

12– Identificazione delle chiavi di uno schema

Prof.ssa Maria De Marsico
demarsico@di.uniroma1.it



SAPIENZA
UNIVERSITÀ DI ROMA



- Utilizziamo il calcolo della chiusura di un insieme di attributi per determinare le chiavi di uno schema R su cui è definito un insieme di dipendenze funzionali F

- Dato il seguente schema di relazione

$$R = (A, B, C, D, E, H)$$

e il seguente insieme di dipendenze funzionali

$$F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$$

Calcolare la chiusura dell'insieme ABH

Applicando l'algoritmo per il calcolo della chiusura di un insieme di attributi, **inizializziamo $Z=ABH$** e al **primo passo prima del while** inseriamo nella chiusura di ABH gli attributi C, D ed E grazie alle dipendenze funzionali **in cui la parte sinistra è contenuta in ABH**, cioè $AB \rightarrow CD$ e $AB \rightarrow E$

In realtà a questo punto potremmo anche fermarci perché abbiamo inserito **tutti** gli attributi dello schema, cioè abbiamo verificato $(ABH)^+ = \{A, B, C, D, E, H\}$.

begin

$R = (A, B, C, D, E, H) \quad F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$

$Z := ABH;$

$S := \{ A/Y \rightarrow V \in F \wedge A \in V \wedge Y \subseteq ABH \} = \{ C \text{ (per la dipendenza } AB \rightarrow \underline{CD}), D \text{ (per la dipendenza } AB \rightarrow \underline{CD}), E \text{ (per la dipendenza } AB \rightarrow \underline{E}) \} = CDE$

controllo while ($S \not\subseteq Z$?): $CDE \not\subseteq ABH$ quindi entriamo nella prima iterazione del while

begin (prima iterazione del while)

$Z := Z \cup S = ABH \cup CDE = ABCDEH$

$S := \{ A/Y \rightarrow V \in F \wedge A \in V \wedge Y \subseteq ABCDEH \} = \{ C \text{ (per la dipendenza } AB \rightarrow \underline{CD}), D \text{ (per la dipendenza } AB \rightarrow \underline{CD} \text{ o per } ABC \rightarrow \underline{D}), E \text{ (per la dipendenza } AB \rightarrow \underline{E} \text{ o per } C \rightarrow \underline{E}) \} = CDE$

end

controllo while: $CDE \subset ABCDEH$ (non abbiamo aggiunto niente di nuovo)

usciamo dal while

end

$ABH += ABCDEH = R$

Chiave?



- $ABH^+ = ABCDEH = R$
- Potrebbe essere una chiave di R?

Ricordiamo le condizioni

Definizione

Dati uno schema di relazione R e un insieme F di dipendenze funzionali

un sottoinsieme K di uno schema di relazione R è **una chiave** di R se:

1. $K \rightarrow R \in F^+$
2. non esiste un sottoinsieme proprio K' di K tale che $K' \rightarrow R \in F^+$

ATTENZIONE: dobbiamo sempre verificare che nessun sottoinsieme di K determini funzionalmente R

$$K' \rightarrow R \in F^+ \Leftrightarrow R \subseteq K'^+_F$$

- Usiamo ancora l'algoritmo sui sottoinsiemi di K !

- $R = (A, B, C, D, E, H)$ $F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$
calcoliamo la chiusura **dei sottoinsiemi di ABH**
- **Osservazione 1:** conviene partire da quelli **con cardinalità maggiore ...**
se la loro chiusura non contiene R, è inutile calcolare la chiusura dei
loro rispettivi sottoinsiemi
- **Osservazione 2:** gli attributi che **non compaiono mai a destra delle**
dipendenze funzionali di F, non sono determinati funzionalmente da
nessun altro attributo ... quindi rimarrebbero **fuori** della chiusura di
qualunque sottoinsieme di R che non li contenesse ... ma ogni **chiave**
deve determinare **tutto** lo schema ..., quindi gli attributi che non
compaiono a destra di nessuna dipendenze funzionale in F **dovranno**
essere sicuramente in ogni chiave
- **Osservazione 3:** le osservazioni 1 e 2 valgono **anche** quando stiamo
cercando la chiave o le chiavi di uno schema
- **Osservazione 4:** l'approccio di **forza bruta** (provare comunque **TUTTI**
i sottoinsiemi) **non è sbagliato ma molto poco efficiente** (facciamo
molti calcoli che già sappiamo essere inutili) **tuttavia ...**
- **Osservazione 5:** ... negli esercizi ogni scorciatoia nei calcoli va
giustificata

Chiave?



- $R = (A, B, C, D, E, H)$ $F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$
- calcoliamo la chiusura **dei sottoinsiemi di ABH**
 - H non è determinato da altri attributi, deve essere parte di **qualunque** chiave \rightarrow i sottoinsiemi di cardinalità 2 da verificare sono **AH e BH**

begin

$R = (A, B, C, D, E, H) \quad F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$

$Z := AH;$

$S := \{ A/Y \rightarrow V \in F \wedge A \in V \wedge Y \subseteq AH \} = \emptyset$ (né AH, né A né H determinano funzionalmente alcun altro attributo)

controllo while ($S \not\subseteq Z$?): $\emptyset \subset AH$ quindi **non** entriamo nella prima iterazione del while

end

$AH^+ = AH \subset R$

• -----

begin

$Z := BH;$

$S := \{ A/Y \rightarrow V \in F \wedge A \in V \wedge Y \subseteq BH \} = \emptyset$ (né BH, né B né H determinano funzionalmente alcun altro attributo)

controllo while ($S \not\subseteq Z$?): $\emptyset \subset BH$ quindi **non** entriamo nella prima iterazione del while

end

$BH^+ = BH \subset R$

- Inutile provare con sottoinsiemi di cardinalità ancora minore, quindi abbiamo verificato che:

1. $ABH \rightarrow R \in F^+$
2. non esiste un sottoinsieme proprio K' di ABH tale che $K' \rightarrow R \in F^+$

- Quindi ABH è una chiave di R



- Dato uno schema R ed un insieme di dipendenze funzionali su R , **possiamo avere più chiavi per R**

Quali chiavi?



- $R = (A, B, C, D, E, H)$ $F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$
- Abbiamo detto che H deve essere incluso in ogni chiave
- Abbiamo già verificato che H, AH e BH non sono chiavi
- Proviamo con CH, DH ed EH

Quali chiavi?



$R = (A, B, C, D, E, H)$

$F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$

begin

$Z := CH;$

$S := \{ A/Y \rightarrow V \in F \wedge A \in V \wedge Y \subseteq CH \} = \{ E \}$ (per la dipendenza $C \rightarrow E$)

controllo while ($S \not\subseteq Z$?): $E \not\subseteq CH$ quindi entriamo nella prima iterazione del while

begin (prima iterazione del while)

$Z := Z \cup S = CH \cup E = CEH$

$S := \{ A/Y \rightarrow V \in F \wedge A \in V \wedge Y \subseteq CEH \} = \{ E \}$ (per $C \rightarrow E$) = E

end

controllo while: $E \subseteq CEH$ (non abbiamo aggiunto niente di nuovo)

usciamo dal while

end

$CH^+ = CEH \subset R$

Quali chiavi?



SAPIENZA
UNIVERSITÀ DI ROMA
DIPARTIMENTO DI INFORMATICA

$R = (A, B, C, D, E, H)$

$F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$

begin

$Z := DH;$

$S := \{ A/Y \rightarrow V \wedge A \in V \wedge Y \subseteq DH \} = \emptyset$ (né DH, né D né H determinano funzionalmente alcun altro attributo)

controllo while ($S \not\subseteq Z ?$): $\emptyset \subset DH$ **quindi non** entriamo nella prima iterazione del while

end

$DH^+ = DH \subset R$

begin

$Z := EH;$

$S := \{ A/Y \rightarrow V \wedge A \in V \wedge Y \subseteq EH \} = \emptyset$ (né EH, né E né H determinano funzionalmente alcun altro attributo)

controllo while ($S \not\subseteq Z ?$): $\emptyset \subset EH$ **quindi non** entriamo nella prima iterazione del while

end

$EH^+ = EH \subset R$

- $R = (A, B, C, D, E, H)$ $F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$
- Dovremmo provare con altri sottoinsiemi di R con **tre attributi incluso H** ma ...
- ... è immediato osservare che **aggiungere A senza B o viceversa** non ci porta ad avere una chiusura che comprende tutto lo schema ...
- ... infatti l'attributo D dipende funzionalmente **SOLO** da sottoinsiemi di R in cui **compaiono sia A che B** ($AB \rightarrow CD, ABC \rightarrow D$)
- ... inoltre né A né B dipendono funzionalmente da altri attributi, nel qual caso avremmo potuto includerli in Z nel corso dell'algoritmo, e da qui poi includere D
- Quindi anche A e B devono essere in ogni chiave
- **Ma allora ABH è l'unica chiave** (aggiungere attributi ad ABH porta a violare il requisito di minimalità)

Dato il seguente schema di relazione

$R = (A, B, C, D, E, G, H)$

e il seguente insieme di dipendenze funzionali

$F = \{ AB \rightarrow D, G \rightarrow A, G \rightarrow B, H \rightarrow E, H \rightarrow G, D \rightarrow H \}$

determinare le 4 chiavi di R.

- **Suggerimento.** Cominciamo dagli attributi che ne **determinano funzionalmente** altri, e da quelli che **NON sono determinati** funzionalmente da altri

Dato il seguente schema di relazione

$$R = (A, B, C, D, E, G, H)$$

e il seguente insieme di dipendenze funzionali

$$F = \{ AB \rightarrow D, G \rightarrow A, G \rightarrow B, H \rightarrow E, H \rightarrow G, D \rightarrow H \}$$

determinare le 4 chiavi di R.

Nel nostro caso

AB (insieme!), G, D e H sono buoni candidati

E non determina nessun attributo, ma è determinato da H

C non determina nessun attributo ma non è neppure determinato da nessun attributo, quindi sarà in ogni chiave

Calcoliamo le chiusure



SAPIENZA
UNIVERSITÀ DI ROMA
DIPARTIMENTO DI INFORMATICA

$R = (A, B, C, D, E, G, H)$ $F = \{AB \rightarrow D, G \rightarrow A, G \rightarrow B, H \rightarrow E, H \rightarrow G, D \rightarrow H\}$

Calcoliamo le chiusure dei candidati che abbiamo individuato e **verifichiamo** la minimalità

$ABC^+ = ABCDEGH = R$

$AC^+ = AC$ $BC^+ = BC$ ABC è una chiave di R

$GC^+ = ABCDEGH = R$ GC è una chiave di R (inutile provare i sottoinsiemi)

$DC^+ = ABCDEGH = R$ DC è una chiave di R (inutile provare i sottoinsiemi)

$HC^+ = ABCDEGH = R$ HC è una chiave di R (inutile provare i sottoinsiemi)

Nota: ABC ha un attributo in più delle altre, ma soddisfa le condizioni per essere chiave dello schema in quanto abbiamo verificato che i suoi sottoinsiemi non lo siano a loro volta



- Per decidere «da dove cominciare» per cercare una chiave, possiamo partire dalle considerazioni fatte prima, iniziando dalle parti sinistre (determinanti) o dai singoli attributi.
- In alternativa si può anche cominciare dagli insiemi individuati dalle dipendenze funzionali : data una dipendenza funzionale $V \rightarrow W$ in F calcoliamo la chiusura dell'insieme di attributi $X = R - (W - V)$
 - La differenza interna evita di considerare dipendenze dovute alla riflessività
 - La differenza esterna esclude gli attributi «solo» dipendenti nella dipendenza in esame
 - Questi attributi potranno essere presi in considerazione de compaiono nella parte sinistra (determinante) di qualche altra dipendenza
- Se la chiusura di X contiene R , abbiamo identificato una superchiave ma ... occorre verificare che i suoi sottoinsiemi non determinino tutto R

- $R = \{A, B, C, D, E\}$
- $F = \{AB \rightarrow C, AC \rightarrow B, D \rightarrow E\}.$
- In base alle dipendenze, andiamo a calcolare le chiusure di ABDE, ACDE, ABCD
 - $(ABDE)^+ = ABDEC = R$
 - $(ACDE)^+ = ACDEB = R$
 - $(ABCD)^+ = ABCDE = R$
- MA SONO TUTTE E TRE CHIAVI?
 - $(AB)^+ = ABC$ mancano D ed E, ma aggiungendo D determiniamo anche E
 - $(ABD)^+ = ABCDE = R$ quindi ABDE e ABCD non sono chiavi
 - $(AC)^+ = ABC$ mancano D ed E, ma aggiungendo D determiniamo anche E
 - $(ACD)^+ = ABCDE = R$ quindi ACDE non è chiave
 - $(D)^+ = DE$ (era evidente)
 - $(E)^+ = E$ (era evidente)
- Le due chiavi sono ABD e ACD

- Dati uno schema di relazione R e un insieme di dipendenze funzionali F , Calcoliamo l'intersezione degli insiemi ottenuti come sopra, cioè degli insiemi $X=R-(W-V)$ con $V \rightarrow W$ in F
- Se l'intersezione di questi insiemi determina tutto R , allora questa intersezione è l'unica chiave di R
- Nel nostro esempio
 - $(ABDE \cap ACDE \cap ABCD)^+ = (AD)^+ = AD$

Quindi avremmo già potuto capire che esiste più di una chiave

- Se l'intersezione di questi insiemi **NON** determina tutto R allora esistono più chiavi che vanno **TUTTE** identificate per il test 3NF
- **RACCOMANDAZIONE IMPORTANTE:** quella descritta è una comoda verifica, ma se nel compito viene richiesto di **verificare che un insieme di attributi sia chiave, o di trovare la chiave, va usata SOLO la definizione** (verifica che la chiusura contenga R e che nessun sottoinsieme abbia la stessa proprietà)



- Una volta individuate le chiavi di uno schema di relazione, possiamo determinare se lo schema è in 3NF

Dato il seguente schema di relazione

$$R = (A, B, C, D, E, G, H)$$

e il seguente insieme di dipendenze funzionali

$$F = \{ AB \rightarrow CD, EH \rightarrow D, D \rightarrow H \}$$

determinare una chiave di R, e, sapendo che la chiave è unica, verificare che R non è in 3NF.

$R = (A, B, C, D, E, G, H)$ $F = \{ AB \rightarrow CD, EH \rightarrow D, D \rightarrow H \}$

Analizzando le dipendenze, **cominciamo col considerare gli insiemi di attributi che compaiono a sinistra** delle dipendenze stesse e calcoliamone le chiusure utilizzando l'algoritmo studiato :

$(AB)^+_F = \{A, B, C, D, H\}$ quindi AB non è superchiave

$(EH)^+_F = \{E, H, D\}$ quindi EH non è superchiave

$(D)^+_F = \{D, H\}$ quindi D non è superchiave

$R = (A, B, C, D, E, G, H)$ $F = \{ AB \rightarrow CD, EH \rightarrow D, D \rightarrow H \}$

- A questo punto osserviamo che **G non compare mai nelle dipendenze, quindi dovrà far parte sicuramente della chiave** dello schema.
- Inoltre **l'attributo E non compare mai a destra delle dipendenze, quindi anche questo attributo non può essere determinato dalla chiave a meno di non farne parte.**
Osservando che alla chiusura di AB mancano appunto questi due attributi, proviamo a calcolare la chiusura dell'insieme ABEG :

$(ABEG)^+_F = \{A, B, E, G, C, D, H\}$ quindi abbiamo trovato una superchiave.

- $R = (A, B, C, D, E, G, H)$ $F = \{ AB \rightarrow CD, EH \rightarrow D, D \rightarrow H \}$
 - Per verificare la seconda condizione, cioè la minimalità della superchiave trovata, dobbiamo controllare che la chiusura di nessun sottoinsieme di ABEG contenga tutti gli attributi dello schema. In teoria dovremmo calcolare le chiusure di tutti i possibili sottoinsiemi di ABEG, in pratica si possono fare delle considerazioni per fare un numero minore di calcoli. Per esempio, tenendo conto del fatto che è inutile verificare la chiusura di sottoinsiemi che non contengono G (abbiamo detto che deve far parte della chiave), cominciamo a calcolare le chiusure di ABG, BEG, AEG:
-
- $(ABG)^+_F = \{A, B, G, C, D, H\}$ quindi ABG non è superchiave
- $(BEG)^+_F = \{B, E, G\}$ quindi BEG non è superchiave
- $(AEG)^+_F = \{A, E, G\}$ quindi AEG non è superchiave

- $R = (A, B, C, D, E, G, H)$ $F = \{ AB \rightarrow CD, EH \rightarrow D, D \rightarrow H \}$
- Poiché non ci sono sottoinsiemi di tre attributi di ABEG che siano superchiave, è inutile calcolare le chiusure di sottoinsiemi con minor numero di attributi (sicuramente non sono in grado di determinare l'intero schema).
- Abbiamo quindi verificato entrambe le condizioni che ci permettono di affermare che **ABEG è chiave dello schema dato**

Per completezza notiamo che

- $(EHG)^+_F = \{E, H, D, G\}$ quindi EHG non è superchiave
- $(DG)^+_F = \{D, G, H\}$ quindi DG non è superchiave

Violazione della 3NF



$R = (A, B, C, D, E, G, H)$ $F = \{ AB \rightarrow CD, EH \rightarrow D, D \rightarrow H \}$

- Trovata la chiave **ABEG**, è sufficiente osservare la presenza **anche della sola** dipendenza $AB \rightarrow CD$ per verificare che R non è in 3NF.
- Infatti è una dipendenza **parziale** (la parte sinistra è contenuta propriamente in una chiave), e decomponendola in $AB \rightarrow C$ e $AB \rightarrow D$, in entrambi i casi l'attributo a destra non è primo e AB non è superchiave)
- Anche solo rimanendo in F , troviamo altre dipendenze che provocano violazioni. Infatti sia $EH \rightarrow D$ che $D \rightarrow H$ sono dipendenze transitive :
 - per entrambe l'attributo a destra non è primo e l'insieme a sinistra non è superchiave
 - data l'unica chiave dello schema, abbiamo che $ABEG-EH = ABG \neq \emptyset$ e $ABEG-D = ABEG \neq \emptyset$

Dato il seguente schema di relazione

$$R = (A, B, C, D, E, H)$$

e il seguente insieme di dipendenze funzionali

$$F = \{ AB \rightarrow CD, C \rightarrow E, AB \rightarrow E, ABC \rightarrow D \}$$

Abbiamo già verificato che **ABH** è una chiave per R.

Sapendo che ABH è l'unica chiave per R, controlliamo se R è in 3NF.

Basta osservare la presenza delle dipendenze parziali (l'insieme di attributi a destra è contenuto propriamente in una chiave) $AB \rightarrow CD$ (decomposta in $AB \rightarrow C$ e $AB \rightarrow D$) e $AB \rightarrow E$.

- le parti destre non sono parti della chiave
- le parti sinistre non sono superchiavi

Un vecchio esempio



SAPIENZA
UNIVERSITÀ DI ROMA
DIPARTIMENTO DI INFORMATICA

Cliente	Nome	C#	Città
	Rossi	C1	Roma
	Rossi	C2	Milano
	Bianchi	C3	Roma
	Verdi	C4	Roma

Ordine	O#	C#	A#	N-pezzi
	O1	C1	A1	100
	O2	C2	A2	200
	O3	C3	A2	150
	O4	C4	A3	200
	O1	C1	A2	200
	O1	C1	A3	100

Dipendenze?

Chiave/i?

3NF?

Conseguenze?



- Gli schemi vanno decomposti
- Dobbiamo trovare una decomposizione «buona»
 - Sottoschemi in 3NF
 - Vengono preservate le dipendenze in F^+
 - Il join delle istanze relative agli schemi della decomposizione consente di riottenere istanze «valide» dello schema di partenza