

# Dátové modely

© HALÁSZ, Jozef

**Zdroj:** RUSKO, Miroslav - HALÁSZ, Jozef, 2011: *Environmentálne orientované informačné systémy*. - Žilina: Strix, Edícia EV-64, Prvé vydanie, ISBN 978-80-89281-76-3, 220 s.

Vytváranie modelov je úzko spojené s ľudským poznávaním sveta, so systematickým objavovaním toho, čo nás obklopuje. Poznatky sú výsledkom pracovnej, spoločenskej a myšlienkovej činnosti ľudí. Pod pojmom poznatok obvykle rozumieme reprodukciu určitej vymedzenej časti reálneho sveta vrátane zákonitostí, ktoré v ňom platia. Ich bezprostrednou funkciou je prevedenie rozptýlených, nejasných predstáv a tušení do všeobecnej formy. Znalosti sú vzájomne previazané, meniteľné a doplniteľné štruktúry poznatkov. Znalosť niečoho znamená vlastniť jemu zodpovedajúcu reprezentáciu v podobe dostatočne verného a presného kognitívneho modelu<sup>1</sup>.

Pomocou modelov zoskupujeme naše poznatky do určitých celkov, pričom určujúcu rolu hrá tak úroveň nášho poznania, ako aj ciele spojené s vytváraným modelom. Vznikajú tak modely globálne a detailné, abstraktné a konkrétne. Veľmi rozmanité sú aj prostriedky na opísanie týchto modelov.

V technickej praxi je pojem modelu chápaný užšie, väčšinou vo väzbe na určitý konkrétny systém alebo proces. Na opis týchto modelov sa používajú umelé jazyky a grafické prostriedky.

**Dátové modely** sú prostriedkom pre poznanie, opísanie a organizáciu dát. Vznikli v súvislosti s potrebou špecifikovať štruktúru dát pri analýze a navrhovaní databázových a informačných systémov.

Pri tvorbe abstraktného modelu reality sa od jeho tvorca vyžaduje, okrem systematického postupu, aj veľa skúseností a intuície. Vytvorený abstraktný model odráža viac alebo menej verne tie prvky reality, ktoré tvoria predmet záujmu užívateľov dát.

Dátový model je matematický formalizmus umožňujúci dátovo orientovaný opis reality.

Dátový model môže byť spracovaný na rôznych úrovniach abstrakcie:

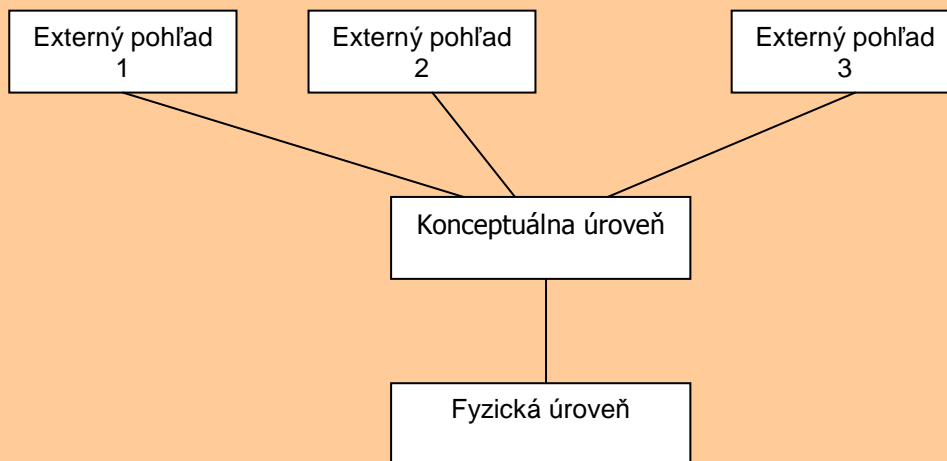
- fyzická úroveň je najnižšia úroveň abstrakcie - podrobne popisuje štruktúru záznamov a spôsob ukladania dát na pamäťových médiách, indexovanie a pod.
- konceptuálna úroveň popisuje dáta a ich vzájomné vzťahy.
- na najvyššej úrovni abstrakcie sa nachádzajú pohľady na informačný systém alebo databázu zo zorného uhla rôznych aplikácií alebo skupín používateľov.

Dátový model je tvorený množinou pojmov, ktoré uľahčujú špecifikáciu štruktúry dát a umožňujú vytvárať operácie pre manipuláciu s dátami.

---

<sup>1</sup> POPPER, M.- KELEMEN, J., 1988: *Expertné systémy*. Bratislava : Alfa. ISBN 80-05-00051-0

Väčšina dátových modelov obsahuje štruktúralnu a operatívnu časť a integritné pravidlá. Štruktúralna časť slúži na opis dátových typov (text, číslo, dátum a pod.) a ich väzieb. Operatívna časť opisuje manipuláciu s údajmi. Integritné obmedzenia môžu byť reprezentované jednoduchými pravidlami určujúcimi, ktoré položky napríklad nemôžu mať zápornú hodnotu, ale aj veľmi komplexnými logickými podmienkami.



Obr. 1 Úrovne abstrakcie dátových modelov

Prvé dátové modely sa využívali na navrhovanie súborov a báz dát. Hlavným štruktúrnym typom v nich bol záznam (record), ktorý sa skladal z položiek, prvkov (item, element) alebo z polí (field).

Veľké rozšírenie zaznamenal najmä hierarchický model dát, ktorý organizoval typy záznamov do stromovej štruktúry a sieťový model dát, ktorý organizoval typy záznamov ako uzly grafu. Ďalším významným vývojovým krokom bol relačný model, navrhnutý E. F. Coddom, ktorý oddelil logickú reprezentáciu dát od ich fyzickej implementácie a poskytol modelovaniu dát matematický základ zavedením pojmov množiny a relácie<sup>2</sup>.

Relačný model organizuje dáta do tabuliek (relácii). Tieto tabuľky obsahujú vo svojich riadkoch elementárne informácie (entity) a v stĺpcoch ich charakteristiky (atribúty).

Jednoduchosť relačného modelu, ktorá je jeho výraznou prednosťou pre tvorbu a činnosť báz dát, je súčasne nevýhodou na úrovni analýzy a návrhu informačného systému, pretože neposkytuje dostatok informácií o vzťahoch medzi dátami.

Tvorba informačných modelov na vyššej úrovni vyžaduje prostriedky, ktoré umožňujú jednoducho a prehľadne vyjadriť komplexné štruktúry a ktoré majú dobrú vypovedaciu schopnosť pre komunikáciu s používateľom. Pre túto úroveň sú výhodnejšie sémantické modely.

## Sémantické dátové modely

Sémantické dátové modely (SDM) opisujú údaje a ich spracovanie z hľadiska používateľa. Kladú dôraz na význam určitých údajov, čiže poskytujú vyššiu úroveň abstrakcie pre modelovanie reality. Pre človeka je samozrejmé, že je nezmyslom porovnávať napr. hmotnosť a evidenčné číslo určitého zariadenia, pričom sa jedná

<sup>2</sup> CODD, E., 1970: *A Relational Model for Large Shared Data Banks*. CACM 13(6) 1970

o syntakticky rovnaké (číselné) údaje. Líšia sa svojím významom - sémantikou. Spracovanie dát závisí od významu, preto je dôležité vytvoriť výstižné sémantické modely, ako východisko pre spracovanie podrobnejších dátových modelov.

Základnými komponentmi sémantických modelov sú objekty, typy objektov a ich charakteristika (vlastnosti). Dôležité sú aj vzťahy medzi objektmi a integritné pravidlá. Prostriedkom pre vytváranie nových (zložitých) typov objektov sú typové konštruktory.

## Entitno-relačný model

V oblasti analýzy a navrhovania informačných systémov sa najviac využíva entitno-relačný model (**ERM**).

Entitno-relačný model je modelom informačného systému v termínoch entít a vzťahov medzi entitami. Technika vychádza z predpokladu existencie „reálneho sveta“, ktorý možno v týchto termínoch modelovať.

Na znázornenie entitno-relačného modelu sa používa entitno-relačný diagram (ERD).

Diagramy vzťahov entít sú jednou z najstarších techník používaných pri tvorbe informačného systému. Svoje počiatky majú v teoretických prácach

- *Ch. Bachmana (1969)*
- *skupiny Codasyl Data Base Task Group (1971),*

ktoré sa zaoberali hierarchickými a sieťovými modelmi dát.

Súčasnú techniku a formu entitno-relačného diagramu navrhol v roku 1976 P. Chen<sup>3</sup>.

## Životný cyklus entity

Diagramová technika na znázornenie životného cyklu, či histórie života entity (entity life history - **ELH**) vznikla pôvodne s cieľom doplniť staticky pohľad ERD o dynamicky aspekt. V určitom zmysle bola ELH pôvodne konkurentom diagramu tokov dát (DFD), pričom bola užšie spojená s logickým modelovaním dát. V súčasnom období mnohé metodiky pripisujú ELH integrujúcu úlohu prostredníka medzi ERD a DFD.

Z hľadiska tradičného modelovania informačného systému pomocou ERD sa entity považujú za statické, časovo nezávislé objekty. V skutočnosti sú entity predmetom veľkého počtu procesov, ktoré menia ich stav.

*Napríklad v knižnici entita KNIHA môže prechádzať viacerými stavmi: najprv sa zadovážá, potom katalogizuje, požičiava čitateľom a nakoniec sa vyraduje.*

ELH je diagramová technika na vyjadrenie využitia určitej entity procesmi alebo udalosťami, ktoré vytvárajú ich systém. Užitočnosť ELH spočíva v tom, že pomáhajú analytikovi lepšie porozumieť entitám. Sú aj prvým krokom pri dokumentovaní podrobného priebehu procesov.

Vytvorenie ELH je postupnosťou nasledujúcich krokov:

- vyhotovenie zoznamu entít z ERD,
- vyhotovenie zoznamu procesov z diagramu dátových tokov,
- konštrukcia matice ELH,
- konštrukcia diagramu ELH pre každú entitu.

---

<sup>3</sup> CHEN, P., 1976: *The Entity-Relationship Model - Toward a Unified View of Data.* – In: ACM Transactions on Database Systems, Vol.1, No.1, 1976, s. 9-36

## Objektovo orientovaný dátový model

Vývoj dátových modelov smeruje k vytvoreniu nástrojov pre najvýstižnejšie a zároveň najprirodzenejšie opísanie reálneho sveta. V každodennom živote informácie o určitých objektoch vyjadrujú nielen ich parametre a charakteristiky, ale spájajú ich s určitým spôsobom spracovania, či zobrazenia. Taktiež vieme, že existujú objekty, ktorých niektoré vlastnosti sú odvodené od iných objektov. Napríklad väčšina organizácií eviduje zhruba rovnaké údaje o svojich pracovníkoch, ale tieto údaje sa ináč spracúvajú a zobrazujú na rôznych oddeleniach.

Myšlienka spájať údaje s metódami ich spracovania vznikla v programátorských kruhoch už pred desiatkami rokov a preto máme dnes dobre prepracované informačné technológie podporujúce objektovo - orientované spracovanie.

Objektovo - orientovaný dátový model tvoria:

- Objekt a identifikátor objektu. Každá entita reálneho sveta sa považuje za objekt, ktorý má jednoznačné pomenovanie. Objekt je samostatná a ďalej nedeliteľná entita, ktorá zapuzdruje dáta a obsahuje zároveň metódy, ktoré sú určené pre prácu s týmito údajmi.
- Atribúty a metódy. Každý objekt je v určitom stave (vyjadrenom množinou hodnôt atribútov) a chová sa určitým spôsobom (určeným metódami, procedúrami spracovania atribútov). Stav a chovanie zapuzdrené v objekte sú dostupné len prostredníctvom správ prenášaných medzi objektmi.
- Trieda je množina užívateľom definovaného typu objektov. Triedy zoskupujú všetky objekty, ktoré zdieľajú tú istú množinu atribútov a metód. Objekt vždy patrí do jednej triedy - je inštanciou danej triedy. Na triedu môžeme nazerať ako na abstraktný dátový typ. Trieda predstavuje šablónu (popis), ktorá slúži pre vytváranie objektov.
- Hierarchia a dedičnosť tried. Odvođením novej triedy (podtrieda) od existujúcej triedy (nadtrieda), táto zdedí všetky atribúty a metódy a navyše môže byť doplnená vlastnými atribútmi a metódami.

Objektové dátové modely majú veľký význam o.i. pre databázové aplikácie v CAD a GIS.

## Environmentálne dátové modely

Každý environmentálny objekt sa dá opísať súborom environmentálnych dátových objektov. Tieto objekty sú abstraktnými entitami, ktoré sa dajú spracovať na počítačoch alebo priamo použiť v procese rozhodovania kompetentnými pracovníkmi. Typickým environmentálnym dátovým objektom je súbor nameraných hodnôt charakterizujúci koncentráciu určitej látky v ovzduší alebo vo vode (odpovedajúci environmentálny objekt).

Environmentálny dátový objekt môže mať analógovú alebo digitálnu podobu. Treba si uvedomiť, že zber rôznych informácií o prírodnom prostredí, za účelom pochopiť určité procesy a mechanizmy pôsobiace v prírode, sprevádza ľudstvo počas celej histórie. Sledovanie hladiny vodných tokov, množstva úrody na poliach, počtu zveri v loveckých revíroch bolo vykonávané od pradávna. Meranie teploty ovzdušia a vody, záznamy o výdatnosti zrážok a pod., boli vykonávané v posledných storočiach už pravidelne. Jednalo sa o lokálne údaje, na ktorých komplexné spracovanie neboli k dispozícii ani teoretické a hlavne ani technické prostriedky. Počítačová technika dnes umožňuje hromadné spracovanie týchto údajov, z ktorých použitím metód matematickej štatistiky a na základe znalosti fyzikálno-chemických princípov sa dajú vytvoriť stále presnejšie modely environmentálnych objektov. Vzhľadom na počítačové spracovanie sú dátové objekty digitalizované, prevádzané do množiny číselných údajov.

Týka sa to aj objektov, ktoré boli tradične zobrazované graficky, napríklad máp a orientačných plánov.

Prvým krokom pri spracovaní dát je sústredenie objektov reálneho sveta do entít, ktoré sú abstraktné a môžu byť spracované počítačmi alebo priamo rozhodovacími subjektmi. V tejto práci sa sústredíme na tie objekty reálneho sveta, ktoré sú environmentálnymi objektmi. Z určitého hľadiska skoro na každý objekt reálneho sveta sa môžeme pozerieť ako na environmentálny objekt, nech sa jedná o živočíchov alebo objekty vytvorené človekom, či prírodou.

Automatizované environmentálne informačné systémy dokážu zozbierať a spracovať obrovské množstvo dát. Automatizovaný zber dát produkuje denne niekoľko terabajtov informácií, ktorých spracovanie tradičnými postupmi je nemožné. Moderné vyhodnocovacie systémy umožňujú poloautomatické alebo automatické filtrovanie, selekciu, katalogizáciu vstupných dát za účelom vymedzenia relevantnej množiny hodnôt. Pri tejto činnosti je výhodné používať tzv. metadáta, t.j. dáta o dátach. Metadáta opisujúce jednotlivé environmentálne dátové objekty budeme nazývať environmentálnymi metadátami.

Formát a obsah metadátových objektov je závislý na štruktúre environmentálnych dát. Množina nameraných hodnôt, ktorá sa vytvorila napr. meraním množstva určitej látky v ovzduší sa dá zhrnúť do tabuľky alebo grafu, sprevádzanej opisom metodiky merania, vyhodnocovania a pod. Na základe uvedeného sa dá podľa Günthera<sup>4</sup> vytvoriť trojúrovňový model environmentálnych informačných systémov.

Každý environmentálny objekt sa dá opísať súborom environmentálnych dátových objektov. Tieto objekty sú abstraktnými entitami, ktoré sa dajú spracovať na počítačoch alebo priamo použiť v procese rozhodovania kompetentnými pracovníkmi. Typickým environmentálnym dátovým objektom je súbor nameraných hodnôt charakterizujúci koncentráciu určitej látky v ovzduší alebo vodnom toku (odpovedajúci environmentálny objekt).

Každý environmentálny dátový objekt môžu opísať environmentálne metadáta.

Dátový tok sa dá rozčleniť na štyri fázy: zber, úschova, analýza dát a správa metadát. t.j.

- Prvá fáza, zber dát sa týka získavania dátových podkladov z prvotných zdrojov. Patria sem tak tabuľky, súbory nameraných hodnôt, ako aj súbory fotografií a snímok. V tomto kroku sa transformujú veľmi rozmanité environmentálne objekty na environmentálne dátové objekty, ktorých štruktúra môže byť jednoduchšia a prehľadnejšia ako štruktúra pôvodných dát. S transformáciou dát je spojené aj určité predspracovanie (agregácia, filtrovanie, štatistické metódy). V prípade obrazovej informácie sa môže jednať aj o automatickú identifikáciu jednoduchých útvarov, ako aj celých objektov. Výsledné dátové objekty sa dajú zakódovať v rôznych formátoch a uložiť v dátových súboroch alebo v databázach.
- Druhá fáza je spojená s uschovaním dát, pričom základom sú štandardné databázové prostriedky. Vzhľadom na heterogénnosť a komplexnosť environmentálnych dát je potrebné tieto prostriedky prispôbiť konkrétnej aplikácii, doplnením špecifických nástrojov. V prvom rade je potrebné vytvoriť mechanizmy spracovania priestorových dátových objektov. Vyžaduje to špecifikáciu nových dátových typov, ako aj nástrojov pre účinné ich spracovanie.

---

<sup>4</sup> GÜNTHER, O., 1998: *Environmental Information Systems*. - Berlin : Springer Verlag

- Tretiu fázu tvorí analýza dát, pri ktorej sa zo zhromaždených dát odvodzujú informácie pre podporu rozhodovacích činností. Analýza dát je tradične spojená s aplikáciou štatistických metód, ku ktorým sa v poslednej dobe pridáva vyťažovanie informácií z dátových skladov (data mining). Analýza je väčšinou cielená, čiže údaje sa zoskupujú a vyhodnocujú so zameraním na konkrétne oblasti a činnosti.
- Štvrtou fázou je tvorba metadát, v ktorej sa definujú abstraktné opisy dátových objektov používané v predchádzajúcich fázach.

Tento komplexný agregatívny proces má poskytovať informácie pre rozhodovanie na rôznych úrovniach, t.j.

- na nižších úrovniach, spojených s menším stupňom agregácie, sa dajú získané informácie použiť pri riešení operatívnych, taktických úloh.
- na vyššej úrovni sa jedná o strategické rozhodnutia vrcholových orgánov.

Podľa uvedeného môžeme špecifikovať charakteristické rysy environmentálnych informačných systémov:

- Množstvo spracovávaných dát je mimoriadne veľké, napr. objem dát z družicového snímkovania môže dosiahnuť denne aj niekoľko terabajtov.
- Správa dát je výrazne heterogénna a značne distribuovaná, pretože zberom, spracovaním a uchovávaním dát sa zaoberá veľký počet rôznych inštitúcií.
- Environmentálne dátové objekty majú dynamický charakter s rôznorodou škálou platnosti v priestore a čase a majú často veľmi komplexnú štruktúru.