Closure of a set of attributes

בהינתן $S \subseteq X$ וסט של תלויות פונקציונאליות F, ניתן באמצעות אלגוריתם למצוא את X^+ - כל השדות התלויים בX לפי F.

מזה נובע הCorollary:

$$X \to Y \Leftarrow F \Leftrightarrow Y \subset X^+$$

כמו כן

$$\left(X^+\right)^+ = X^+$$

Applications of closure שימושים

1.

זה דומה לRecognizer(משהו שבודק אם מילה שייכת לשפה) - נתונה תלות, בודקים אם היא נכונה.

$X \to Y$ `holds on R?`

$Y \subset X^+ \Rightarrow$`holds`
$Y \not\subset X^+ \Rightarrow$`doesn't hold`

2.

זה דומה לGenerator(משהו שמפיק מילים של שפה) - נתון X וקבוצה של כללים, ומפיקים את התלויות הנובעות מהם.

$R(S), K \subset S$
K `is superkey?`

$K^+ = S \Rightarrow$`yes`
$K^+ \subsetneq S \Rightarrow$`no`

Armstrong's Axioms אסקיומות

1. רפלסקיביות - $Y \subseteq X \Rightarrow X \to Y$(Trivial FD)
2. Augemntation - $X \to Y \Rightarrow \forall_Z XZ \to YZ$ $\qquad X \cup Z \to Y \cup Z$
3. טרנזיטיביות - $X \to Y, Y \to Z \Rightarrow X \to Z$ $\qquad Z := Z - X$

סגור של תלויות פונקציונאליות Closure of a set of FDs

יהי יחס R, F קבוצה של תלויות פונקציונאליות שמתקיימות בR.

הגדרה

כל קבוצת תלויות פונקציונאליות $F1 \equiv F$ נקראת בסיס.

הגדרה

בסיס B נקרא בסיס מינימלי אם הוא מקיים את התנאים הבאים:

1. $\forall_{f \in B}$singleton right side - בצד ימין של כל תלות בB יש רק שדה אחד.
2. $\forall_{f \in B} B - f \not\equiv B$ - אי אפשר להוריד אף תלות מB בלי להחליש אותו.
3. $f = X_1 X_2 \to Y, X_2 \to Y$

הערה: תלויות טריוויאליות לא שייכות לבסיס מינימלי(לפי אקסיומה שנייה)

דוגמה

$$R(A, B, C)$$

$$F = \{A \to B, A \to C, B \to A, B \to C, C \to A, C \to B\}$$

רוצים למצוא את כל התלויות הנגזרות:
נוסיף בסיבים שקולים לפי ההנחות:

$$F_1 = F \cup \{AB \to C, AC \to B, BC \to A\}$$

$$F_2 = F_1 \cup \{A \to BC, B \to AC, C \to AB\}$$

נוסיף את התלויות הטריוויאליות:

$$F_3 = F_2 \cup \{A \to A, B \to B, C \to C\}$$

כל אלו בסיסיםפ - אבל לא מינימליים. נבנה בסיסים מינימליים:

$$B_1 = \{A \to B, B \to A, B \to C, C \to B\}$$

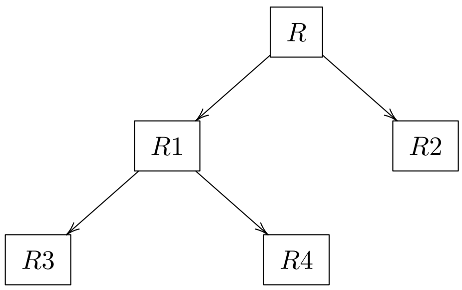
$$B_2 = \{A \to B, B \to C, C \to A\}$$

פירוק יחסים

Flights

Flight	Class	Price	Airpl	Num

ראינו שהיחס הזה עושה לנו הרבה בעיות של כפילות. נרצה לפרק אותו, אבל בצורה נכונה, כך שכל פעם מה שאנחנו מקבלים שקול למקור. יהיה ניתן לבנות מהפירוק הזה עץ, שבו כל אב שקול לאוסף הבנים שלו.



הטלת תלויות פונקציונאליות Projectin FDs

$$R\left(S\right), F$$

$$S1 \subset S, R1 = \pi_{S1} R$$

$$F1 \Leftarrow F$$

איך נחשב את $F1$?

שיטה

נתון יחס $R\left(A, B, C, D\right)$, וקבוצת תלויות פונקציונאליות $F = \{A \to B, B \to C, C \to D\}$.

$$R1\left(A,C,D\right),F1=?$$

לפי טרנזיטיביות, $A\rightarrow B\rightarrow C\rightarrow D$ לכן $A\rightarrow C\rightarrow D$, ולכן $F1=\{A\rightarrow C,C\rightarrow D\}$

אלגוריתם

Input: $R\left(S\right)$,$R1\left(S1\right)$,$R1=\pi_{S1}R$,F (on R)
Output: $F1,\ F1\Leftarrow F$
//Initialization
F1=$\emptyset$
//Basis
for(all $X\subset S1$){

compute X_F^+;
F1:=F1+$X\rightarrow Y$ |[1] $X\rightarrow Y$ nontrivial, $Y\in X_F^+$, $Y\in S1$

}
//Min basis
while($\exists f\in 1|f\Leftarrow F1-f$) F1:=F1-f
while($\exists X\rightarrow Y||X|>1,Z=X-Xi$)

if($Z\rightarrow Y\Leftarrow F$){

$X\rightarrow Y$:=$Z\rightarrow Y$

}

דוגמה

$$R\left(A,B,C,D\right)$$

$$F=\{A\rightarrow B,B\rightarrow C,C\rightarrow D\}$$

$$R1=\left(A,C,D\right)$$

$$F1=?$$

כבר עשינו את זה קודם ־ אבל לא פורמלית. עכשיו נעשה את זה לפי האלגוריתם.
הסגור של A יתן לנו סכימה ־ $A^+=\{A,B,C,D\}$. לכן אנחנו לא צריכים להתבונן ב$AB,AC,AD,ABC,\ldots$.
הסגור של C הוא $C^+=\{C,D\}$

[1] משמעות הסימן | הוא if. כלומר לקחת רק את האיברים שעונים על התנאים שמפורטים אחריו.

ובאותו אופן, $D^+ = \{D\}$.

אפשר לסכם את זה(לאחר הורדת התלויות הטריוויאליות, ואלו שלא נמצאות ב$R1$) באמצעות הבסיס $\{A \to C, A \to D, C \to D\}$. אבל התלות $A \to D$ נובעת מתלויות אחרות, ולכן ניתן להוריד אותה.

RDB(Relational Database) design

בגלל הבעיות ביחסים גדולים, רוצים לפרק ליחסים יותר קטנים:

$$\begin{tikzcd} & R\left(S\right) \arrow[dl] \arrow[dr] & \\ R1\left(S2\right) & & R2\left(S2\right) \end{tikzcd}$$

רוצים שיתקיים

$$S1 \subset S \qquad S2 \subset S$$

$$R1 = \pi_{S1} R \qquad R2 = \pi_{S2} R$$

$$S1 \cup S2 = S$$

כדי לבנות דבר כזה משתמשים ב

אלגוריתם נירמול Normalization Algorithm

מתחילים מצורה נורמלית ראשונה - $1NF$ - כל המידע ביחס אחד.

$$1NF \longrightarrow 2NF$$

ב$2NF$ יש פחות בעיות, אבל עדיין יש, ולכן ממשיכים לפרק:

$$1NF \longrightarrow 2NF \longrightarrow 3NF \longrightarrow BCNF \longrightarrow 4NF \longrightarrow \cdots$$

בקורס הזה נעצור ב$4NF$, אבל אפשר להמשיך עוד.