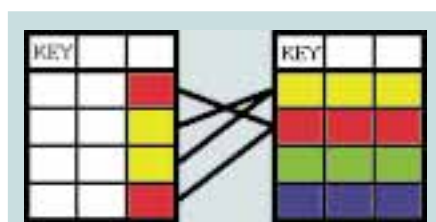


Databáze standardu SQL, díl 4.

Není vazba jako vazba

V posledním díle našeho seriálu jsme se věnovali dokončení části o normálních formách tabulek a začali jsme problematiku entit a relací. Pokračujme!

Připomeňme si graficky, jak vypadá relace třetího typu, tedy typu N : 1 – vidíme ji znázorněnou na obrázku 1.

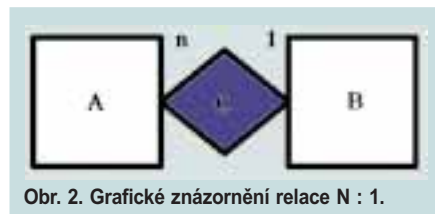


Obr. 1. Relace N : 1 mezi dvěma tabulkami.

Tabulku vpravo si můžeme představit jako vhodný číselník organizovaný podle unikátního klíče. Jednotlivé řádky, položky tabulky, jsou výrazně barevně označeny tak, aby bylo zřejmé, že každý řádek tabulky má jednoznačnou hodnotu klíče a je jednoznačně tímto klíčem určen. Tabulka vlevo se pouze odkazuje do tohoto číselníku. Vztah N : 1 je typický pro dvě tabulky, z nichž tabulka připouštějící víceznačnosti nemá obecně charakter číselníku, zatímco tabulka nepřipouštějící víceznačnosti má charakter číselníku s unikátním klíčem. Na obrázku 1 je barevně znázorněn vztah mezi tabulkami. V tabulce vlevo ve třetím, tedy neklíčovém sloupci jsou uvedeny odkazy na klíče v pravé tabulce. Odkazy ukazují na různé hodnoty klíče a reprezentují vztah mezi oběma tabulkami. Prakticky můžeme ukázat relaci mezi dvěma entitami na vztahu entity ZVIRE a entity DRUH. Entita ZVIRE popisuje jednotlivá zvířata zoologické zahrady z hlediska jejich evidenčních čísel, křestních jmen a čísel druhu. Klíčem v této tabulce je evidenční číslo zvířete, druhou položkou je křestní jméno a třetí pak číslo druhu zvířete. Entita DRUH je popsána tabulkou, kde unikátním klíčem je číslo druhu, v druhém sloupci je český název druhu a ve třetím sloupci jeho latinský název. Pak je jasné, že obě dvě tabulky jsou v 5NF a existuje mezi nimi vztah N : 1. Tento vztah je velmi přijatelný, protože v první tabulce může být několik zvířat jednoho druhu (například několik tygrů indických) a ta se odkazují na jednu jedinou položku tabulky druhého prostřednictvím čísla druhu. Cesta od dat v tabulce zvířat do tabulky druhů je velmi

jednoznačná a technologicky snadná. Naproti tomu cesta z tabulky druhů do tabulky zvířat je víceznačná a časově náročná. Pro nás je však podstatné, že obě dvě tyto cesty mají pro databázové systémy smysl. Od praktického databázového systému očekáváme, že bude umět tento vzájemný vztah entit co nejrychlejším způsobem, tedy v co nejkratším čase, realizovat. Je dobré umět vhodným a názorným způsobem popsat relaci mezi entitou ZVIRE a entitou DRUH. Velmi vhodné bude naučit se takovéto vztahy znázorňovat graficky. Na obrázku 2 vidíme znázornění vztahu N : 1 pomocí nově zavedených symbolů.

Zde jsme písmenem A označili entitu ZVIRE, písmenem B je označena entita DRUH a písmenem C je označena relace mezi zvířaty a druhy.



Obr. 2. Grafické znázornění relace N : 1.

Tuto relaci budeme později záměrně vyjadřovat slovem. Všimněte si znázornění poměru N : 1 pomocí symbolů u znaku relace. Tím je jednoznačně určeno, kterým směrem je vyhledávání snazší. Jsem-li totiž na vhodné položce v tabulce entity A, potom prostřednictvím relace jednoznačně určím, která je odpovídající položka v tabulce popisující entitu B.

Ideální vztah je N : 1

Jistě jste si položili otázku, proč je zrovna tak výhodné vytvářet relace typu N : 1 v porovnání s relacemi 1 : 1 nebo M : N. Hlavní důvod je v rychlosti přístupu k informacím mezi tabulkami. Pokud známe evidenční číslo zvířete a chceme zjistit co nejrychleji, jak se vlastně jmenuje v tabulce DRUH, pak máme celou řadu možností. Pokud by kartotéka druhů byla v papírové formě, byl by v ní chaos a nebyla by seřazena podle čísel druhů, potom hrozí, že v takové kartotéce budeme hledat velmi dlouho, než zjistíme, o které zvíře vlastně jde. Takové kartotéky většinou ani v nepočítačové

formě neprovozujeme. Obdobně tomu bude i v rámci SQL, kde se dává přednost organizovaným informacím s rychlým přístupem. Pokud je kartotéka seřazena podle rostoucích čísel druhů, pak stačí obrazně řečeno píchnout doprostřed a posoudit, jestli číslo druhu v polovině kartotéky je větší, menší nebo stejné než číslo druhu zvířete, které hledáme. Tím se celý problém rozpadne na dva poloviční problémy, které řešíme stejnou metodou, tedy píchnutím do poloviny. Latinské přísloví k tomu už 2000 let pragmaticky navádí slovy: Divide et impera (Rozděl a panuj). Malý příklad jen dokreslí efektivitu uvedeného postupu.

Máme-li kartotéku obsahující 100 000 druhů zvířat seřazenou podle čísla zvířete, potom systematické prohledávání od začátku do konce vyžaduje v průměru 50 000 porovnání hodnot čísla druhu v kartotéce s hodnotou hledanou. V optimistickém případě by sice už první porovnání mohlo vést k úspěchu, ale nejen pesimista musí počítat s prohledáním celé tabulky. Střední hodnota počtu porovnání je přesně uprostřed. Tabulka druhů je však seřazena podle čísla druhu – první půlení metodou rozděl a panuj pak tedy vede na hledání v 50 000 položkách. Druhé půlení zúží problém na 25 000 a po sedmnácti půleních nám zůstane už jen jedna hledaná položka. Místo 100 000 porovnání v nejhorsím případě tak provedeme maximálně 17 porovnání, což představuje zkrácení doby vyhledávání na 0,017 %!!! Nedivme se proto moudrým lidem a dokonalým databázovým systémům, že uchovávají informace v rozumně organizovaném tvaru. Moderní informační technologie umožňují dokonce s využitím technik HASH zajistit přístup k datům podle unikátního klíče prostřednictvím jedné až tří operací, což představuje zkrácení doby hledání v uvedené tabulce na 0,003 %!!! Počet přístupů na disk vůbec nesouvisí s počtem uchovávaných dat. Často také slyšíme o B-stromech nebo o digitálních stromových strukturách, které rovněž usnadňují přímý přístup k datům. Pokud vý-



Obr. 3. Lidé a jejich trvalé bydliště.

sledkem analýzy jsou tabulky v 5NF a relace N : 1, máme zejména v SQL velkou šanci k využití těchto silných a rychlých vyhledávacích technologií. Věnujme se nyní v klidu analýze relací mezi jednotlivými entitami.

Jazykový koutek

ER model je založen na jednoduchém pravidle spočívajícím ve splnění čtyř podmínek:

1. Vyjádřit fakta pomocí tabulek v 5NF.
2. Tabulky představující entity pojmenovat výstižným podstatným jménem v jednotném čísle.

3. Mezi entitami vytvořit relace typu N : 1 a výstižně je pojmenovat slovesem nebo předložkou.

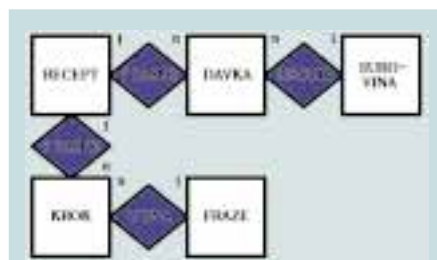
4. Čteme-li slova označující první entitu, relaci od ní a druhou entitu ve směru od N k 1, pak musí takto sestavená věta dávat smysl.

Splněním uvedených pravidel v rámci modelované reality získáme dokonalou analýzu. Na obrázku 3 je první ukáзка volby vhodných názvů v ER modelu.

Je zde znázorněn vztah mezi třemi entitami: CLOVEK, BYDLISTE a POSTA. Entita CLOVEK je dána tabulkou jednotlivých lidí v 5NF s přímým přístupem podle unikátního rodného čísla. Entita BYDLISTE je opět reprezentována tabulkou v 5NF a představuje trvalé bydliště. Entita POSTA představuje tabulku pošt v 5NF s unikátním poštovním směrovacím číslem. Přitom v jednom trvalém bydlišti může bydlet několik lidí, ale jeden člověk má právě jedno trvalé bydliště. Proto je relace mezi entitou člověk a bydlištěm typu N : 1 a velmi výstižně se dá charakterizovat slovesem MA. Člověk skutečně má trvalé bydliště. Zde je důležité, aby název entity, která je v rámci relace víceznačná, tvořil podmět příslušné věty o tom, že někdo někde bydlí. Relaci nelze číst v obráceném směru. Věta Bydliště má člověka neodpovídá realitě, protože jedno trvalé bydliště může být obsazeno více lidmi. Obdobně je to se vztahem entit BYDLISTE a POSTA. Bydliště MA příslušnou poštu, na kterou je třeba odesílat dopisy a nikoli obráceně. Na obrázku 3 vidíme, že analýza byla provedena velmi precizně, protože existují jednoznačné názvy entit a relací, které mají správný směr a dávají z hlediska českého jazyka smysl. Navíc je zřejmá hierarchická struktura, kde číselník pošt je na nejvyšší pozici, pod ním je číselník jednotlivých míst, kde lze trvale bydlet, a pod tím je teprve číselník lidí majících trvalá bydliště.

Číselníky kralují

Pokud budeme zásadně dávat přednost relacím N : 1, pak zcela nutně druhá entita v relaci musí mít charakter číselníku s přímým přístupem podle unikátního klíče. Z toho plyne, že číselníky budou častými entitami databázových systémů nezávisle na charakteru řešené úlohy. Nyní se můžeme vrátit k příkladům rozebíraným z hlediska struktury tabulek v předchozích dílech seriálu a rozebrat příslušné entity a relace. Na obrázku 4 vidíme v grafické formě schéma kuchyňské databáze receptur.

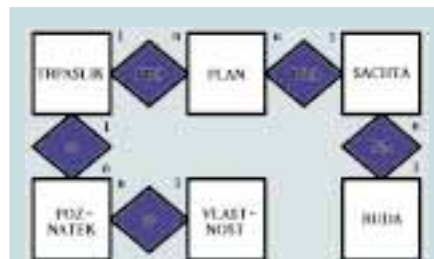


Obr. 4. ER model databáze kuchyňských receptur.

Jsou zde použity následující entity: RECEPT jako číselník receptů, SUROVINA jako číselník surovin, FRAZE jako číselník kulinařských frází, spojovací entita DAVKA určující dávkování surovin v rámci jednotlivých receptů a spojovací entita KROK, která říká, v jakém kroku kterého receptu se používají jednotlivé fráze. Nyní stačí přečíst ve správném pořadí od N k 1 všechny relace mezi entitami. První relace je mezi entitami DAVKA a RECEPT a je možno ji komentovat větou: Dávka V RAMCI receptu. Druhou relaci lze komentovat větou: Dávka URČUJE surovinu, jak jinak než jejím druhem. Další relaci je možno popsat větou: Krok V RAMCI receptu, protože jeden krok pracovního postupu se týká vždy jednoho receptu. Poslední relace je dána větou: Krok UZIVA fráze, neboť v každém kroku popisu je užitá jedna fráze. Všimněte si toho, že entita DAVKA je podřízena dvěma stranám. Na jedné straně entitě SUROVINA a dále také entitě RECEPT. Podobně entita KROK je podřízena entitám RECEPT a FRAZE.

Sněhurčino pozorování mrňousů a plánování jejich důlní činnosti lze rovněž popsat ER modelem. Na obrázku 5 vidíme v návaznosti na úvahy z předchozích dílů definitivní verzi databázového systému s několika drobnými vylepšeními.

Sněhurka se rozhodla definovat několik entit. Entitu TRPASLIK jako číselník trpaslíků orientovaný podle jejich evidenčních čísel,



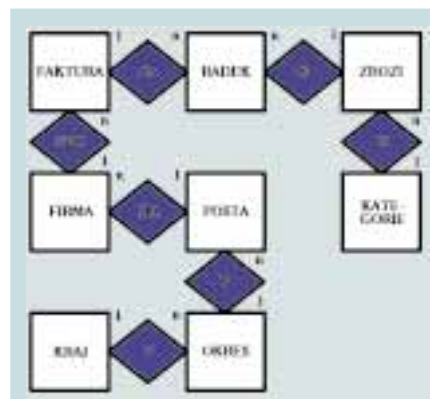
Obr. 5. ER model Sněhurčina impéria.

entitu VLASTNOST jako číselník trpasličích vlastností podle čísla vlastnosti, entitu SACHTA jako číselník míst, kde se těží, a entitu RUDA jako číselník rud organizovaný podle identifikačního čísla rudy. Dále Sněhurka zavedla pomocnou entitu PLAN podřízenou entitám TRPASLIK a SACHTA a poslední pomocnou entitu POZNATEK podřízenou entitám TRPASLIK a VLASTNOST. Entita PLAN je dána tabulkou obsahující den, číslo trpaslíka, číslo šachty, plánovaný a skutečný výkon. Entita POZNATEK je realizována tabulkou, kde je uvedeno pouze číslo trpaslíka a číslo vlastnosti. Opět stačí pojmenovat jednotlivé relace mezi entitami ve správném směru. O jednom trpaslíkovi může totiž být více poznatků, takže je zřejmé, že relace mezi entitou POZNATEK a entitou TRPASLIK je poznatek O trpaslíkovi. Entita POZNATEK má vztah i k entitě VLASTNOST. Tato relace je poznatek O trpasličí vlastnosti. Další jedno-

duchá relace je mezi entitou SACHTA a RUDA, kde šachta slouží NA těžbu rudy. Entita PLAN je opět podřízena entitám SACHTA a TRPASLIK. Plán je PRO trpaslíka v jedné relaci, zatímco tentýž plán je i PRO šachtu, na které se těží. Na obrázku 5 je celý databázový systém, kde nejnižší v hierarchii stojí entity POZNATEK a PLAN, zatímco entity VLASTNOST, TRPASLIK, SACHTA a RUDA mají charakter číselníků, přičemž číselník rud je nadřazen číselníku šachet.

Obyčejná faktura

Představme si obyčejnou fakturu, kterou vystavuje firma vyrábějící a dodávající potravinářské zboží do různých prodejen. Z takového

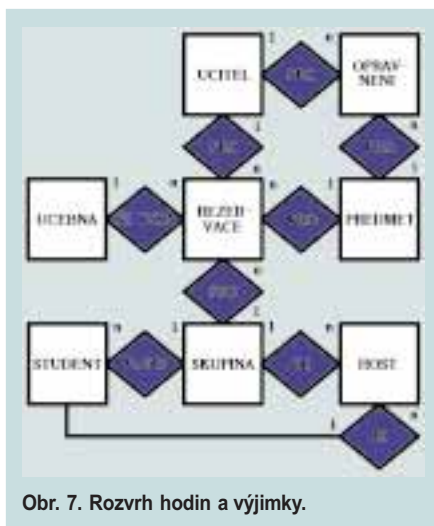


Obr. 6. ER model fakturace.

faktury můžeme přímo a nepřímo vyčíst různé informace, které by měly sloužit k rozhodování a k dalším rozborům. Vždyť z umístění firmy můžeme také nepřímo pochopit, v jakém se nachází okrese nebo kraji, a můžeme tak prognózovat spotřebu našeho sortimentu zboží podle regionů, což je velmi cenná informace, která nás může vést k zamyšlení, zda se nějaké regiony něčím liší nebo zda nás někde neohrožuje naše konkurence. Bude patrně nutné jednotlivé údaje na faktuře uvedené uvést do několika různých tabulek. Na posledním obrázku 6 vidíme situaci po dekompozici do ER modelu. Zvidavý čtenář pochopí nutnost zavedení entity RADEK, která popisuje jednotlivé řádky faktury, jež mají vztah nejen k entitě FAKTURA, ale i k entitě ZBOZI. Zavedení entit OKRES, KRAJ a KATEGORIE zboží usnadní tvorbu statistických přehledů pro potřebu marketingu.

Hvězdicovité uspořádání

Doposud uvedené příklady ER modelů měly velmi jednoduchou strukturu a jednotlivé entity a relace střídaly v přímých řetězcích. Realita je ovšem poněkud rozmanitější. Při konkrétní analýze problému a tvorbě ER modelu mohou entity a relace vytvořit hvězdicovité struktury, cykly, vícenásobná propojení nebo rekurentní spojení uspořádání. Běžný rozvrh hodin v rámci konkrétní školy je praktickou ukázkou jak hvězdicovitého, tak cyklického uspořádání entit a relací. Na obrázku 7 můžeme vidět strukturu ER modelu rozvrhu hodin.



Obr. 7. Rozvrh hodin a výjimky.

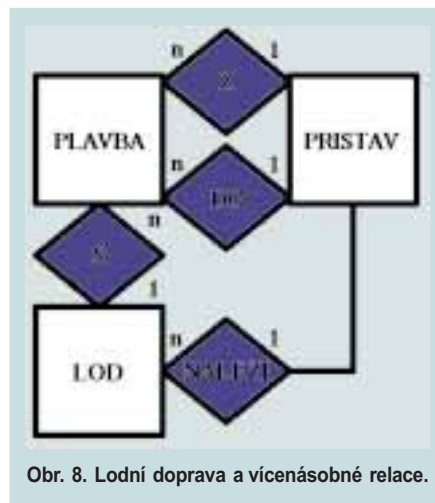
Nejprve uvedu výčet jednotlivých entit. Entita UCITEL obsahuje číslo učitele jako unikátní klíč, jméno učitele a osobní údaje typu datum narození, stav a podobně. Entita UCEBNA obsahuje číslo učebny jako unikátní klíč, její název a lokalizaci v rámci školy. Entita STUDENT obsahuje číslo studenta jako unikátní klíč, jméno studenta a číslo studijní skupiny, do které patří. Entita SKUPINA představuje číselník studijních skupin obsahující číslo skupiny jako unikátní klíč a její název. Entita PREDMET je číselníkem vyučovaných předmětů obsahujícím unikátní čísla předmětů a jejich názvy. Velmi zajímavá je entita OPRAVNENI, obsahující pouze číslo učitele a číslo předmětu, který je oprávněn učit. Jde o spojovací entitu v 5NF s unikátní dvojicí sloupců. Taková vazba může sloužit ke kontrole, zda všichni, co něco učí, na to mají konkrétní oprávnění. Další spojovací entita s názvem HOST popisuje hostování studentů v jednotlivých skupinách. V tabulce je uvedeno pouze číslo skupiny a číslo hostujícího studenta z jiné skupiny. Taková dvojice popisuje individuální zápis studenta do skupiny, ke které kmenově nepatří. Takto se chovají běžně studenti vysokých škol při zanedbání povinností nebo při širším studijním zábrhu. Poslední entita, která vytváří hvězdicovité uspořádání celého ER modelu, je entita REZERVACE, která nám připomíná onen tradiční barevný lísteček v ručně vytvářeném rozvrhu hodin, na kterém je napsáno vše potřebné. Rezervace tedy obsahuje zkratku dne v týdnu, pořadové číslo vyučovací hodiny, číslo učebny, číslo učitele, číslo předmětu a číslo skupiny studentů, která touto rezervací má zajištěno vše potřebné. Na obrázku 7 si samostatně prostudujte názvy jednotlivých relací ve správném směru, abyste si uvědomili význam pojmenování relací v ER modelu. Jiný kraj, jiný mrav. Ve škole, kam jste chodili, byly možná jiné zvyklosti, hostování bylo zakázáno nebo to byla pouze dohoda mezi učitelem a žákem. To je první námět pro vaše zamyšlení a tvorbu. Co kdyby bylo třeba evidovat tituly jednotlivých učitelů? Každý občan, tedy i učitel, může mít několik titulů. Pokuste se rozšířit ER mo-

del i o problematiku titulů učitelů. Konečně třetí námět k zamyšlení vyplývá z příslušnosti učitele i studenta k lidskému rodu. Pokuste se nahradit entity UCITEL a STUDENT entitou CLOVEK a navrhnete strukturu relací.

Vícenásobné relace

Mezi dvěma entitami může být stanoveno několik různých relací. Na obrázku 8 vidíme ER model z oblasti lodní dopravy. Jde o popis vztahů mezi loděmi a přístavy.

Lodě totiž neustále plují z jednoho přístavu do druhého a navíc jeden z těchto přístavů mají jako svůj mateřský. Pokud bychom nezavedli kromě entity LOD a entity PRISTAV též entitu PLAVBA, těžko bychom mohli učinit rozumnou dekompozici systému. Entita LOD je číselníkem lodí s unikátním evidenčním číslem, kde u každé lodi je též její název a číslo mateřského přístavu. Entita PRISTAV je číselníkem přístavů s evidenčním číslem přístavu jako unikátním klíčem, s názvem přístavu, zeměpisnou délkou a šířkou. Entita PLAVBA určuje, jaká loď a mezi kterými přístavy pluje v daném časovém období. Na čtenáři je, aby navrhl strukturu takové tabulky v 5NF s vhodně složeným unikátním klíčem. Relace mezi entitami LOD a PRISTAV se týká příslušnosti lodí k mateřským přístavům. Relace mezi entitami PLAVBA a LOD popisuje spojení plaveb s příslušnou lodí. Všimněte si, že relace mezi entitami PLAVBA a PRISTAV je dvojnásobná. Vícenásobnost relace mezi dvěma entitami je dovolena, pokud má každá z relací jiné jméno. V ER



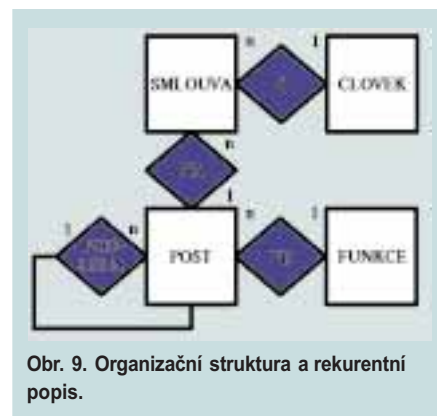
Obr. 8. Lodní doprava a vícenásobné relace.

modelu lodní dopravy je celkově realizována trojnásobná vazba mezi lodí a přístavem – plavba probíhá z přístavu, plavba probíhá do přístavu, loď náleží přístavu a plavba se koná s lodí.

Rekurentní relace

Pokud jste někdy chodili do práce, pak jste možná zjistili, že jste jen malým kolečkem nebo velkým kolem, které personálnímu oddělení pomáhá zaplnit organizační schéma. Každá větší firma má svého systemizačního pavouka, který popisuje vzájemnou přímou podřízenost pracovních postů. S výjimkou generálského postu má každý post jednoznač-

ně určen svůj nadřízený post, a tak je popsána celá struktura firmy jako nepravidelně rozvětvený strom. K popisu jednotlivých postů je třeba znát jejich označení. To je dáno poměrně omezeným repertoárem názvů vykonávaných funkcí, jako například technolog, svačínář, kontrolor, analytik nebo asistent manažera. Vztah jednotlivých „lidiček“ k postům, které zastávají, je dán pracovními smlouvami. Pro demonstraci rekurentní relace je na obrázku 9 uveden ER model organizační struktury firmy spolu se skutečným naplněním jednotlivých pracovních pozic jednotlivými lidmi.



Obr. 9. Organizační struktura a rekurentní popis.

Je zde uvedena entita CLOVEK jako číselník lidí podle rodných čísel, entita FUNKCE jako číselník pracovních funkcí podle jejich čísla a entita POST jako číselník pracovních pozic. Číselník POST zahrnuje unikátní číslo postu, číslo pracovní funkce a číslo nadřazeného postu. Uvědomme si, že entita POST se odkazuje sama na sebe, a tím je elegantně vybudována celá stromová struktura nadřízenosti postů. Poslední spojovací entita je entita SMLOUVA, která obsahuje pouze číslo postu, rodné číslo osoby, údaje o typu smlouvy, období trvání smlouvy. Co musí být v tabulce SMLOUVA unikátním klíčem, aby byla v 5NF? Počítejte s tím, že jistý John ve firmě začínal jako uklízeč, nyní je ředitelem vývoje, přitom ještě dělá svařeče na vedlejší pracovní poměr a neví se, kdy dostane padáka. Naše schéma navíc umožňuje neobsazení jednotlivých postů v rozporu s plánem, výkon funkcí se stejným názvem v rámci několika různých postů s respektováním nadřízenosti a podřízenosti. Mezi entitami SMLOUVA a CLOVEK je relace smlouva s člověkem. Mezi entitami SMLOUVA a POST je relace smlouva na pracovní post. Mezi entitami POST a FUNKCE je relace post je zastáván ve funkci. Novinkou je rekurentní relace entity POST sama na sebe, kdy post podléhá jinému, tedy nadřízenému postu. Nic složitějšího nás už nemůže v rámci ER modelu potkat.

Rekurentní vícenásobná relace

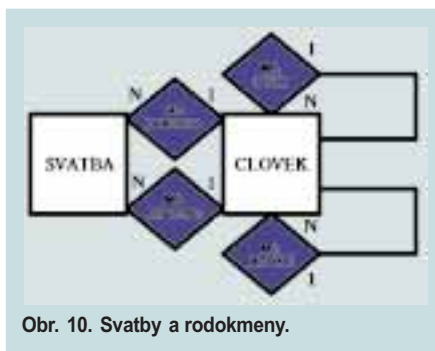
Jednoduché, vícenásobné a rekurentní relace spojují vhodným způsobem entity v rámci ER modelu. Přitom porozumění světu relací je klíčem k úspěšné dekompozici a posléze k velmi efektivnímu kladení SQL dotazů. Pro po-



Ná této straně je celostránková reklama!

vzbuzení chuti k analýze reálných problémů je zde předložen ER model rodokmenu. Omezené výrazové prostředky při grafickém znázorňování rodokmenů vedou tradičně jen k rozvíjení mužských větví. V nejlepším případě jsou ještě doplněny manželky a dcery hlavních mužských postav. Otcové a matky manželek spolu s manžely a dětmi dcer jsou jedno velké tabu. Na obrázku 10 je uveden ER model veškerých příbuzenských vztahů, který bez omezení zachycuje realitu.

K popisu stačí dvě entity a čtyři relace. Entita CLOVEK je číselníkem lidí podle identifikačního čísla. Kromě osobních dat je v něm uvedeno i identifikační číslo otce a identifikační číslo matky. To umožní snadné vytvoření dvojité vzájemně srostlé hierarchické struktury s užitím dvojnásobné rekurentní relace.



Obr. 10. Svatby a rodokmeny.

Entita CLOVEK se poprvé odkazuje sama na sebe pomocí otcovské relace. Podruhé se entita CLOVEK odkazuje sama na sebe pomocí mateřské relace. Pokud zavřeme oči nad některými úředními úkony, jsme s rodokmenem hotovi a můžeme v něm najít i nevlastní neteře starší 17 let. Zavedeme-li navíc entitu SVATBA, bude třeba vytvořit dvojnásobnou relaci k entitě CLOVEK. Každá tradiční svatba má právě jednoho ženicha a právě jednu nevěstu. Návrh příslušné tabulky v 5NF je ponechán na čtenáři. Dále můžete model rozšířit o svědka ženicha, svědka nevěsty a rozdíl mezi biolo-

OBJEKT	VLASTNOST	HODNOTA
Micka	je	siamská kočka
siamská kočka	je	šelma
šelma	je	savec
myš	je	hlodavec
hlodavec	je	savec
Micka	má ozdobu	mašle
mašle	barva	modrá
siamská kočka	barva očí	modrá
siamská kočka	ráda	myš
šelma	obživa	lov
hlodavec	obživa	zrní
savec	pije	mléko

Tabulka č. 2 : Kočičí sémantická síť.

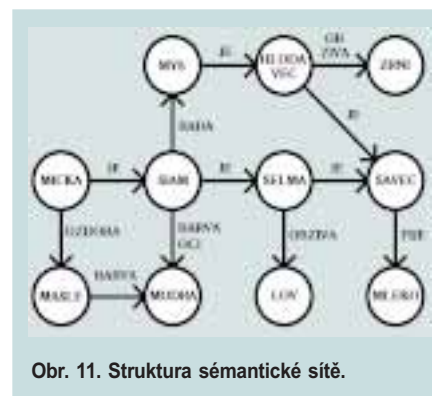
gickým a statutárním otcem. Pokud budete trvat na družičkách, pěstounech a adoptivních rodičích, pak budete nuceni ke změně analytického myšlení a uvědomění si podstaty a pomíjivosti vztahů nejen mezi lidmi.

ER model sémantické sítě

Marvin Minsky z MIT počátkem sedmdesátých let vytvořil teorii rámců a sémantických sítí, a tak prokázal, že každou lidskou znalost je možné vyjádřit unikátní trojicí: OBJEKT – VLASTNOST – HODNOTA. Vzpomeňme si na první třídu a naše první znalosti získané četbou a pozorováním okolí: „Ema má maso.“ „My se máme.“ „3 x 2 = 6.“ „Učitelka je chytřejší než žák.“ „Já jsem žák.“ Nebo „Spolužák Quido je vůl.“ V následující tabulce 1 vidíme rafinovanost zápisu těchto znalostí.

Asi je vám trochu podezřelé zavedení objektů příklad 7, povzdech 2 a skupina 3. Učíme-li se malou násobilku zpaměti, pak sedmý příklad je samostatným objektem, o kterém se ví, že jeho členové mají hodnoty dva a tři, výsledek má hodnotu šest a operace má hodnotu násobení. Druhý povzdech v první školní den jako objekt se týká hodnoty chytrosti, na prvním vítězném místě je učitelka a na druhém

místě třída objektů žák jako hodnota. Z toho, že já jsem žák a Quido je také žák, ani náhodou neplyne, že jsme spolužáci. Naše spolužáctví plyne ze společného členství ve skupině 3, která má za cíl učení se. Sémantické sítě se často znázorňují graficky pomocí orientovaného grafu. Jednotlivé uzly představují objekty, třídy objektů nebo hodnoty vlastností, což jsou rovněž objekty. Jednotlivé hrany před-



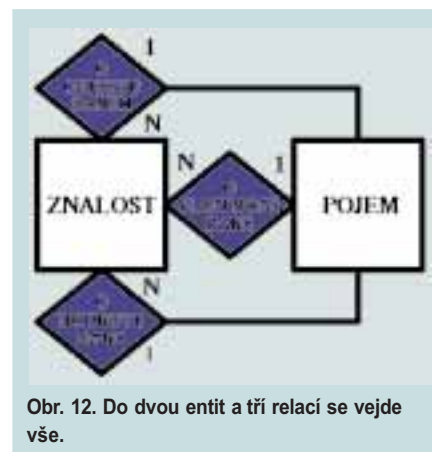
Obr. 11. Struktura sémantické sítě.

stavují vlastnosti objektů nebo tříd. Pokud hrana směřuje od objektu ke třídě, představuje vlastnost JE PRVKEM. Směřuje-li od třídy ke třídě, představuje vlastnost JE PODTŘÍDOU. Na obrázku 11 a v tabulce 2 je uvedena část kočičí sémantické sítě.

Nyní už patrně věříte, že sémantická síť dovede pomocí slovně vyjádřených trojic pojmů vyjádřit libovolná fakta. Je nezbytné zavést čí-

OBJEKT	VLASTNOST	HODNOTA
Ema	vlastnit	maso
my	mít se	dobře
příklad 7	člen	3
příklad 7	člen	2
příklad 7	výsledek	6
příklad 7	operace	násobení
povzdech 2	srovnání	chytrost
povzdech 2	první	učitelka
povzdech 2	druhý	žák
já	jsem	žák
Quido	je	žák
Quido	je	vůl
skupina 3	cíl	učení se
skupina 3	je	my
skupina 3	člen	já
skupina 3	člen	Quido

Tabulka č. 1 : Vzpomínka na první třídu.



Obr. 12. Do dvou entit a tří relací se vejde vše.

CISLO_POJ	TEXT_POJ
0	je
1	Micka
2	siamská kočka
3	šelma
4	myš
5	hlodavec
6	savec
7	mašle
8	má ozdobu
9	barva
10	modrá
11	barva očí
12	ráda
13	obživa
14	lov
15	zrní
16	pije
17	mléko

Tabulka č. 3 : Entita POJEM.

selník slovního vyjádření pojmů, tak jak je uvedeno na obrázku 12.

Entita POJEM představuje číselník dovolených pojmů s unikátním číslem pojmu, za kterým následuje text slovního vyjádření pojmu. Texty se mohou opakovat, přestože popisují různé pojmy. V sémantické síti klidně může být sedm Aniček a Anička může zlobit Aničku, aniž by zlobila sama sebe. Druhá en-

tita ZNALOST je potom tvořena unikátní uspořádanou trojicí čísel pojmů a celá trojice tvoří unikátní klíč. Trojnásobná relace mezi entitami představuje znalost o objektu daném pojmem, znalost o vlastnosti dané pojmem a znalost o hodnotě dané pojmem. V tabulce 3 je uvedeno konkrétní naplnění číselníku POJEM z kočičí sémantické sítě. Obdobně je v tabulce 4 uveden výpis spojovací entity ZNALOST.

Pokusme se odhadnout paměťové nároky velké sémantické sítě obsahující dva miliony pojmů a sto milionů znalostí. To odpovídá síti s miliardou hran a přibližně jedním milionem uzlů, kde z každého uzlu vede přibližně sto hran do jiných uzlů. Potom k očíslování pojmů stačí datový typ LONG INTEGER o délce čtyři bajty a textový popis pojmu může mít například délku 50 bajtů. Tabulka pojmů zabírá na disku zhruba 100 MB a tabulka znalostí má velikost zhruba 1,2 GB. K této základní spotřebě paměti je třeba přičíst ještě nejméně jednou tolik pro indexové soubory nezbytné pro rychlý přístup k pojmům a znalostem. Nyní se můžeme v jiném světle dívat na rodokmen jako na jednoúčelovou sémantickou síť, kde člověk coby objekt má vlastnosti otec a matka, jejichž hodnotami jsou jiní lidé. Každá svatba je samostatným objektem, jehož vlastnosti jsou ženich, nevěsta, svědek ženicha, svědek nevěsty, družba, družička, host, oddávající, úřad a datum. Vzhledem k unikátnosti trojice CIS_OBJ, CIS_VLA a CIS_HOD můžeme mít libovolný počet hostů, družiček

CIS_OBJ	CIS_VLA	CIS_HOD
1	0	2
2	0	3
3	0	6
4	0	5
5	0	6
1	8	7
7	9	10
2	11	10
2	12	4
3	13	14
5	14	15
6	16	17

Tabulka č. 4 : Entita ZNALOST.

a družbů na jedné svatbě. Jedna svatba je jeden uzel sémantické sítě, ze kterého vychází až několik set orientovaných hran. Hlavním cílem konstrukce sémantických sítí není jen skladovat informace v univerzálním tvaru, ale umožnit formulaci průřezových otázek s deduktivním odvozením odpovědi. Pokud budu mít nadbytek mléka a nedostatek myší, budu se zajímat o objekty, které pijí mléko a není o nich známo, že mají rádi myši. Takové objekty jsou savec, hlodavec, myš a šelma. Pokud bychom do sémantické sítě navíc přidali informace z první třídy a tvrzení „Učitelka je člověk“, „Žák je člověk“ a „Člověk je savec“, pak odpověď bude rozšířena ještě o objekty člověk, učitelka, žák, Quido a já.

Jaromír Kukul