Chapitre 5
Dépendances fonctionnelles

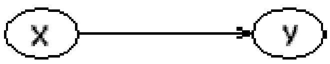
1. Dépendances fonctionnelles

i. Définition

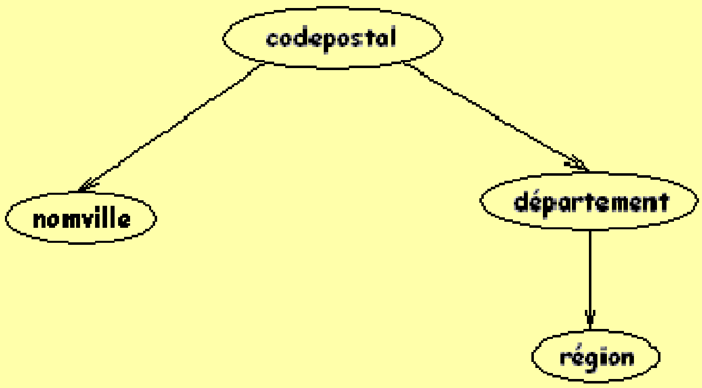
Afin de simplifier et d'éviter les redondances inutiles, on est amené à effectuer des décompositions. Le problème est alors d'éviter les pertes d'informations ou la création d'informations incohérentes. Pour résoudre ce problème, Codd (le père du modèle relationnel) a introduit la notion de dépendance fonctionnelle.

Soit une relation R et X et Y deux sous ensembles de R (X et Y peuvent être des attributs "simples" ou des attributs "composés", c'est à dire une liste d'attributs).

On dit que Y est fonctionnellement dépendant de X si la valeur de X permet de déterminer la valeur de Y de manière unique. On note X->Y et on dit que la connaissance de X entraîne celle de Y. On peut représenter une dépendance fonctionnelle par un graphe orienté :



exemple 1 :
Soit la relation **LOCALITE (nomville, codepostal, département, région)** qui décrit une ville. On peut envisager les dépendances fonctionnelles suivantes :
codepostal -> nomville car un codepostal correspond à une seule ville bien définie.
codepostal -> département car le code postal permet la détermination unique du département
département -> région car un département n'appartient qu'à une seule région.
Par contre, **ville -> codepostal** n'est pas une dépendance fonctionnelle car une ville peut avoir plusieurs codes postaux. De même, **nomville -> département** n'est pas une dépendance fonctionnelle car dans deux départements différents on pourrait trouver le même nom de ville. Le graphe des dépendances fonctionnelles est :



On notera que l'on a aussi la dépendance fonctionnelle **codepostal -> région** puisque le code postal permet de déterminer le département et qu'un département n'est que dans une seule région. Ceci montre une propriété des dépendances fonctionnelles : la transitivité.

ii. Propriétés des dépendances fonctionnelles

Les dépendances fonctionnelles (DF) possèdent un certain nombre de propriétés (que nous ne démontrons pas ici mais beaucoup sont évidentes) :

- **réflexivité** : tout ensemble d'attributs détermine lui-même ou une partie de lui-même
  X -> X ou (X,Y) -> X

par exemple nomville -> nomville
(nomville, département) -> nomville

- **augmentation** : Si X -> Y alors (X,Z) -> (Y,Z)

par exemple codepostal -> nomville
donc (codepostal, région) -> (nomville, région)

- **transitivité** : Si X -> Y et si Y-> Z alors X -> Z

par exemple codepostal -> département et département -> région
donc codepostal -> région
Les propriétés de réflexivité, d'augmentation et de transitivité sont appelées axiomes d'Armstrong.

- **pseudo-transitivité** : Si X -> Y et si (T,Y) -> Z alors (T,X) -> Z

par exemple codepostal -> nomville et (département, nomville) -> région
donc (département,codepostal) -> région

- **union** : Si X -> Y et si X -> Z alors X -> (Y,Z)

par exemple codepostal -> nomville et codepostal -> département
donc codepostal -> (nomville, département)

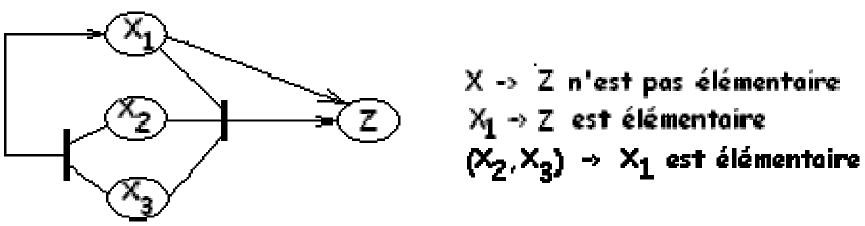
- **décomposition** : Si X -> Y et si Z est inclus dans Y alors X -> Z

par exemple codepostal -> (nomville, département)
donc codepostal -> département et codepostal -> nomville

iii. Dépendances fonctionnelles élémentaires

Pour simplifier l'étude des dépendances fonctionnelles, il est commode de ne prendre en considération que les dépendances fonctionnelles élémentaires.
Une dépendance fonctionnelle X -> Z est élémentaire s'il n'existe pas d'attribut Y inclus dans X tel que Y -> Z.

Si X est un attribut, la dépendance fonctionnelle X -> Z est automatiquement élémentaire.
Si X est un agrégat d'attributs, par exemple X = (X1, X2, X3) et si X -> Z et si X1 -> Z, la dépendance fonctionnelle X ->Z n'est pas élémentaire.



Exemple 2 :
Dans l'exemple considéré plus haut
(codepostal, département) -> nomville et codepostal -> nomville sont des dépendances fonctionnelles, mais seule la seconde est élémentaire.

On notera que les dépendances fonctionnelles élémentaires (DFE) ne possèdent plus que les propriétés de réflexivité et de transitivité.

iv. Fermeture

La fermeture d'un ensemble d'attributs X, notée X+, représente l'ensemble des attributs de R qui peuvent être déduits de X à partir d'une famille de dépendances fonctionnelles en appliquant les axiomes d'Armstrong. Ainsi, Y sera inclus dans X+ si et seulement si X -> Y.

Algorithme de recherche de la fermeture de X :

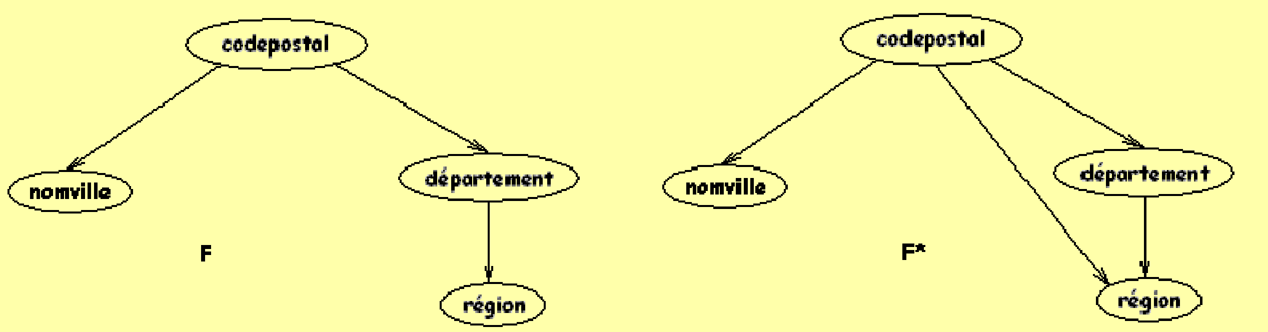
1) initialiser X+ à X,
2) trouver une dépendance fonctionnelle de F possédant en partie gauche des attributs inclus dans X+,
3) ajouter dans X+ les attributs placés en partie droite de la dépendance fonctionnelle,
4) répéter les étapes à partir de 2) jusqu'à ce que X+ n'évolue plus.

Exemple 3 :
Soit la famille de dépendances fonctionnelles F = {A-> C, A -> E, B -> D, C -> D, (B,E) -> G, (D,E) -> C} et cherchons la fermeture de AD
On part avec X+ = AD
On a A-> C donc X+ = ADC
On a A -> E donc X+ = ADCE
On a C -> D, (D,E) -> C ce qui ne change rien. Donc la fermeture de C est X+ = ADCE

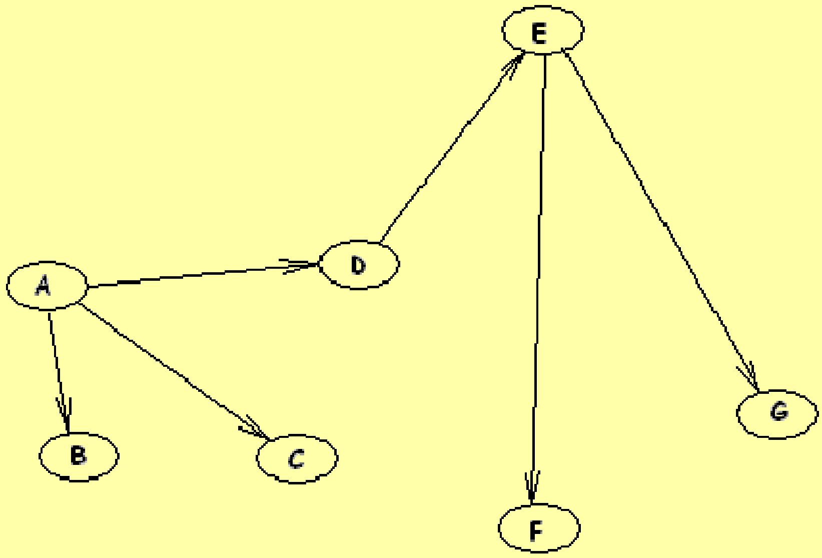
v. Fermeture transitive

Considérons, pour un ensemble d'attributs donné, un ensemble F de DFE. En construisant, par transitivité, toutes les DFE possibles à partir de F, on obtient un ensemble F* de DFE appelé fermeture transitive de F.

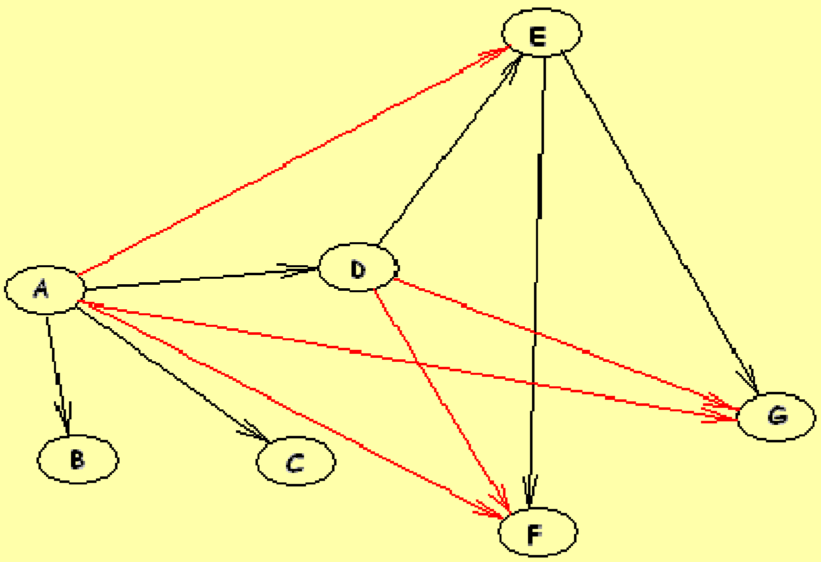
exemple 4 :
en reprenant l'exemple précédent, on a



exemple 5:
Considérons les dépendances fonctionnelles définies par le graphe ci-dessous :



La fermeture transitive est décrite par le graphe suivant (on a ajouté en rouge les dépendances fonctionnelles supplémentaires)

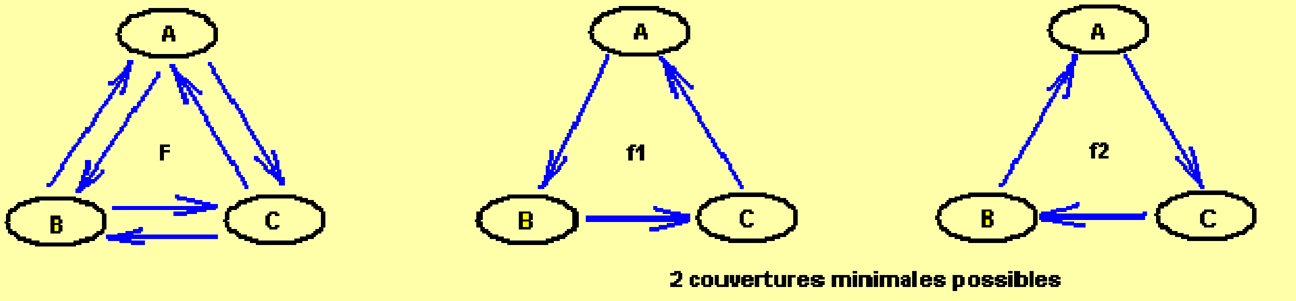


Deux ensembles de DFE, F et F', peuvent avoir la même fermeture transitive : F* = F'* . On dira dans ce cas que F et F' sont équivalents.

vi. Couverture minimale

Etant donné un ensemble F de DFE, on peut envisager de retrouver (par transitivité) F à partir d'un sous-ensemble de F. Le **plus petit sous-ensemble f** (c'est à dire comportant le moins de DFE) **permettant de reconstruire F est appelé une** couverture minimale **de F.** Signalons que cette couverture minimale n'est pas unique en général. Notons également que f et F sont équivalents : f* = F*

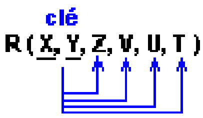
Exemple6:



Dans la couverture minimale, il n'y a plus que des **dépendances fonctionnelles directes**. On dit qu'une dépendance fonctionnelle X -> Z est directe si il n'existe pas d'attribut ou groupe d'attributs Y tels que X --> Y et Y -> Z.

vii. Clé d'une relation

La notion de dépendance fonctionnelle permet de préciser la notion de clé d'une relation que nous avons introduite plus haut : une clé K d'une relation R est un attribut ou un groupe d'attributs tel qu'il existe une dépendance fonctionnelle élémentaire entre K et tout autre attribut de R.



(X,Y) est la clé de la relation R