# Fondements des bases de données

## Couvertures d'ensembles de dépendances fonctionnelles

Équipe pédagogique BD

La notion de couverture est une relation d'équivalence entre des ensembles de contraintes.

### Couverture d'un ensemble de DFs

Soit $\Sigma$ et $\Gamma$ deux ensembles de DFs, $\Gamma$ est une couverture de $\Sigma$ *ssi*

$$\Gamma^{+} = \Sigma^{+}$$

- Une couverture d'un ensemble de DF est donc une représentation alternative
- Mais qui possède exactement la même sémantique.
- C'est exactement le même ensemble de DF qui est implicite.
- On a intérêt à choisir de bons représentants au sein des classes.

# Des critères pour de bon ensembles équivalents

## Propriétés des couvertures

- un ensemble $F$ de DF est dit non redondant s'il n'existe pas de couverture $G$ de $F$ telle que $G \subseteq F$ avec $G \neq F$.
- un ensemble $F$ de DF est dit minimum s'il n'existe pas de couverture $G$ de $F$ tel que $|G| \leq |F|$.
- $F$ est dit optimal s'il n'existe pas de couverture $G$ de $F$ avec moins d'attributs que dans $F$.

## Propriétés immédiates [1]

- une couverture minimum est non redondante ;
- une couverture optimum est minimum.

1. On utilise souvent aussi la contraposée de chaque implication.

# Algorithme : couverture minimum

**Data**: $F$ un ensemble de DF
**Result**: $G$ une couverture minimum de $F$
$G := \emptyset$
**for** $X \rightarrow Y \in F$ **do**
    $G := G \cup \{X \rightarrow X^+\}$
**end**
**for** $X \rightarrow X^+ \in G$ **do**
    **if** $G - \{X \rightarrow X^+\} \vdash X \rightarrow X^+$ **then**
        $G := G - \{X \rightarrow X^+\}$
    **end**
**end**
**return** G

- Cet algorithme est polynomial dans le nombre de DF dans $F$ et le nombre d'attributs dans $F$.
- La couverture minimum calculée par l'algorithme n'est pas forcément unique : d'autres couvertures peuvent avoir le même nombre de DF, mais être différentes.
- Parmi celles-ci, certaines sont optimum ; malheureusement, leur calcul est un problème difficile dans le cas général (NP-Complet).

# Exemple

$$F = \left\{ \begin{array}{lll} AB \rightarrow C & C \rightarrow A & BC \rightarrow D \\ ACD \rightarrow B & D \rightarrow EF & ABE \rightarrow C \\ CF \rightarrow BD & CE \rightarrow AF & \end{array} \right\}$$

1. A partir des df $X \rightarrow Y$ de $F$, construire $G$ l'ensemble des règles de la forme $X \rightarrow X^+$.
   - $AB \rightarrow AB^+ = ABCDEF = R$
   - $C \rightarrow C^+ = AC$
   - $BC \rightarrow BC^+ = R$
   - $ACD \rightarrow ACD^+ = R$
   - $D \rightarrow D^+ = DEF$
   - $ABE \rightarrow ABE^+ = R$
   - $CF \rightarrow CF^+ = R$
   - $CE \rightarrow CE^+ = R$

# Supprimer de $G$ les règles inutiles ;

| $G$ | |
|---|---|
| | $AB \to AB^+ = ABCDEF = R$ |
| | $C \to C^+ = AC$ |
| | $BC \to BC^+ = R$ |
| | $ACD \to ACD^+ = R$ |
| | $D \to D^+ = DEF$ |
| | $ABE \to ABE^+ = R$ |
| | $CF \to CF^+ = R$ |
| | $CE \to CE^+ = R$ |

## Supprimer de $G$ les règles inutiles ;

$AB \to AB^+ = ABCDEF = R$ utile ?
Oui si $AB^+_{G \setminus AB \to R} \neq R$.

| $G$ | |
|---|---|
| | $AB \to AB^+ = ABCDEF = R$ |
| | $C \to C^+ = AC$ |
| | $BC \to BC^+ = R$ |
| | $ACD \to ACD^+ = R$ |
| | $D \to D^+ = DEF$ |
| | $ABE \to ABE^+ = R$ |
| | $CF \to CF^+ = R$ |
| | $CE \to CE^+ = R$ |

- $AB^+ = AB \neq R \Rightarrow AB \to AB^+ = ABCDEF = R$ est utile !

# Supprimer de $G$ les règles inutiles ;

| $C \to C^+ = AC$ utile ? | |
|---|---|
| $G$ | |
| | $AB \to AB^+ = ABCDEF = R$ |
| | $C \to C^+ = AC$ |
| | $BC \to BC^+ = R$ |
| | $ACD \to ACD^+ = R$ |
| | $D \to D^+ = DEF$ |
| | $ABE \to ABE^+ = R$ |
| | $CF \to CF^+ = R$ |
| | $CE \to CE^+ = R$ |

- $C^+ = C \neq AC \Rightarrow C \to C^+ = AC$ est utile !

# Supprimer de $G$ les règles inutiles ;

$BC \to BC^+ = R$ utile ?

| $G$ |  |
|---|---|
| | $AB \to AB^+ = ABCDEF = R$ |
| | $C \to C^+ = AC$ |
| | $BC \to BC^+ = R$ |
| | $ACD \to ACD^+ = R$ |
| | $D \to D^+ = DEF$ |
| | $ABE \to ABE^+ = R$ |
| | $CF \to CF^+ = R$ |
| | $CE \to CE^+ = R$ |

- $BC^+ = R \Rightarrow BC \to BC^+ = R$ n'est pas utile ! . On la supprime de $G$.

# Supprimer de $G$ les règles inutiles ;

| $ACD \to ACD^+ = R$ utile ? | |
|---|---|
| $G$ | |
| | $AB \to AB^+ = ABCDEF = R$ |
| | $C \to C^+ = AC$ |
| | $ACD \to ACD^+ = R$ |
| | $D \to D^+ = DEF$ |
| | $ABE \to ABE^+ = R$ |
| | $CF \to CF^+ = R$ |
| | $CE \to CE^+ = R$ |

- $ACD^+ = R \Rightarrow ACD \to R$ n'est pas utile ! . On la supprime de $G$.

# Supprimer de $G$ les règles inutiles ;

$D \rightarrow DEF$ utile ?

| $G$ | |
|---|---|
| | $AB \rightarrow AB^+ = ABCDEF = R$ |
| | $C \rightarrow C^+ = AC$ |
| | $D \rightarrow D^+ = DEF$ |
| | $ABE \rightarrow ABE^+ = R$ |
| | $CF \rightarrow CF^+ = R$ |
| | $CE \rightarrow CE^+ = R$ |

- $D^+ = D \neq DEF \Rightarrow D \rightarrow DEF$ est utile.

# Supprimer de $G$ les règles inutiles ;

$ABE \to R$ utile ?

| $G$ | |
|---|---|
| | $AB \to AB^+ = ABCDEF = R$ |
| | $C \to C^+ = AC$ |
| | $D \to D^+ = DEF$ |
| | $ABE \to ABE^+ = R$ |
| | $CF \to CF^+ = R$ |
| | $CE \to CE^+ = R$ |

- $ABE^+ = R \Rightarrow ABE \to R$ est inutile. On la supprime

# Supprimer de $G$ les règles inutiles ;

$CF \to R$ utile ?

| $G$ | |
|---|---|
| | $AB \to AB^+ = ABCDEF = R$ |
| | $C \to C^+ = AC$ |
| | $D \to D^+ = DEF$ |
| | $CF \to CF^+ = R$ |
| | $CE \to CE^+ = R$ |

- $CF^+ = ACF \neq R \Rightarrow CF \to R$ est utile.

# Supprimer de $G$ les règles inutiles ;

$CE \rightarrow R$ utile ?

| $G$ | |
|---|---|
| | $AB \rightarrow AB^+ = ABCDEF = R$ |
| | $C \rightarrow C^+ = AC$ |
| | $D \rightarrow D^+ = DEF$ |
| | $CF \rightarrow CF^+ = R$ |
| | $CE \rightarrow CE^+ = R$ |

- $CE^+ = ACE \neq R \Rightarrow CE \rightarrow R$ est utile.

Une[2] couverture $G$ de $F$ est :

| $G$ | |
|---|---|
| | $AB \rightarrow ABCDEF = R$ |
| | $C \rightarrow AC$ |
| | $D \rightarrow DEF$ |
| | $CF \rightarrow R$ |
| | $CE \rightarrow R$ |

2. Attention, une couverture minimum n'est pas unique.

# Exercice

Calculer les couvertures des ensembles suivants

$F_1$

| | | |
|---|---|---|
| $AB \rightarrow D$ | $C \rightarrow A$ | $BC \rightarrow D$ |
| $D \rightarrow EF$ | $BE \rightarrow C$ | $CF \rightarrow B$ |
| $CE \rightarrow A$ | $CE \rightarrow G$ | |

$F_2$

| | | |
|---|---|---|
| $AB \rightarrow C$ | $C \rightarrow A$ | $BC \rightarrow D$ |
| $D \rightarrow EF$ | $BE \rightarrow C$ | $CF \rightarrow B$ |
| $CE \rightarrow F$ | | |

# Calcul d'une couverture canonique

Pour décomposer selon $F$, on va utiliser un ensemble $F'$ qui soit :

- Couverture de $F$ : $F^+ = F'^+$,
- Minimal : on ne peut pas retirer de DF en préservant toujours la couverture,
- Sans attributs redondants, ni à droite ni à gauche,
- Regroupé : il n'y a pas deux DF avec la même partie gauche.

On a vu des algorithmes qui permettent de produire une telle couverture. Ces étapes sont nécessaires pour assurer que les algorithmes vont bien produire un bon schéma !

# Réduction du nombre d'attribut pour un ensemble de DF

**Data**: Un ensemble *minimum* de DF $F$ sur $R$.

$Min := F$

`/* Réduction des parties gauches */`

**for** $X \rightarrow Y \in Min$ **do**

  $W := X$

  **for** $A \in X$ **do**

    **if** $Min \models (W - A) \rightarrow X$ **then** $W := W - \{A\}$

  **end**

  $Min := (Min - \{X \rightarrow Y\}) \cup \{W \rightarrow Y\}$

**end**

`/* Réduction des parties droites */`

**for** $X \rightarrow Y \in Min$ **do**

  $W := Y$

  **for** $A \in Y$ **do**

    $G := (Min - \{X \rightarrow Y\}) \cup \{X \rightarrow (W - A)\}$

    **if** $G \models X \rightarrow Y$ **then** $W := W - \{A\}$

  **end**

  $Min := (Min - \{X \rightarrow Y\}) \cup \{X \rightarrow W\}$

**end**

**return** $Min$

## Exemple de réduction

Soit l'ensemble de DFs $\Sigma$ [3] :

$$\Sigma = AB \rightarrow ABCDF;\ B \rightarrow BCD;\ DE \rightarrow F;\ E \rightarrow D$$

Couverture minimum $G$ :
$\{AB \rightarrow ABCDF;\ B \rightarrow BCD;\ E \rightarrow DEF\}$

- Réduction des parties gauches :

$$Min = AB \rightarrow ABCDF;\ B \rightarrow BCD;\ E \rightarrow DEF$$

- Réduction des parties droites :

$$Min = AB \rightarrow F;\ B \rightarrow CD;\ E \rightarrow DF$$

3. Attention, sur cet exemple, $\Sigma$ n'est pas une couverture minimum.

# Conclusion

## Ce qu'il faut retenir

- Couverture d'un ensemble de DFs : une représentation alternative véhiculant la même sémantique.
- Propriétés d'une (bonne) couverture : non redondante, minimum, optimum.
- Algorithme de calcul d'une couverture minimum.
- Algorithme de calcul d'une couverture "canonique" (minimum + réduction parties gauches et droites).

*Fin.*