

Adatbázisok - 1. előadás

Horváth Árpád <horvath.arpad@amk.uni-obuda.hu>

Óbudai Egyetem
Alba Regia Műszaki Kar (AMK)
Székesfehérvár

2015. október 15.

Köszönet

A tárgyat korábban Kottyán László tanította. Köszönöm neki, hogy az általa elkészített előadásanyagot megkaphattam tőle, és nem kellett nulláról indulnom a segédlet kidolgozásában. Köszönöm továbbá a német nyelvű Normalisierung wikipedia-szócikk példáit a szócikk szerzőinek.

Vázlat

- 1 Tematika
- 2 Fogalmak
 - Egyed-kapcsolat diagram
 - Relációs adatmodell
- 3 Használt szoftverek

Tematika (1)

- Adatbázisok tervezése
 - Relációs adatmodell
 - A tervezés folyamata
 - Fizikai modell elkészítése (PostgreSQL, psql, pgAdmin)

Tematika (2)

- Adatbázisok használata, kezelése
 - SQL
 - Nézetek, függvények, stb.
 - Python kliens használata

Tematika (3)

- Térbeli adatbázisok
 - OGC szabványok, fogalmak
 - Térbeli adatbázisok létrehozása (PostGIS)
 - Térbeli függvények
 - GIS kliens használata (QGIS)

Vázlat

- 1 Tematika
- 2 Fogalmak
 - Egyed-kapcsolat diagram
 - Relációs adatmodell
- 3 Használt szoftverek

Adatbázis

- Adatmodell által meghatározott struktúrába szervezett adatok összessége, amely szoftverrel kezelhető és perzisztens adattárolást biztosít.
- Alapvető adatbázis műveletek (CRUD):
 - Create - adat létrehozása, rögzítése az adatbázisban,
 - Read - adat kinyerése, olvasása az adatbázisból,
 - Update - adat módosítása, frissítése az adatbázisban,
 - Delete - adat törlése az adatbázisból.

Adatmodell

- Az adatmodell az adatbázis adatelemeinek szerkezetét, az elemek kapcsolatát és elemeken alkalmazható műveleteket írja le.
- Egy definíció szerint: *Az adatmodell egyedek, tulajdonságok és kapcsolatok halmaza, amely absztrakt módon tükrözi a valós objektumoknak, azok jellemzőinek, viszonyainak elvont kategóriáit.*

Adatmodellek

- (i) Hálós
- (ii) Hierarchikus
- (iii) Relációs
- (iv) Objektum-reláció
- (v) Objektumorientált
- (vi) Multidimenzionális
- (vii) Semi-structured
- (viii-) stb.

A többi nem kell ismerni.

Adatbázis-kezelő rendszerek típusai

- Irodai (pl. MS Access, LibreOffice Base)
- Beágyazott (embedded) (pl. SQLite, Berkeley DB)
- Adatbázis kiszolgálók (pl. Microsoft SQL Server, Oracle, DB2, MySQL, MongoDB, PostgreSQL)
- stb.

Vázlat

- 1 Tematika
- 2 Fogalmak
 - Egyed-kapcsolat diagram
 - Relációs adatmodell
- 3 Használt szoftverek

E-K diagram

- Peter Pin-Shan Chen, Egyed-kapcsolat modell
- Egyed-kapcsolat diagram :
 - A relációs adatbázisok modellezésére, tervezésére használható

Egyed

Az **egyed** minden olyan dolog, amely minden más dologtól megkülönböztethető és amelyről adatokat tárolunk. Egyedek például a következők: hallgató, tanár, tantárgy, autó, könyv. Az egyed konkrét dolgok absztrakciója, általánosítása. Az egyed által képviselt konkrét elemek halmazát *egyedhalmaznak* is nevezik. Például a hallgatók egyed konkrét hallgatóinak halmaza egy egyedhalmaz. Egy adott hallgató a hallgató egyed egy *előfordulása*, vagy értéke.

Tulajdonság

- A **tulajdonság**, más szóval **attribútum**, az egyed leírására szolgál, az egyed egy jellemzője. A tulajdonságok teszik lehetővé, hogy az egyed előfordulásait megkülönböztethessük egymástól. A hallgató egyed tulajdonságai lehetnek például: név, lakóhely, iskola, szak, Neptun kód, stb.
- **Többértékű tulajdonság**: azok a tulajdonságok, amelyeknek több értéke lehet egy egyed esetén. Például egy munkahelyi nyilvántartásban a dolgozó egyed végzettség tulajdonsága lehet többértékű, ha egy dolgozónak több végzettsége van.
- **Összetett tulajdonság**: ha a tulajdonság több attribútumból tevődik össze. Például a cím tulajdonság, amely irányítószámból, helység névből, utcából és házszámból épül fel.

Kulcs

- Azt a tulajdonságot vagy tulajdonságok csoportját, amely egyértelműen meghatározza, hogy az egyed mely értékéről, előfordulásáról van szó **kulcsnak** nevezzük. Ha a kulcs egy attribútumból áll, akkor *egyszerű kulcsról*, ha több attribútumból tevődik össze *összetett kulcsról* beszélünk. Azok a tulajdonságok, amelyek a kulcsot alkotják az *elsődleges attribútumok*, amelyek nem részei a kulcsnak a *másodlagos attribútumok*.
- Az egy egyednél előforduló, lehetséges kulcsokat *kulcsjelölteknek* hívjuk, ezek közül azt amelyiket kiválasztjuk az egyed értékeinek azonosítására az *elsődleges kulcs*, a többi kulcsot *másodlagos kulcsnak* vagy *alternatív kulcsnak* nevezzük.

Idegen kulcs

- **Idegen vagy külső kulcsnak** nevezzük egy egyedben azt az attribútumot, amely egy másik egyedben elsődleges kulcs. Az idegen kulcsok az egyedek közötti kapcsolatok kialakításánál játszanak szerepet.

Kapcsolat

- Az egyedek közötti viszonyt fejezi ki, úgy mint a valóságos objektumok esetében.
- A kapcsolatok fajtája határozza meg, hogy egy egyedhalmaz elemei milyen módon kapcsolódnak a másik egyedhalmazhoz, vagyis az adott egyed előfordulásaihoz hány előfordulás kapcsolódik a másik egyedből.

1:1 kapcsolat

- **Egy-egy (1:1) típusú kapcsolat:** egy egyed egy előfordulásához a másik egyed pontosan egy előfordulása kapcsolódik és ez fordítva is igaz. Tehát kölcsönösen egyértelmű megfeleltetés áll fenn az egyedek előfordulásai között.
- pl.: Férj-Feleség

1-N kapcsolat

- **Egy-több (1:N) típusú kapcsolat:** az egyik egyed mindegyik előfordulásához a másik egyed több előfordulása is tartozhat.
- pl.: Anya-Gyermek

N-M kapcsolat

- **Több-több (N:M) típusú kapcsolat**: az egyik egyed mindegyik előfordulásához a másik egyed több előfordulása is tartozhat és viszont.
- Az N:M típusú kapcsolatoknál megadható a maximális kardinalitási szám, amely azt határozza meg, hogy egy egyed előfordulásához a másik egyed maximum hány előfordulása kapcsolódhat. Ehhez hasonlóan definiálható a minimális kardinalitás.
- pl.: Hallgató-Tantárgy

Vázlat

- 1 Tematika
- 2 Fogalmak
 - Egyed-kapcsolat diagram
 - Relációs adatmodell
- 3 Használt szoftverek

Codd publikációja

- Edgar F. Codd, IBM
- 1970: A Rational Model of Data for Large Shared Data Banks
- Matematikai háttér: relációs algebra, relációs kalkulus

Codd publikációja

- Edgar F. Codd, IBM
- 1970: A Rational Model of Data for Large Shared Data Banks
- Matematikai háttér: relációs algebra, relációs kalkulus

Mondjunk példát relációra!

Kételemű (binér) relációk

Számok között:

- $=, \leq, <$

Egyenesek között:

Egyenes és sík között:

Kételemű (binér) relációk

Számok között:

- $=, \leq, <$
- osztó (a osztója b -nek)

Egyenesek között:

Egyenes és sík között:

Kételemű (binér) relációk

Számok között:

- $=, \leq, <$
- osztó (a osztója b -nek)
- akár ez is lehet: a négyzetgyöke b -nek

Egyenesek között:

Egyenes és sík között:

Kételemű (binér) relációk

Számok között:

- $=, \leq, <$
- osztó (a osztója b -nek)
- akár ez is lehet: a négyzetgyöke b -nek

Egyenesek között:

- párhuzamos, metsző, kitérő

Egyenes és sík között:

Kételemű (binér) relációk

Számok között:

- $=, \leq, <$
- osztó (a osztója b -nek)
- akár ez is lehet: a négyzetgyöke b -nek

Egyenesek között:

- párhuzamos, metsző, kitérő

Egyenes és sík között:

- az egyenest tartalmazza a sík

Kételemű (binér) relációk

Számok között:

- $=, \leq, <$
- osztó (a osztója b -nek)
- akár ez is lehet: a négyzetgyöke b -nek

Egyenesek között:

- párhuzamos, metsző, kitérő

Egyenes és sík között:

- az egyenest tartalmazza a sík
- az egyenes metszi a síkot

Kételemű (binér) relációk

Számok között:

- $=, \leq, <$
- osztó (a osztója b -nek)
- akár ez is lehet: a négyzetgyöke b -nek

Egyenesek között:

- párhuzamos, metsző, kitérő

Egyenes és sík között:

- az egyenest tartalmazza a sík
- az egyenes metszi a síkot
- az egyenes párhuzamos a síkkal

Descartes-szorzat

Definíció (Descartes-szorzat)

Legyenek H_1, \dots, H_n halmazok. Ekkor a

$$D = \{(h_1, \dots, h_n) \mid h_i \in H_i\}$$

(rendezett elem-n-esek) halmazát a H_i halmazok

Descartes-szorzatának nevezzük és a $H_1 \times \dots \times H_n$ kifejezéssel jelöljük.

Példa

Legyen $A = \{1, 2\}$ és legyen $B = \{a, b\}$. Ekkor a Descartes-szorzat

$$A \times B = \{(1, a), (1, b), (2, a), (2, b)\}.$$

A reláció foglalma

Definíció

Legyen $H_1 \times \dots \times H_n$ a $H_i (1 \leq i \leq n)$ halmazok Descartes-szorzata. Ekkor a

$$\rho \subseteq H_1 \times \dots \times H_n$$

halmazt a H_i halmazokon értelmezett relációnak nevezzük.

Mi köze az eddig ismert relációknak a mostani fogalomhoz?

Első példa a relációra

Példa

Legyen $A = B = \{1, 2, 3\}$. Reláció-e ekkor

$$\rho = \{(a_i, b_i) \mid a_i < b_i \text{ és } a_i \in A \text{ valamint } b_i \in B\}?$$

Igen, mert

$$\rho = \{(1, 2), (1, 3), (2, 3)\} \subseteq A \times B.$$

Ilyenkor jelölhetem úgy, hogy $(2, 3) \in \rho$. De kétváltozós relációnál gyakran a reláció jelét a két érték közé rakjuk: $2\rho 3$. Vagy ρ helyett más jelet használva: $2 < 3$.

Még egy példa

Példa

Legyen $A = \{\text{Lánczos Kornél, Eszterházy Péter, Neumann János, Csoóri Sándor}\}$, T a magyar települések halmaza, E az évszámok halmaza, F a foglalkozások halmaza. Válogassunk ki úgy elemnégyeseket, amelynél az első tag ember A -ból van, a második az ember születési helye, a harmadik a sz. éve, a negyedik a foglalkozása.

Ez reláció, mert része az $A \times T \times E \times F$ halmaznak.

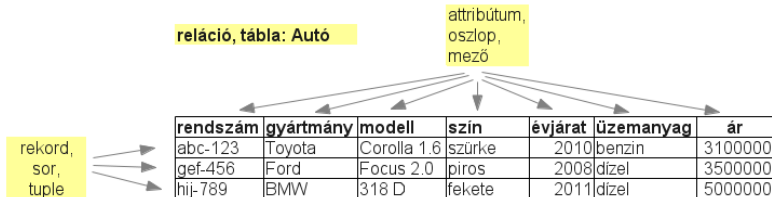
(Lánczos Kornél,	Székesfehérvár,	1893,	fizikus)
(Eszterházy Péter,	Csákvár,	1950,	író)
(Neumann János,	Budapest,	1903,	matematikus)
(Csoóri Sándor,	Zámoly,	1930,	költő)
(Csoóri Sándor,	Zámoly,	1930,	politikus)

És ez már hasonlít is egy szokványos adatbázis-táblához.

Relációs adatmodell

- A relációs adatbázisban a **reláció** egy táblázat. A táblázat szó helyett, általában a **tábla** elnevezés használatos.
- **Oszlop** vagy **mező**: az adat leírására szolgál, névvel és típussal rendelkezik.
- **Értéktartomány**, az oszlopok által felvehető értékek halmaza.
- Az oszlopok számát a **reláció fokának**, a sorok számát a **reláció számosságának** nevezzük.
- **Rekord**: egy egyed egy konkrét előfordulását adja meg, a rekordok sorrendje a táblában tetszőleges.
- Táblák közötti kapcsolatok: 1-1, 1-N, N-M.

Példa: Tábla



Vázlat

- 1 Tematika
- 2 Fogalmak
 - Egyed-kapcsolat diagram
 - Relációs adatmodell
- 3 Használt szoftverek

PostgreSQL

- A PostgreSQL egy nyílt forráskódú adatbázis-kezelő rendszer, amely számos operációs rendszeren (Windows, Linux, Mac OS X, AIX, BSD, HP-UX, SGI IRIX, Solaris, Tru64) használható.
- A PostgreSQL fejlesztése a kaliforniai Berkeley Egyetemen (University of California, Berkeley) kezdődött Postgres néven. 1989 júniusában jelent meg az adatbázis-kezelő 1.0 verziója, majd a 4. verzió megjelenését követően az egyetemi projekt befejeződött. 1996-tól a PostgreSQL Global Development Group közösség nyílt forráskódú projektként indította újra a fejlesztést.

PostGIS

- A térbeli adatok tárolására és kezelésére a PostGIS kiterjesztés telepíthető.
- A PostGIS a Refractions Research Inc. nyílt forráskódú projektje abból a célból, hogy a PostgreSQL adatbázis-kezelő részére biztosítsa a térbeli adatkezelési képességeket az OGC szabványoknak megfelelően.