

PROJEKTOVÁNÍ DATOVÉ ZÁKLADNY METODOU HIT MODELU

Jana Kubalová

1. Úvod

Otázka metodiky návrhu báze dat vyvstane zpravidla až poté, co jsou nalezeny odpovědi na otázky typu Zavést vůbec banku dat - ano či ne? a Jaký má být systém řízení báze dat? Jestliže se rozhodneme pro banku dat, ptáme se dále jaká data integrovat v datové základně, aby jich nebylo zbytečně mnoho, neboť se chceme vyvarovat uchovávání údajů duplicitních anebo redundantních a zároveň žádáme, aby veškeré požadavky, kladené na funkční způsobilost banky dat byly úspěšně splněny, jinými slovy - chceme zachovat právě ty prvky, které se budou účastnit žádaných vazeb.

Nastíněný problém pomáhá řešit metodika projektování HIT modelu, jakožto odpovídajícího obrazu sledované skutečnosti, transformované prostřednictvím pseudoatributů. Pseudoatributy definuje projektant na základě analýzy dialogů, vedených se znalcem modelované reality. Na základě pseudoatributů vytváří projektant datové jádro a vlastní model, přičemž se nespolehá na svou intuici v zásadní otázce, co je popis objektů a co je popis vztahů mezi nimi a nevyžadují se na něm žádné zkušenosti v projektování báze dat. HIT metodika projektování poskytuje dostatečně rozvinutý teoretický aparát, o který je možné se opřít a navíc formou hitovských diagramů nabízí jednoduchý společný jazyk pro rychlý účelný dialog znalce problému s projektantem modelu.

Účelem tohoto příspěvku není a neohce být vyzdvihování předností HIT metodiky projektování ani porovnávání této metodiky s jinými. Příspěvek demonstruje způsob aplikace této metodiky v praktickém návrhu databázové části, popisující členskou základnu konkrétního sledovaného systému, který by obecně vzato i v daném pohledu mohl být jakýkoliv, neboť je zde z pochopitelných důvodů omezení reprezentován velmi zjednodušeně jen dvěma základními prvky.

2. Teoretický aparát metodiky projektování HIT modelů

Pojem HIT model značí, že model bude homogenní, integrovaný a typový, tedy staví na typech, které se integrují do homogenní struktury na základě logických závěrů z dialogu mezi odborníkem a projektantem. Tyto závěry se zobrazí v množině atributů, z níž se pak podle jistých pravidel vybírají atributy a izolované objektové typy do dílčích datových jader. Tato jádra se pak slučují za současného vypouštění atributů a typů, vymykajících se pravidlům o integraci do datového jádra.

2.1. Klasifikace typů

Každý atribut vyjadřuje vztah mezi dvěma či více prvky, tzv. typy. Atributu n -tého stupně se zúčastní n typů. Typy se dělí na objektové - popisované a hodnotové - popisující, z nichž ty, u nichž známe úplný výčet hodnot jakých mohou nabývat, se nazývají číselníkovými.

2.1.1. Objektový typ

Objektový typ je výraz pro označení reálného objektu sledovaného na dané rozlišovací úrovni. V databázové části o členských základech jsme sledovali tři objektové typy: PRACOVNÍK, ORGANIZACE a PŘIHLÁŠKA. Objektový typ PŘIHLÁŠKA je spoluurčující pro objektové typy PRACOVNÍK a) nebo ORGANIZACE, protože typ ČÍSLO ve funkci primárního klíče typu PŘIHLÁŠKA může spoluurčit konkrétní objekt PRACOVNÍK a) nebo ORGANIZACE.

2.1.2. Hodnotový typ

Vlastnosti objektového typu, rozeznatelné na dané rozlišovací úrovni, se popisují příslušnými hodnotami. Tyto hodnoty se integrují v množině hodnotových typů. Objektový typ PRACOVNÍK je popisován hodnotovými typy DÁLNOPIŠ, TELEFON, ÚTVAR, (kde je pracovník zaměstnán) a JMÉNO (pracovníka). Objektový typ PŘIHLÁŠKA je popsán hodnotovými typy ČÍSLO (příhlášky) a DATUM (vyplnění příhlášky). Objektový typ ORGANIZACE je popsán hodnotovými typy ICOORG (kód organizace podle IČO, NÁZEV (příhlášeného členu - jednotlivými členy mohou být i závody, pobočky, odloučená pracoviště), ADRESA, AKTIV (vyjádření aktivity členu podle stanovené definice), RESQRT, VBJ.

2.2. Klasifikace atributů

Atribut A je funkce zobrazení množiny \mathcal{L} prvků jednoho typu do množiny prvků jiného typu, přičemž zobrazení je transformované z množiny reálných světů ω a z množiny časových okamžiků

\mathcal{T} :

$$A: (\omega \rightarrow (\mathcal{T} \rightarrow (\mathcal{L} \rightarrow B)))$$

Báze dat, kterou projektujeme, bude vyuklivána v podmínkách pro nás nominálního světa, přičemž v konkrétním časovém okamžiku budou v ní integrovaná data pro nás aktuální, proto se jakožto projektanti můžeme omezit na funkci

$$\mathcal{L} \rightarrow B$$

Tato funkce se v terminologii HIT metodiky nazývá pseudatribut, ale běžně se ustálil také pojem atribut. Jako každá jiná funkce má svůj definiční obor a obor hodnot, jakých nabývá. I bez podrobnější znalosti zkoumaného problému každý poměrně přesně určí, jakých hodnot může nabývat typ TELEFON, DATUM, ADRESA. Důležité je to, že HIT metodika připouští jen takové atributy, které zahrnují ve svém definičním oboru nebo oboru funkčních hodnot espoň jeden objektový typ. Požadavek splňující na příklad tyto atributy:

$$A1: (\omega \rightarrow (\mathcal{T} \rightarrow (\text{ORGANIZACE} \rightarrow \text{ICOORG})))$$

ICOORG dané ORGANIZACE; $0, n : 0, 1$

$$A2: (\omega \rightarrow (\mathcal{T} \rightarrow (\text{RESORT} \rightarrow (\text{ORGANIZACE} \rightarrow \text{BOO}))))$$

ORGANIZACE, spadající pod daný RESORT; $1, 1 : 0, n$

$$A3: (\omega \rightarrow (\mathcal{T} \rightarrow (\text{ORGANIZACE} \oplus \text{PŘIHLÁŠKA} \rightarrow \text{ÚTVAR})))$$

ÚTVAR, s něhož byla odeslána PŘIHLÁŠKA dané ORGANIZACE;
 $0, n : 1, 1$

$$A4: (\omega \rightarrow (\mathcal{T} \rightarrow (\text{ÚTVAR} \oplus \text{ICOORG} \rightarrow (\text{NABÍDKA} \rightarrow \text{BOO}))))$$

NABÍDKA, vyslovená ORGANIZACÍ s daným kódem IČO, potvrzená daným útvarem; $0, n : 0, m$

$$A5: (\omega \rightarrow (\mathcal{T} \rightarrow (\text{ICOORG} \rightarrow (\text{ÚTVAR} \oplus \text{NABÍDKA} \rightarrow \text{BOC}))))$$

aktivita organizace s daným kódem IČO, vyjádřená konkrétními NABÍDKAMI, potvrzenými v jednotlivých ÚTVARECH; $1, n : 0, m$

A6: (C) → (V → (ICOORG ⊗ KÓD ODPADU → (DATUM ⊗ NABÍDKA → BOO))))

aktivita dané ORGANIZACE hodnocená s ohledem na určitý odpad, daný KÓDEM ODPADU, vyjádřená NABÍDKOU potvrzenou v ÚTVARU; 0,n : 0,m

POZNÁMKA 1:

Poslední dva atributy spojují navíc dva dosud nedefinované prvky, a to objektový typ NABÍDKA ve významu nabízeného odpadu nějaké organizace, který je využitelný jakožto druhotná surovina v jiné organizaci a hodnotový typ KÓD ODPADU ve významu číselníkového typu, vyjadřujícího kódem klasifikací druhu odpadu.

POZNÁMKA 2:

Atributy A1 až A6 představují šest druhů atributů, které HIT metodika uznává z praktického hlediska.

Všech šest popsaných atributů je zobrazeno hitovskými diagramy v obrázkové části, kde je uveden znak))) , tzv. form. Tento znak indikuje, že hodnota atributu může být reprezentována n -tíci prvků příslušného typu. S tím souvisí i to, že atribut z oboru funkčních hodnot se zobrazí do množiny Booleovských hodnot BOO.

U popsaných atributů uvedená čísla ve tvaru a,c : b,d má význam klasického poměru atributu; a:b vyjadřuje dolní poměr, c:d udává horní poměr atributu. Konstanty a, b, c, d se pro atribut A, jehož obor funkčních hodnot je \mathcal{X}_A a definiční obor je \mathcal{D}_A , určí podle následující definice:

a = 1 právě tehdy, když každý prvek z \mathcal{X}_A je obrazem aspoň jednoho prvku z \mathcal{D}_A , jinak a = 0

b = 1 právě tehdy, když každý prvek z \mathcal{D}_A se zobrazuje spoň do jednoho prvku z \mathcal{X}_A , jinak b = 0

c = 1 právě tehdy, když každý prvek z \mathcal{X}_A je přiřazen maximálně jednomu prvku z \mathcal{D}_A , jinak c = n

d = 1 právě tehdy, když každý prvek z \mathcal{D}_A se zobrazí maximálně do jednoho prvku z \mathcal{X}_A , jinak d = m

2.2.1. Rozložitelnost atributů

Klasický poměr atributu indikuje důležité vlastnosti daného atributu, které nemusejí být vždy na první pohled patrné. Horní poměr mluví o tom, zda navržený atribut je rozložitelný nebo ne. Platí: jestliže $n = 1$ a zároveň $m = 1$ a zároveň v \mathcal{X}_A je více než jeden výskyt typů, nebo jestliže $n = 1$ a zároveň $m = 1$ a zároveň v \mathcal{D}_A je více než jeden výskyt typů, pak je atribut A rozložitelný.

Atributy A1 až A6 nejsou rozložitelné.

2.2.2. Odvoditelnost atributů

Možnost vzájemné odvoditelnosti dvou atributů je závislá na vybavenosti banky dat odpovídajícím aplikačním programovým vybavením (APV). Jestliže APV disponuje algoritmem, kterým lze atribut B získat z atributu A, pak atribut B je odvoditelný z atributu A. V zásadě platí, že je-li možné z atributu A složitosti s odvodit atribut B složitosti ne vyšší než $s - 1$ s horním poměrem $c : 1$ nebo $1 : d$, pak atribut A je rozložitelný a atribut B je z něj odvoditelný. Posouzení odvoditelnosti či neodvoditelnosti atributů je někdy náročné. Právě v tomto bodě i HIT metodika vyžaduje dobrou intuici nebo předchozí zkušenosti.

2.3. Podmínky výběru typů a vztahů mezi nimi do báze dat

Poté co jsme vyprojektovali atributy, které vystihují dostatečně přesně pohled na systém z hlediska znalce modelovaného problému a které jsou nerozložitelné a neodvoditelné navzájem, můžeme předpokládat, že jsme neopomenuli zahrnout do struktury žádné důležité údaje a také, že jsme vyloučili duplicitní vztahy mezi nimi.

S ohledem na výše uvedené se vytvoří čisté datové jádro za současného sjednocení názvů typů, odstranění nekonzistencí, snížení počtu objektových typů, (je-li to možné) a sloučení všech dílčích datových jader, (podílelo-li se na řešení více lidí). Z čistého datového jádra je nutno přesným postupem odvodit zhuštěné datové jádro, které znázorňuje vztahy mezi objektovými typy a počty atributů k nim se vázajících. Tímto požadavkem je řízen postup výběru typů i atributů do zhuštěného datového jádra. Toto

jádro reprezentuje konceptuální model. Ten dále podléhá transformaci do logického návrhu tzv. sítě typů rekordů neboli zobecněného implementačního schématu.

3. Závěr

HIT metodika projektování popisuje všechny vztahy a vlastnosti prvků modelu jako funkce, popis objektů a popis vztahů mezi nimi jsou dobře teoreticky zdůvodněny a určují se při transformaci modelu do vlastní implementace. Metodika doporučuje projektantovi ve všech fázích vývoje modelu konzultovat dosažené výsledky s odborníkem, který zná modelovaný problém. Odborníkem doporučené změny je dovoleno realizovat nejvýše na úrovni konceptuálního návrhu a v dalších transformacích postupovat podle pravidel HIT metodiky.

Ten, kdo je zvyklý pracovat způsobem Chenova modelu, mási více spoléhat na svou dobrou intuici. Ten, kdo by chtěl sestavit bázi dat jen na základě svých zkušeností - bez teoretického aparátu, měl by vědět, že přístup jakoukoliv metodikou je lepší než přístup intuitivní.

Seznam použité literatury

1. F. Krejčí, M. Pendrych:

Datový model HIT a jeho aplikace při projektování datových struktur (příručka pro experty a projektanty informačních systémů), výzk. zpráva, OVC VUT Brno, 1982

2. P. Materna, F. Krejčí, J. Zlatuška, O. Felix, J. Pokorný: HIT - databázový model. Sborník SOFSEM '81, vyd. ÚVT UJEP Brno a VVS Bratislava, 1981

3. M. Pendrych, F. Krejčí:

Hit - metodika projektování, OVC VUT Brno, 1984

