

729G28: Webprogrammering och databaser

Fö 4: Normalisering

Eva L. Ragnemalm

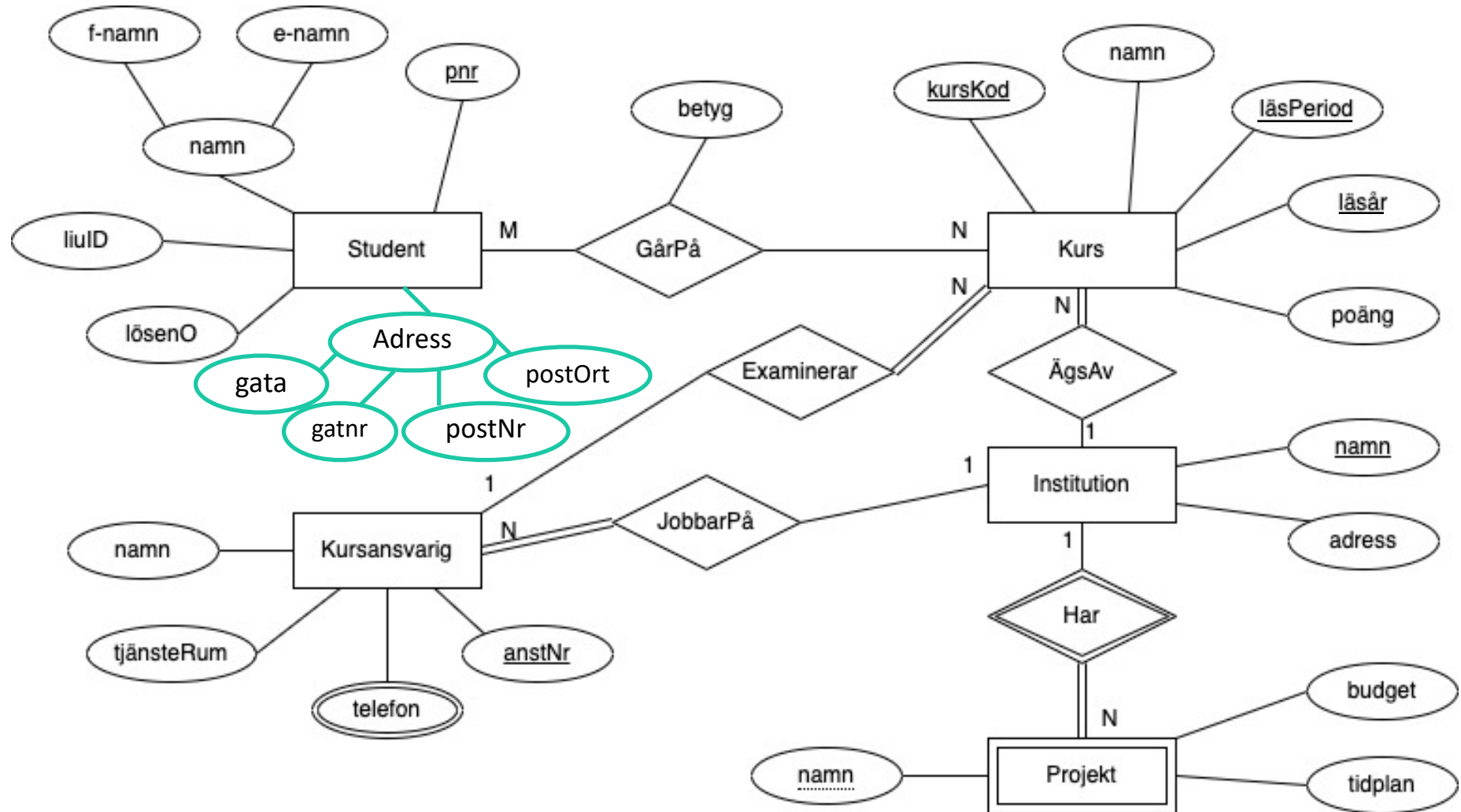
Relationsteori - mer definitioner

Normalformer (1NF, 2NF, 3NF och BCNF)

Varför normalisera?

- Metod att skydda oss från ”dum” design
 - Lagra samma data flera ggr i onödan
 - Inte kunna lagra viss data
 - Otydlig betydelse av en tupel/rad – som kan komma från otydlig betydelse av entitetsinstans

Universitetet: ER-diagram



Universitetet, med adress:

4

Student

<u>pnr</u>	e-namn	f-namn	liuID	lösen	gata	gatnr	postnr	postort
------------	--------	--------	-------	-------	------	-------	--------	---------

Kurs

<u>kurskod</u>	namn	<u>läsår</u>	period	poäng	kursansv	institution
----------------	------	--------------	--------	-------	----------	-------------

Kursansvarig

namn	<u>anstnr</u>	tjrum	inst
------	---------------	-------	------

Telefon

<u>anstnr</u>	<u>telnr</u>
---------------	--------------

Institution

<u>namn</u>	adress
-------------	--------

Projekt

<u>inst</u>	<u>namn</u>	tidsplan	budget
-------------	-------------	----------	--------

Registrerad-På

<u>student</u>	<u>kurskod</u>	<u>läsår</u>	betyg
----------------	----------------	--------------	-------

Integritetsvillkor från deltagande::

Kurs: kursansv, inst Not Null

Anställd: inst Not Null

Problem med Kurs

Kurs

<u>kurskod</u>	namn	<u>läsÅr</u>	period	poäng	kursansv	institution
729G28	Webbpgm o Datab	2015	ht1	6	Erik Prytz	IDA
729G28	Webbpgm o Datab	2022	ht2	6	Eva Ragnemalm	IDA
732G16	Databaser, design	2020	vt2	7,5	Eva Ragnemalm	IDA
TSBK07	Datorgrafik	2022	ht1	6	Ingemar Ragnemalm	ISY
TDDE43	Design av int. sys.	2024	ht1	12	Eva Blomkvist	IDA
TGTU90	Vetenskap filosofi	?	?	4	Maria Eidenstam	TEMA

Normalisering

En metod att finna redundans i **en** relation

Normaliserings-villkor = Normalformer

Begrepp:

- atomära attribut, supernycklar, kandidatnycklar och primärnycklar
- nyckelattribut
- funktionella beroenden och determinant
- informationsbevarande relationsschemauppdelning

1:a Normalform (1NF)

Alla attribut ska vara **atomära** (odelbara)

- Relationsmodellen antar att alla attribut är odelbara.
- 2:a normalform, 3:e normalform, BCNF baseras på **Funktionella beroenden** i relationen.

Repetition: nycklar i relationsmodellen

- Supernyckel
 - en attributmängd som bestämmer en tupel
- Kandidatnyckel
 - en minimal supernyckel
- Primärnyckel
 - den kandidatnyckel som väljs ut.

Nyckelattribut

- Attribut som ingår i **någon kandidatnyckel** till relationen.
- Student (pnr, e-namn, f-namn, liuid, lösen, gata, gatNr, postNr, postOrt)
 - pnr kan identifiera en rad
 - men liuid kan det också, är också **kandidatnyckel!**
 - pnr och liuid är **nyckelattribut** i Student

Nyckelattribut forts.

- Nyckelattribut markeras i relationsschemat med *
 - Student (pnr*, e-namn, f-namn, liuid*,... postort)
 - Kurs (namn, kurskod*, läsår*, period, poäng, kursansvarig, institution)

Student

<u>pnr</u> *	e-namn	f-namn	liuid*	lösen	gata	gatnr	postnr	postort
--------------	--------	--------	--------	-------	------	-------	--------	---------

Kurs

namn	<u>kurskod</u> *	<u>läsår</u> *	period	poäng	kursansv	institution
------	------------------	----------------	--------	-------	----------	-------------

Funktionellt beroende

X är en delmängd av attributen i en relation.

- Om X entydigt bestämmer värdet på ett attribut Y (i samma relation), kallas det att "Y är **funktionellt beroende** av X" eller att "X bestämmer Y".
- $X \Rightarrow Y$
- X kallas **determinant** i det funktionella beroendet.

Exempel

Ex: Student

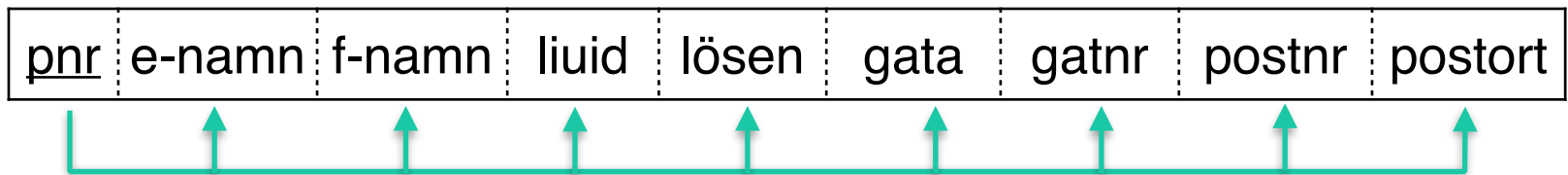
$\{pnr\} \Rightarrow e\text{-namn}, \{pnr\} \Rightarrow f\text{-namn}, \{pnr\} \Rightarrow lösen, \{pnr\} \Rightarrow liuid,$
 $\{pnr\} \Rightarrow gata, \{pnr\} \Rightarrow gatNr, \{pnr\} \Rightarrow postNr, \{pnr\} \Rightarrow postOrt,$
 $\{pnr\} \Rightarrow \{pnr\}$

Kan också skrivas:

$\{pnr\} \Rightarrow \{pnr, e\text{-namn}, f\text{-namn}, liuid, lösen, gata, gatNr,$
 $postNr, postOrt\}$

Och ritas:

Student



Exempel funktionellt beroende

- Determinanten kan vara flera attribut

ex: Kurs

$\{\text{kursKod}, \text{läsår}\} \Rightarrow \{\text{namn}, \text{period}, \text{poäng}, \text{kursansv}, \text{institution}\}$

- Determinanten får innehålla "onödiga" attribut

ex: Student:

$\{\text{pnr}, \text{e-namn}\} \Rightarrow \{\text{f-namn}, \text{liuid}, \text{lösen}, \text{gata}, \text{gatNr}, \text{postNr}, \text{postOrt}\}$

- Determinanten behöver inte vara en nyckel

ex Student:

$\{\text{postNr}\} \Rightarrow \{\text{postOrt}\}$

Funktionella beroenden, forts

- Funktionella beroenden baseras på relationens semantik. Ett fb ska gälla i alla möjliga instanser av databasen!
- OBS:
 - Att det finns ett fb $X \Rightarrow Y$ betyder **inte** att det finns ett fb $Y \Rightarrow X$
 - ex: $\text{pnr} \Rightarrow \text{e-namn}$ betyder inte att $\text{e-namn} \Rightarrow \text{pnr}$

Funktionella beroenden kan härledas

- Transitivitet
 - om $X \Rightarrow Y$ och $Y \Rightarrow Z$ så $X \Rightarrow Z$
- Reflexivitet:
 - $X \Rightarrow X$ alltid.
 - Obs: Om $Y \subseteq X$ så $X \Rightarrow Y$
- Augmentation:
 - Om $X \Rightarrow Y$ så har vi också $\{XZ\} \Rightarrow YZ$

Fullt funktionellt beroende

Givet: $X \Rightarrow Y$.

- Om inget attribut kan tas bort ur X utan att vi förlorar det funktionella beroendet (dvs X är minimal), kallas det **fullt funktionellt beroende**, ffb.

Fullt funktionellt beroende, exemplet Kurs

Kurs

<u>kurskod*</u>	namn	<u>läsår*</u>	period	poäng	kursansv	institution
-----------------	------	---------------	--------	-------	----------	-------------

- Nyckeln:
 $\{\text{kurskod, läsår}\} \Rightarrow \{\text{namn, period, poäng, kursansv, institution}\}$
- MEN: $\{\text{kurskod, läsår}\} \Rightarrow \text{poäng}$ är FB men ej FFB
– för ett program får inte ändra poäng hur som helst!
- OBS att verkligheten styr!

Andra normalformen - 2NF

- Definition: Ett relationsschema R är i 2NF om det är i 1NF och varje icke-nyckelattribut A i R är **FFB** av varje **kandidatnyckel** i R .
- Eller som fråga: Finns det något icke-nyckelattribut som bestäms av en del av en kandidatnyckel?

2NF Exemplet kurs

Kurs (kurskod*, namn, läsår*, period, poäng, kursansv, institution)

$\{\text{Kurskod}^*, \text{läsår}^*\} \Rightarrow \text{kursansv}$ FFB

$\{\text{Kurskod}^*, \text{läsår}^*\} \Rightarrow \text{period}$ FFB

MEN:

$\{\text{Kurskod}^*, \text{läsår}^*\} \Rightarrow \text{namn}$ inte FFB

$\{\text{Kurskod}^*, \text{läsår}^*\} \Rightarrow \text{poäng}$ inte FFB

$\{\text{Kurskod}^*, \text{läsår}^*\} \Rightarrow \text{institution}$ inte FFB

- Uppfyller ej 2NF

Informationsbevarande relationsschemauppdelning

Om vi delar upp en relation R i relationerna R_1 och R_2 så blir uppdelningen **informationsbevarande**

- **om** R_2 innehåller en determinant X plus de attribut som bestäms av den
- **och** R_1 innehåller ”resten av attributen i R ” plus determinanten i det funktionella beroendet.
- (och inget av attributen som nu bara finns i R_2 ingår i en determinant som bestämmer attribut i R_1).

Informationsbevarande relationsschemauppdelning, kontroll

- SQL-operationen **R join S on S.a=R.a**
- sätter ihop två relationer (R och S) till en, där varje rad ur den högra relationen sätts ihop med varje rad ur den vänstra, för de rader där villkoret ($S.a=R.a$) är uppfyllt.
(a är attribut i R och S som har samma domän)
- Om $R_1 \text{ join } R_2 \text{ on } R_1.x=R_2.x = R$
där x är determinanten i det funktionella beroendet
- så är uppdelningen informationsbevarande!

Informationsbevarande relationsschemauppdelning, inte

Person(pnr, namn, adress)
med funktionella beroenden:
 $\text{pnr} \Rightarrow \text{pamn}$, $\text{pnr} \Rightarrow \text{adress}$,
inte $\text{namn} \Rightarrow \text{adress}$

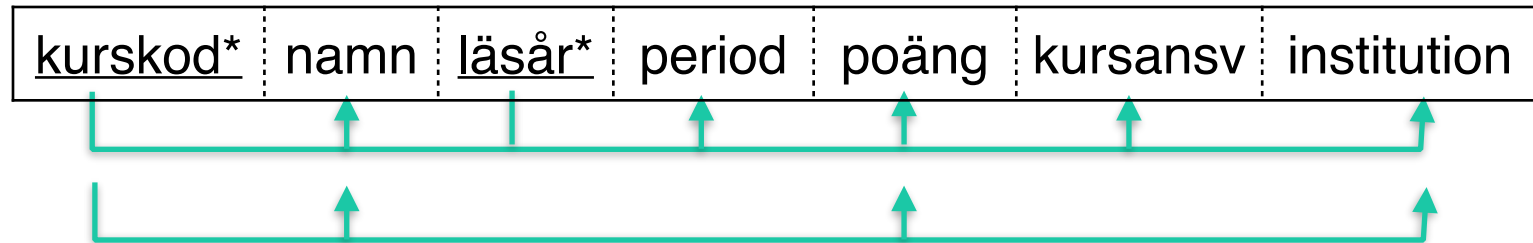
pnr	namn	adress
7908101234	Anna	Ågatan 3
8112237890	Anna	Rydsv.12

pnr	namn
7908101234	Anna
8112237890	Anna

namn	adress
Anna	Ågatan 3
Anna	Rydsv.12

Exemplet Kurs

Kurs



Ger:

Kurs1(Kurskod, namn, poäng, inst)

Kurs2(Kurskod, läsår, period, kursansv)

Tredje normalform - 3NF

- Definition: Ett relationsschema R är i 3NF
om det är i 2NF
och inget icke-nyckelattribut är ffb av något annat icke-nyckelattribut
- Frågan för 3NF:
 - Finns ffb mellan icke-nyckel-attribut?

Exemplet Student

Student (pnr*, e-namn, f-namn, liuid*, lösen, gata, postnr, postort)

Funktionella beroenden:

$\{pnr^*\} \Rightarrow \{namn, lösen, liuid, gata, postnr, postort\}$

$\{liuid^*\} \Rightarrow \{pnr, namn, lösen, gata, postnr, postort\}$

$\{postnr\} \Rightarrow postort$

- Uppfyller INTE 3NF

Boyce-Codd normalform - BCNF

- Ett relationsschema R är i BCNF
om det är i 3NF
och alla determinanter är kandidatnycklar.
- Frågan för BCNF: är varje determinant en kandidatnyckel?

3NF/BCNF

- Inte alltid möjligt att transformera ett schema till BCNF och behålla alla beroendena.
- 3NF har de flesta av BCNF's fördelar.
- Det är inte självklart att man måste uppfylla BCNF.
- OBS: 3NF tillåter någon sorts redundans som BCNF inte gör (funktionella beroenden mellan nyckelattribut).

Frågor?