit, mohou být dokumenty přímo propojeny s konkrétními business objekty.

Další informace o obsahu, jako kdo jsou jeho autoři nebo další osoby za něj zodpovědné, slouží při *správě*, a jsou také užitečné při posuzování relevance a důvěryhodnosti při vyhledávání.

Pro *tvorbu obsahu* bývají formulována pravidla týkající se práv a povinností rolí aktérů tvorby. Tato pravidla se mohou týkat struktury obsahu, předpisů vzhledu, užití jednotlivých typů obsahu ke skladbě dalšího obsahu. Bývají před-vytvářeny návrhy pro určité typy obsahů.

Oba pohledy na danou problematiku, správa dokumentů i správa obsahu, se prolínají v potřebě evidovat odběratele a spravovat distribuci novinek.

4. Pojmy vzoru

Tato kapitola objasňuje specifický slovník užitý v analytickém vzoru „Správa obsahu, správa dokumentů“ z tohoto článku.

4.1 Obsah/dokument, typ obsahu/dokumentu, sorta

Obsah/dokument. Poněvadž na tentýž objekt lze nahlížet jako na obsah ve smyslu zmíněném v předchozí kapitole, a zároveň jako na dokument o něčem, je zde zvoleno označení obsah/dokument.

Typ obsahu/dokumentu. Formální vlastnosti obsahu či dokumentu mohou být samostatně popsány v typu obsahu či dokumentu.

Sorta obsahů/dokumentů. Obsahy či dokumenty mohou být zařazeny dle významu či účelu do sort.

Typ obsahu/dokumentu tedy popisuje strukturní elementy a jejich vlastnosti, sorta obsahu/dokumentu se týká tématu nebo zaměření.

Strukturní element, předpis strukturního elementu. Výraz „strukturní element“ je třeba ujasnit, protože se běžně používá ve dvou různých významech: strukturní element popsáný v typu obsahu či dokumentu, tedy typ strukturního elementu, jenž se má objevovat v obsahu či dokumentech daného typu, a strukturní element konkrétního obsahu či dokumentu. Poněvadž je potřeba oba významy odlišit, pro první význam (typ strukturního elementu) použijeme označení “předpis strukturního elementu”, a výraz “strukturní element” ponecháme pro druhý význam, tj. pro strukturní element konkrétního obsahu či dokumentu.

4.2 Role k obsahu/dokumentu, účastník

Role k obsahu/dokumentu. Obsah či dokument má autora či autory, korektory, redaktory, překladatele, ... To všechno jsou možné role k tomu obsahu či dokumentu. Vedle toho existují ještě jiné role: role ve správě obsahů/dokumentů, role ve věcné oblasti, pro kterou má správa dokumentů sloužit, a další.

Účastník je termín sjednocující osoby či jiné subjekty, jež mohou být obsazeny do rolí.

5. Pod-problematiky správy dokumentů

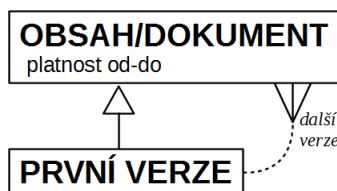
Problematika správy obsahu/dokumentů zahrnuje několik dílčích pod-problematik: verzování, struktura, umístění jednoho v jiném, sortování a tagování, auroři a jiné role, odběratelé a distribuce.

5.1 Verzování

Cílem řešení není efektivní zápis změn mezi verzemi, ale pouze evidence skutečnosti, že jeden obsah/dokument je verzí jiného.

Ačkoli návaznosti mezi verzemi téhož obsahu/dokumentu mohou být různě komplikované, přesto vždy vznikne nejprve první verze, na kterou všechny následující navazují. Tato první verze může pak být pojítkem celé množiny následujících verzí.

Řešení znázorňuje model na obrázku 1.

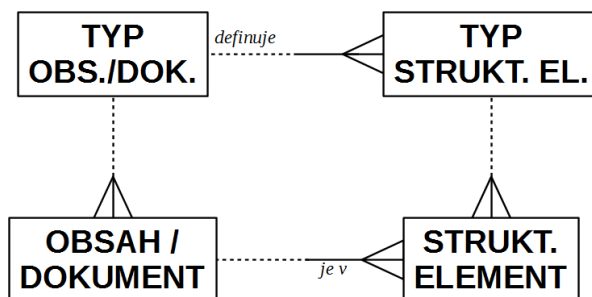


Obr 1. Verze obsahu (autorka)

5.2 Struktura obsahu/dokumentu

Obsah/dokument může mít strukturu, již lze řídit, a jejíž prvky lze samostatně spravovat a zpřístupňovat jejich obsah.

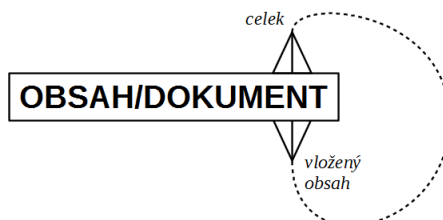
Typ obsahu/dokumentu může požadovat určitou strukturu s nějakými strukturálními prvky. Konkrétní obsah/dokument odpovídající danému typu obsahu/dokumentu má strukturální prvky, jež jsou odpovídajícího typu. Model na obrázku 2.



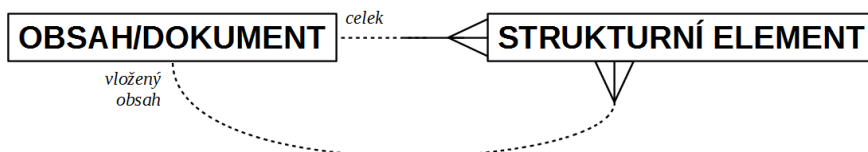
Obr 2. Struktura dokumentu (autorka)

5.3 Umístění obsahu či dokumentu v jiném

Dokument či obsah může zahrnout jiný obsah či dokument, a to buď volně, nebo do předem určeného strukturálního elementu. První případ řeší model na obrázku 3a, druhý případ řeší model na obrázku 3b.

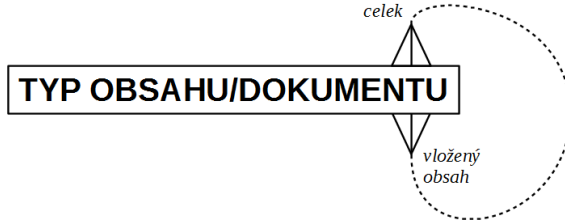


Obr 3a. Zahrnutí obsahu (autorka)

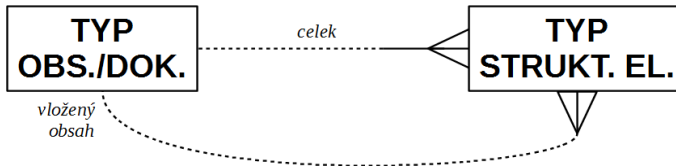


Obr 3b Zahrnutí obsahu ve struktuře (autorka)

Pro umístování obsahů/dokumentů v jiných mohou být stanovena pravidla na úrovni typů, buď pro volné umístování (obrázek 3c), nebo v závislosti na struktuře (obrázek 3d).



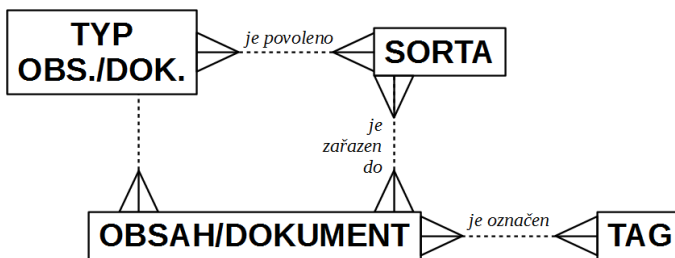
Obr 3c. Zahrnutí typu obsahu (autorka)



Obr 3d. zahrnutí typu obsahu ve struktuře (autorka)

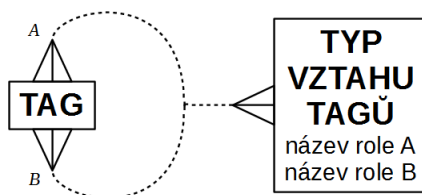
5.4 Sortování a tagování

Obsahy/dokumenty lze zařazovat do sort. V jedné sortě mohou obecně být obsahy/dokumenty různých typů, a obsahy/dokumenty jednoho a téhož typu se mohou vyskytovat v různých sortách. Pro vyhledávání označujeme obsah/dokumenty tagy. Libovolným tagem lze označit libovolný obsah/dokument, bez ohledu na typ či sortu. Model na obrázku 4a.



Obr 4a. – Typy, sorty a tagy (autorka)

Mezi jednotlivými tagy mohou být dány vztahy nějakého typu – např. “širší-užší”, “množina-prvek”, “entita-atribut”, “agregát-část”, či specifické vztahy pro oblast zájmu. Model na obrázku 4b.



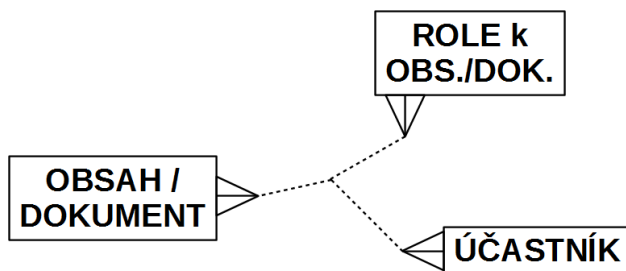
Obr 4b. Vztahy tagů (autorka)

Poznámka: Nejčastěji se používá jediný typ vztahu “nadřazený-podřazený”, tagy jsou tedy obvykle organizovány do hierarchií.

Poznámka: Tagováním lze vyřešit provázání obsahu/dokumentu k business objektům. Pak je potřeba zajistit konzistenci mezi systémem tagů a aktuálním stavem business objektů.

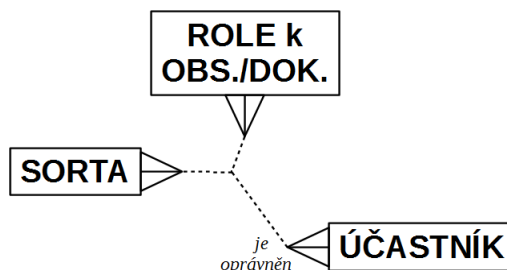
5.5 Autoři a jiné role

Evidence, kdo byl v jaké roli (autor, korektor...), v jakém obsahu/dokumentu – obrázek 5a.



Obr 5a. Zodpovědnosti k obsahu (autorka)

Na úrovni sort mohou být pro role k obsahu vyslovena pravidla – obrázek 5b.



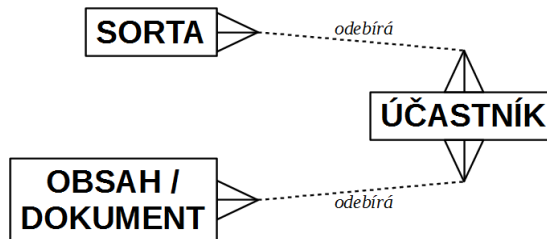
Obr 5b. Práva v sortě (autorka)

Podobně mohou být definována práva a evidovány zodpovědnosti za obsah strukturních elementů. Práva mohou být udělována nikoli přímo jednotlivým účastníkům, ale skupinám. Podrobněji je tato problematika rozebrána v (Palovská, H., 2015).

5.6 Odběratelé a distribuce

Účastník se může přihlásit k odběru sorty (odebírá všechny další obsahy/dokumenty v té sortě) nebo dokumentu (odebírá všechny jeho další verze) – obrázek 6.

Logování distribuce může zahrnovat i informaci o formě, jakou byl obsah/dokument odeslán.



Obr 6. Odběratelé obsahu (autorka)

6. Využitelnost vzoru

Model na obrázku 7a shrnuje problematiku návaznosti obsahů/dokumentů, model na obrázku 7b shrnuje oprávnění a evidenci obsazení rolí k obsahu/dokumentům, model na obrázku 7c shrnuje problematiku věcného tématu obsahu/dokumentů.

Některé CMS neodlišují typy a sorty obsahu/dokumentů. Rozlišení navržené v předloženém vzoru umožňuje znovupoužití šablon struktury (zde nazvaných typy obsahu/dokumentů) pro různé účely (zde reprezentovaných sortami). Práva účastníkům nakládat s obsahem/dokumentem se odvíjejí jednak od meritů věci (sorty objektů/dokumentů), jednak od jejich kvalifikace (typ obsahu/dokumentu). Tyto dva aspekty jsou nezávislé. Obvyklý systém rolí redakčního systému je proto vhodné rozšířit do dvou ortogonálních systémů (co kdo umí a co kdo smí).

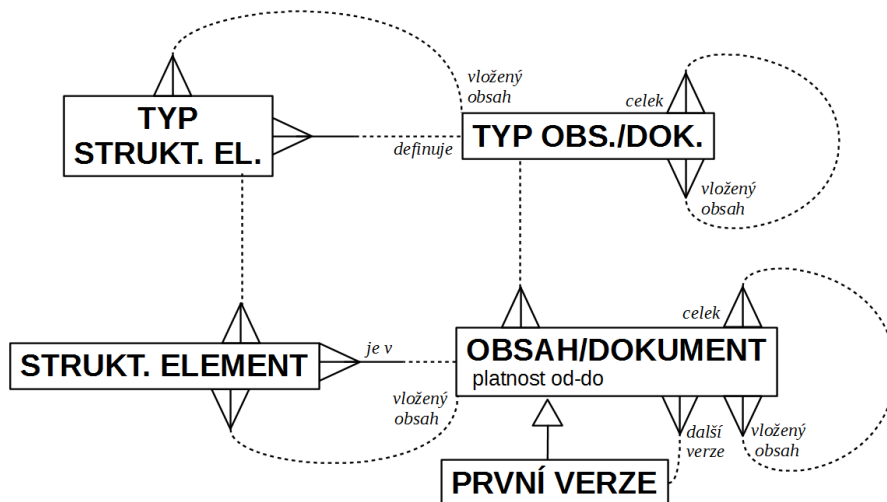
Odlišení typů obsahu/dokumentů a sort umožňuje zefektivnit návrh šablon vzhledu obsahu/dokumentů, kdy odděleně řešíme technickou nasaditelnost šablony a vzhled pro daný účel.

Systémy tagů používají obvykle jediný druh závislosti, a to "širší-užší význam". Sémantika skutečných vztahů témat je však bohatší, tak jak to zachycuje předložený vzor. V závislosti na potřebách aplikační oblasti je možno zvolit jednodušší či složitější řešení, počínaje systémem tagů bez vzájemných vztahů, přes hierarchie tagů, složitější vazby témat, až ke generickému řešení zde předloženého vzoru.

Závislosti verzí obsahu/dokumentů lze řešit konvencí v identifikátorech, kdy první verze nese název odvozený z cíle či účelu a další verze k tomu přidávají označení zahrnující závislosti verzí na sobě. Vytvoření složené identifikace verzí pak umožňuje zachytit jak jejich vztahy, tak příslušnost k záměru první verze. Jiné systémy, jež usilují o efektivní nakládání s verzemi, se zabývají změnami od verze k verzi, v takovém systému je nutné návaznost verzí pevně evidovat.

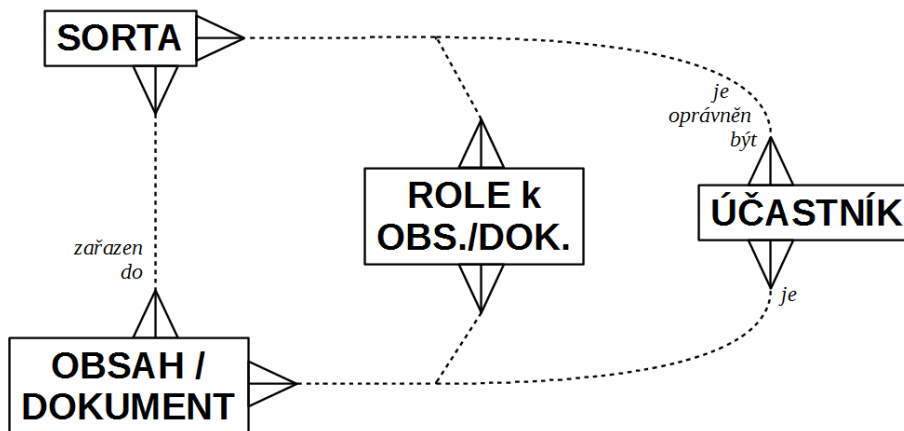
Strukturální předpisy obsahu/dokumentů (zde nazvané typy obsahu/dokumentů) lze řešit podobně jako metadata strukturované databáze (jako “katalog dat”), či prostředky obvyklými pro systémy semi-strukturovaných dat (jako xml schéma, regulární výrazy...)

Předložené řešení se nezabývá otázkami archivace a správy kopií, ty spadají do problematiky spisové služby.



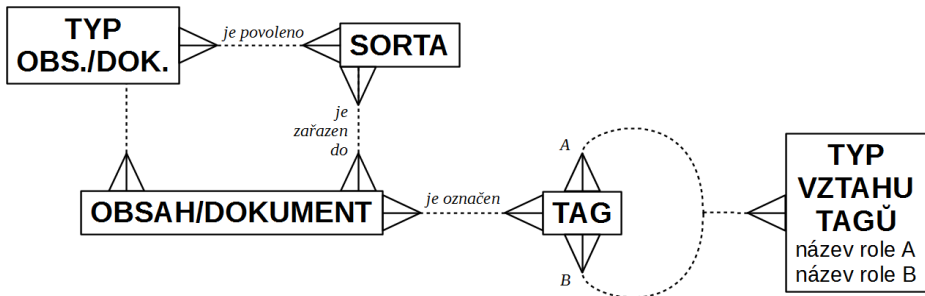
Obr 7a. Návaznost obsahů či dokumentů (autorka)

V obsahu/dokumentech zařazených do sorty ... je účastník ... oprávněn být k nim v roli...



V obsahu/dokumentu ... je účastník ... v roli...

Obr 7b. Oprávnění a evidence obsazení rolí k obsahu/dokumentům (autorka)



Obr 7c. Věcné téma obsahu/dokumentů (autorka)

Literatura

Blaha, M., 2010: *Patterns of Data Modeling*, 1st ed., CRC Press

Hay, D.C., 1996: *Data model patterns : conventions of thought*, 1st ed., Dorset House

Palovská, H., 2011: *Dvoutvářné entity* [Online]. Dostupné na:

<http://krokodata.vse.cz/DM/DveTvare>

Palovská, H., 2014: K čemu jsou dobré databázové návrhové vzory?, *Systémová integrace*, 21 (4), pp. 90-95. [Online]

Dostupné na: http://www.cssi.cz/cssi/system/files/all/SA_2014_04_08_Palovska.pdf

Palovská, H., 2015: Řízení přístupových práv – databázový návrhový vzor, *Systémová integrace*, 22 (4), pp. 46-52. [Online]

Dostupné na: http://cssi.cz/cssi/system/files/all/SI_2015_4_04_Palovska.pdf

Palovská, H., 2016a: Plánování a evidence skutečné spotřeby – databázový návrhový vzor, *Systémová integrace* 23 (1), pp. 66-73. [Online] Dostupné na: http://www.cssi.cz/cssi/planovani_evidence_skutecne_spotreby_databazovy_navrhovy_vzor

Palovská, H., 2016b: Evidence průběhu pracovních procesů – databázový návrhový vzor, *Systémová integrace* 23 (4), pp. 74-82. [Online] Dostupné na:

http://www.cssi.cz/cssi/system/files/all/SI_2016_3-4_Palovska.pdf

Palovská, H., 2017: Užívání vzácných zdrojů – databázový návrhový vzor, *Systémová integrace* 24 (1-2), pp. 87-95. [Online] Dostupné na:

http://www.cssi.cz/cssi/system/files/all/SI_2017_1-2_Palovska.pdf

JEL Classification: M10, C88