Бази от данни

Функционални зависимости (първа част)

доц. д-р Димитър Димитров

Въведение

- БД може да бъде лошо моделирана
  - Излишества
  - Неконсистентност
- Например при проектиране, при което не се започва с E/R, а направо с релационен модел
- Релационните схеми могат да бъдат подобрени чрез ограничения, напр.
  - Функционални зависимости
  - Референтна цялостност
  - Многозначни зависимости

Да си припомним

- С всеки атрибут на релация е свързан **домейн**
  - Множество от допустими стойности
- Релационният модел изисква стойностите на атрибутите да бъдат атомарни
  - Studio (name: string, address: string)

Математическа дефиниция

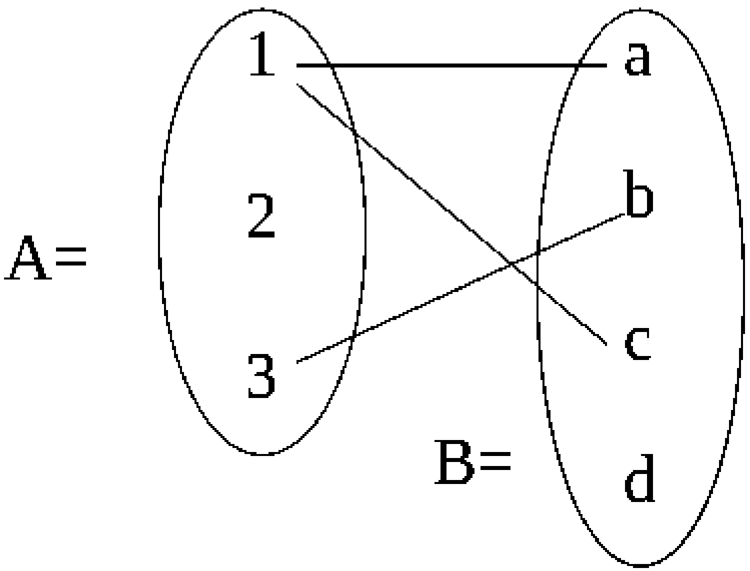
- n-арна **релация** върху множествата A_1, A_2, …, A_n е множеството от наредени n-торки $<a_1, \ldots, a_n>$, където $a_i \in A_i$ за $i=1, \ldots, n$
- Подмножество на декартовото произведение $A_1 \times A_2 \times \ldots \times A_n$ (което от своя страна е множеството от всички наредени n-торки от описания вид)

Примери

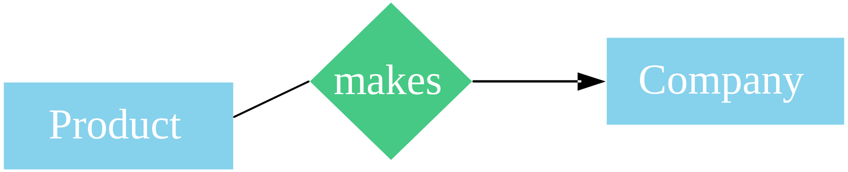
A={1,2,3}

B={a,b,c,d},

R = {(1,a), (1,c), (3,b)}



makes е подмножество на Product x Company



Формална дефиниция

- Нека r е релация
- Да означим с R нейната схема - $R(A_1, A_2, ..., A_n)$
- Тогава r (или $r(R)$) е математическа релация от степен n върху домейните $\mathrm{dom}(A_1)$, $\mathrm{dom}(A_2)$, …, $\mathrm{dom}(A_n)$
- Подмножество на декартововото произведение на домейните, които дефинират R. Това може да се изрази чрез:
- $r(R) \subseteq (\mathrm{dom}(A_1) \times \mathrm{dom}(A_2) \times ... \times \mathrm{dom}(A_n))$
- $r(R)$ е множество от n-tuples, $r = \{t_1, t_2, \ldots, t_m\}$
- Всеки n-tuple t е подреден списък от n стойности $t = \langle v_1, v_2, \ldots, v_n \rangle$, където v_i, $1 \leq i \leq n$, е елемент от $\mathrm{dom}(A_i)$ или специалната стойност NULL

Нотация

- $R (A_1, A_2, \ldots, A_n)$
- $r = \{t_1, t_2, \ldots, t_m\}$
- $t = <v_1, v_2, \ldots, v_n>$
- $t[A_1]$ или $t.A_1$
- Имена на релации: Q, R, S
- Екземпляри на релации: q, r, s
- Кортежи (tuples): t, u, v

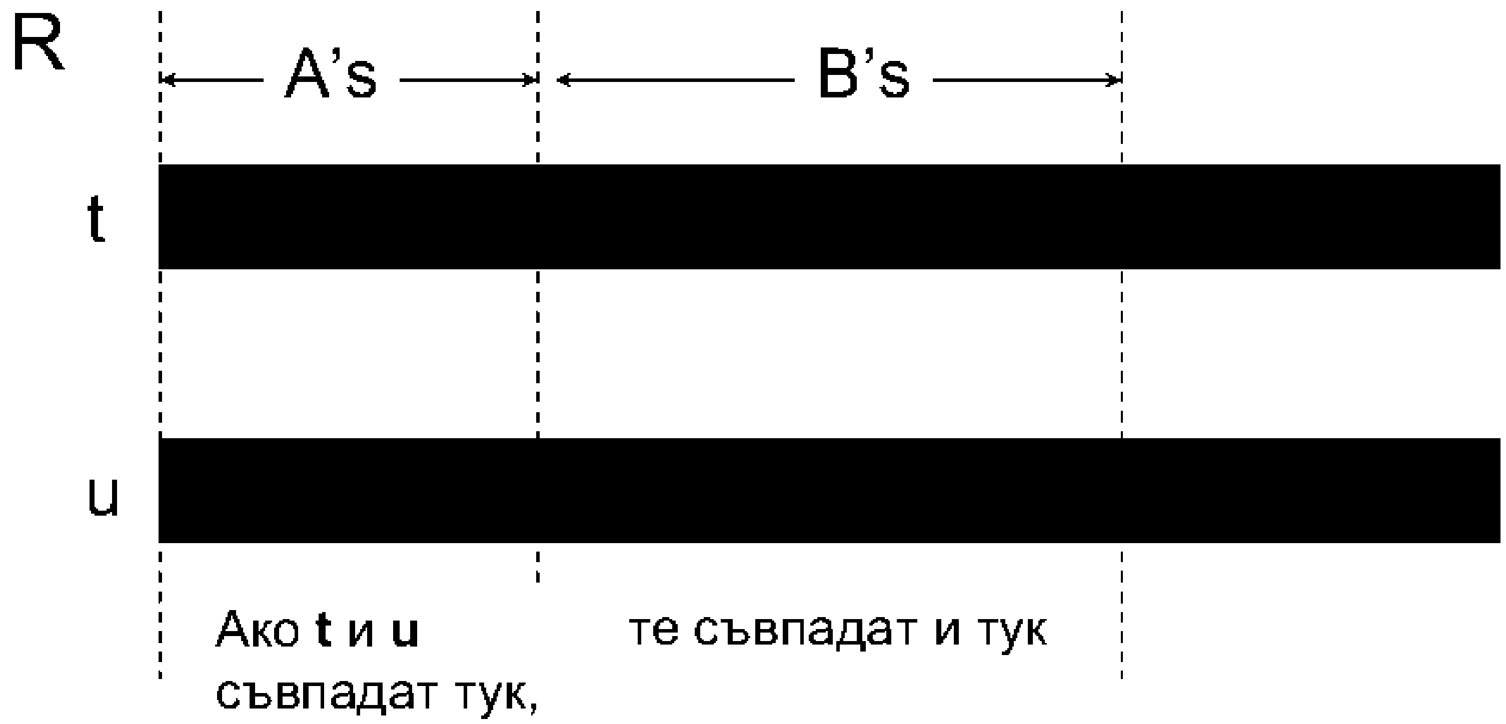
Функционални зависимости

- Най-важните ограничения в релационния модел
- Знанията за функционални зависимости са от особена важност при проектиране на БД за елиминиране на аномалии и излишества

Дефиниция

- Нека *R* е релация, а *A* и *B* са две множества от атрибути на *R*
- <u>*A* функционално определя *B*</u> тогава и само тогава, когато <u>за всеки два кортежа</u> от *R* е изпълнено, че <u>ако те съвпадат по атрибутите *A*</u>, то те <u>съвпадат и по *B*</u>
- Означение: $A \rightarrow B$

- Нека $A \rightarrow B$
- За всеки два кортежа t и u на R:



Пример 1

Employee ID	Employee name	Department ID	Department name
0001	John Doe	1	Human Resources
0002	Jane Doe	2	Marketing
0003	John Smith	1	Human Resources
0004	Jane Goodall	3	Sales

- Employee ID → Employee Name
- Employee ID → Department ID
  - или: Employee ID → {Employee Name, Department ID}
- Department ID → Department Name

Источник: Wikipedia

Пример 2

Movies					
title	***year***	***length***	***filmType***	***studioName***	***starName***
Star Wars	1977	124	color	Fox	Carrie Fisher
Star Wars	1977	124	color	Fox	Mark Hamill
Star Wars	1977	124	color	Fox	Harrison Ford
Mighty Ducks	1991	104	color	Disney	Emilio Estevez
Wayne's World	1992	95	color	Paramount	Dana Carvey
Wayne's World	1992	95	color	Paramount	Mike Myers

- title, year → length
  - Защо и year?
  - Защото гледаме схемата, а не конкретния екземпляр
- title, year → filmType
- title, year → studioName
- Но не и title, year → starName

Пример 3

- Дадена е релация със следните атрибути:
  - code – код на курс
  - name – име на курс
  - period – период, в който даден курс е преподаван (обикновено един курс се преподава много пъти)
  - #students – брой студенти, посещаващи курс
  - teacher – отговорник на курса (винаги ли е един и същ?)
  - room – име на стая
  - #seats – брой места в стая
  - weekday – ден от седмицата (в даден период лекциите по БД са били всеки вторник и четвъртък)
  - hour – час
- Да намерим всички ФЗ

Пример 3 – решение

- code → name
- code, period → #students
- code, period → teacher
- room → #seats
- code, period, weekday → hour
- code, period, weekday → room
- room, period, weekday, hour → code

Защо функционални?

- $A \to B$ означава, че стойностите на B са функция на стойностите на A
- Какво е функция в математиката?
- Съществува функция, която по дадени стойности на всеки от атрибутите $A_1, \ldots, A_n$ съпоставя <u>единствена</u> стойност на $B_1, \ldots, B_m$

Нотация

- *A*, *B*, *C*, … – множества от атрибути
- A_1, A_2, … – индивидуални атрибути
- *F* – множеството на функционалните зависимости
- *f* – индивидуална ФЗ

ФЗ и схема

- Функционалната зависимост е твърдение за <u>схемата</u> на релацията, а не за конкретен екземпляр
- ФЗ не могат да се определят чрез просто преглеждане на данните
  - Да си припомним пример №2
- ФЗ са свойства на семантиката на атрибутите
- Всички данни във всеки възможен екземпляр удовлетворяват съществуващите ФЗ
  - Важи за всяко друго ограничение

Ключове ⚷

- Множеството от атрибути $K = \{A_1, A_2, \ldots, A_n\}$ на релацията R наричаме ключ за релацията, ако:

    1) K функционално определя всички атрибути на R

    2) За никое подмножество на K не е изпълнено условие 1)

Бележки по дефиницията

- В една релация може да има няколко различни (кандидат) ключа
- Тъй като релациите са множества, не е възможно два различни кортежа от *R* да съвпадат по всичките си атрибути
  - Ако това е така, то значи става въпрос за един и същ кортеж
- Не съществуват два кортежа от *R*, които да съвпадат по атрибутите на *K*
  - Разбира се, може да съвпадат по част от атрибутите, напр. филми от една и съща година
- Може да кажем, че множеството *K* е минимално
  - Ако премахнем който и да е негов атрибут, вече не е ключ

Ключ – пример

- StarMovies (<u>title</u>, <u>year</u>, length, filmtype, studioName, <u>starName</u>)
- Ключ: {title, year, starName}
- Кои са подмножествата на ключа и изпълняват ли условие 1) ?

Movies

title	*year*	*length*	*filmType*	*studioName*	*starName*
Star Wars	1977	124	color	Fox	Carrie Fisher
Star Wars	1977	124	color	Fox	Mark Hamill
Star Wars	1977	124	color	Fox	Harrison Ford
Mighty Ducks	1991	104	color	Disney	Emilio Estevez
Wayne's World	1992	95	color	Paramount	Dana Carvey
Wayne's World	1992	95	color	Paramount	Mike Myers

Суперключ

- Да си припомним:
  - Множеството от атрибути $K = \{A_1, A_2, \ldots, A_n\}$ на релацията R наричаме ключ за релацията, ако:
    1) K функционално определя всички атрибути на R
    2) За никое подмножество на K не е изпълнено условие 1)
- Ако K удовлетворява 1), но не и 2), K е суперключ
  - Супермножество (надмножество) за ключа
  - За ключовете в E/R модела няма изискване за минималност

Суперключ – пример

- StarMovies (<u>title</u>, <u>year</u>, length, filmtype, studioName, <u>starName</u>)

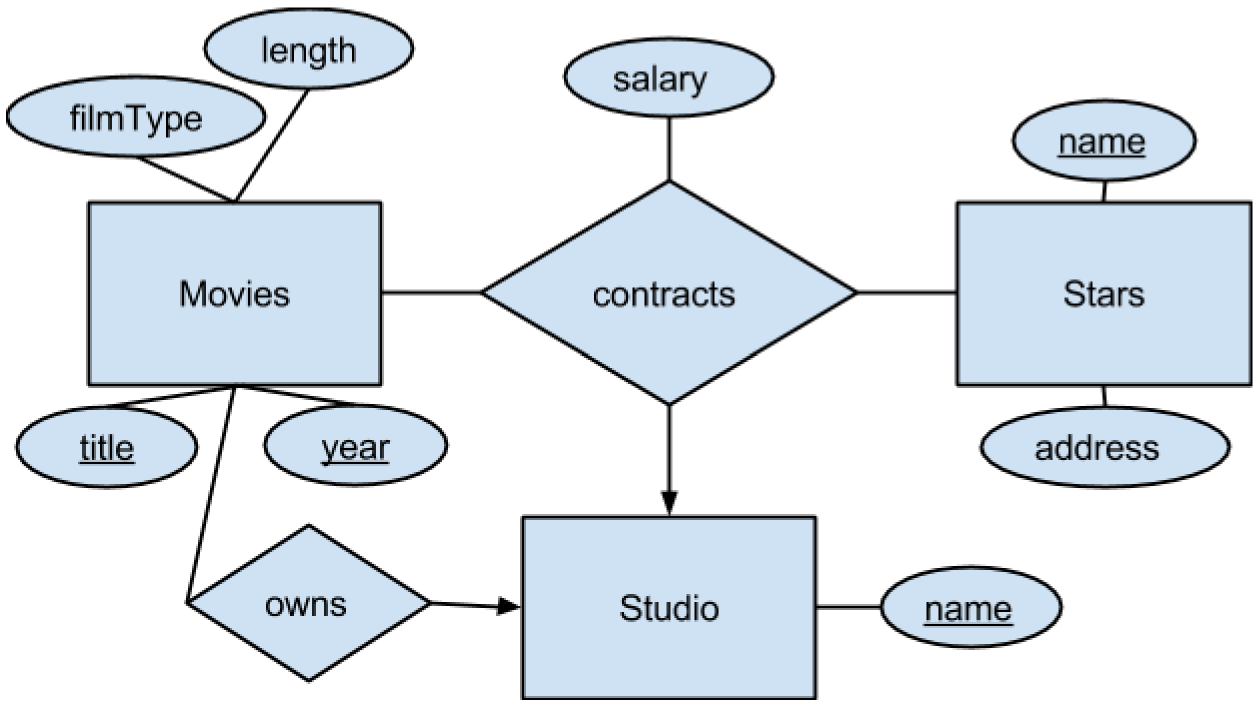
- Някои суперключове:
  - {title, year, starName, length}
  - {title, year, starName, studioName}

Откриване на ключове

- Когато релационната схема е получена от преобразуване на E/R модел, ключът може да се предвиди
- Нека си припомним
  - Множество от същности
  - Връзка “много към много”
  - Връзка “много към едно”, преобразувана в отделна релация
  - Връзка “едно към едно”, преобразувана в отделна релация – имахме избор

Задача

- Да преобразуваме в релационен модел и да определим ключовете



Решение

- Movies (title, year, length, filmType)
- Studios (name)
- Stars (name, address)
- Contracts (title, year, starName, studioName, salary)
  - Небинарна връзка, Studios е откъм “едно”-страната
- Owns (title, year, studioName)

Първичен ключ

- Някои релации имат повече от един възможен ключ
- Избираме един от тях за първичен (primary)
- Изборът е важен
- Понякога* се въвежда сурогатен ключ
  - При липса на естествен ключ сред атрибутите на релацията
  - При наличие на няколко равностойни ключа

Добър и лош дизайн

- Да помислим (засега само интуитивно) кой дизайн е по-добър:
  - Data (Id#, Name, Address, C#, Description, Grade)

Id	Name	Address	C	Description	Grade
124	Jones	Phila	Phil7	Plato	A
456	Smith	NYC	Phil7	Plato	B
789	Brown	Boston	Math8	Topology	C
124	Jones	Phila	Math8	Topology	A
789	Brown	Boston	Eng12	Chaucer	B

  или

  - Student (Id#, Name, Address)
  - Course (C#, Description)
  - Enrolled (Id#, C#, Grade)

Проблеми
на първия вариант

- Излишества при първия вариант
  - Име и адрес
- Информацията за курс зависи от наличието на студент
- ФЗ при първия вариант:
  - Id# → Name, Address
  - C# → Description
  - Id#, C# → Grade

Определяне на всички ФЗ

- Как да открием всички ФЗ, ако знаем само някои от тях?
- Да си припомним: не можем да ги открием с претърсване на данните в конкретна инстанция на релация
  - Може погрешно да определим несъществуващи ФЗ

Аксиоми на Армстронг (1)

- Някои ФЗ могат да бъдат получени като логически следствия от други ФЗ
- ...чрез прилагане на правила, познати като аксиоми на Армстронг
- Основни правила:
  - Рефлексивност
  - Разширение
  - Транзитивност

Аксиоми на Армстронг (2)

Нека W, X, Y и Z са множества от атрибути на дадена релация

A1. Рефлексивност (Reflexivity)

– Ако $Y \subseteq X$, то $X \rightarrow Y$

A2. Разширение, попълнение (Augmentation)

– Ако $X \rightarrow Y$, то $XW \rightarrow YW$

A3. Транзитивност (Transitivity)

– Ако $X \rightarrow Y$ и $Y \rightarrow Z$, то $X \rightarrow Z$

Примери

Аксиоми на Армстронг (3)

- Да се опитаме да докажем аксиомите

- Пищов:

$A \to B \Leftrightarrow$ за всеки два кортежа от R е изпълнено, че ако те съвпадат по атрибутите A, то те съвпадат и по B

A1. Ако $Y \subseteq X$, то $X \to Y$

A2. Ако $X \to Y$, то $XW \to YW$

A3. Ако $X \to Y$ и $Y \to Z$, то $X \to Z$

Д-во на Аксиома 1: рефлексивност

- Ако $Y \subseteq X$, то $X \to Y$
- Доказателство: всеки два кортежа *t* и *u* съвпадат по всички атрибути на *X*, следователно те съвпадат и по всяко подмножество на *X*, включително *Y*

Доказателство на Аксиома 2: разширение

- Ако $X \rightarrow Y$, то $XW \rightarrow YW$
- Доказателство:
  - Да допуснем противното – че съществуват два кортежа *t* и *u*, които съвпадат по всички атрибути на *XW*, но не съвпадат по *YW*
  - *t* и *u* задължително съвпадат по W
  - Следователно *t* <> *u* по някой от атрибутите на *Y*, което противоречи на $X \rightarrow Y$

Доказателство на Аксиома 3: транзитивност

- Ако $X \rightarrow Y$ и $Y \rightarrow Z$, то $X \rightarrow Z$
- Доказателство:
  - Да допуснем, че съществуват два кортежа (x, y_1, z_1) и (x, y_2, z_2), които съвпадат по всички атрибути на X
  - $X \rightarrow Y$, следователно щом съвпадат по всички атрибути на X, задължително съвпадат и по всички атрибути на Y, т.е. $y_1 = y_2$
  - $Y \rightarrow Z$, следователно $z_1 = z_2$
  - Двата кортежа съвпадат

Следствия от аксиомите на Армстронг

- Обединение
  - Ако $X \rightarrow Y$ и $X \rightarrow Z$, то $X \rightarrow YZ$
- Псевдотранзитивност
  - Ако $X \rightarrow Y$ и $WY \rightarrow Z$, то $XW \rightarrow Z$
- Декомпозиция
  - Ако $X \rightarrow Y$ и $Z \subseteq Y$, то $X \rightarrow Z$
- Примери
- Доказателство – чрез аксиомите на Армстронг

Обединение

- Ако $X \to Y$ и $X \to Z$, то $X \to YZ$
- Доказателство:
  - $X \to Y$, следователно $X \to XY$ (A2)
  - $X \to Z$, следователно $XY \to ZY$ (A2)
  - $X \to XY$, $XY \to ZY$, следователно $X \to ZY$ (A3)

Псевдотранзитивност

- Ако $X \rightarrow Y$ и $WY \rightarrow Z$, то $XW \rightarrow Z$
- Доказателство:
  - $X \rightarrow Y$, следователно $WX \rightarrow WY$ (A2)
  - $WX \rightarrow WY$ и $WY \rightarrow Z$, следователно $WX \rightarrow Z$ (A3)

Декомпозиция

- Ако $X \rightarrow Y$ и $Z \subseteq Y$, то $X \rightarrow Z$
- Доказателство:
  - $X \rightarrow Y$ и $Z \subseteq Y$, следователно $Y \rightarrow Z$ (A1)
  - $X \rightarrow Z$ (A3)

Правило за разделяне и обединение

- Ако в дясната част на ФЗ имаме множество от атрибути, можем да поставим всеки от тях в нова ФЗ (и обратното)
- Ако $A \rightarrow B_1, B_2, \ldots, B_n$,

  в сила са и следните ФЗ:

  $A \rightarrow B_1$
  $A \rightarrow B_2$
  $\ldots$
  $A \rightarrow B_n$
- И обратното – ако имаме $A \rightarrow B_1, \ldots, A \rightarrow B_n$, то в сила е и $A \rightarrow B_1, \ldots, B_n$
- Можем ли декомпозираме лявата част?

Пример

- Employee ID → {Employee Name, Department ID}

- Employee ID → Employee Name
- Employee ID → Department ID

Тривиални зависимости

- ФЗ $A_1A_2\ldots A_n \to B$ се нарича *тривиална*, ако атрибутът B съвпада с някой от атрибутите $A_1A_2\ldots A_n$
- Ако вдясно имаме множество от атрибути $B_1B_2\ldots B_m$:
- *Тривиална*: атрибутите $B_1B_2\ldots B_m$ са подмножество на $A_1A_2\ldots A_n$
  - title, year → title
- *Нетривиална*: поне един атрибут от $B_1B_2\ldots B_m$ не е подмножество на $A_1A_2\ldots A_n$
  - title, year → year, length
- *Напълно нетривиална*: нито един от атрибутите $B_1B_2\ldots B_m$ не е част от $A_1A_2\ldots A_n$
  - title, year → length
- Във всяка релация има (тривиални) ФЗ

Правило на тривиалната зависимост

- Имаме право от дясната част на ФЗ да премахнем тези атрибути, които принадлежат на лявата част:
  - title, year → ~~year~~, length

    →

  - title, year → length

Задачи

Дадени са следните ФЗ в R(A,B,C,D,E,F):
$A \rightarrow B$
$A \rightarrow C$
$CD \rightarrow E$
$CD \rightarrow F$
$B \rightarrow E$

1. Докажете, че $AD \rightarrow EF$
2. ACD суперключ ли е на R? Ако да, кои атрибути можем да премахнем?
3. Вярно ли е, че $BC \rightarrow F$?

Решения на задачите

1. Чрез псевдотранзитивност доказваме, че AD→E и т.н.
2. Да, C
3. Не

Въпроси?