Fondements des Bases de données : *relation exemple*

M1 - Informatique

Lhouari Nourine, Karima Ennaoui et *Simon Vilmin*.

Institut d'informatique, ISIMA

2020-2021

Slides inspirées du cours de Jean-Marc Petit :
`https://perso.liris.cnrs.fr/jmpetit/ferme/doku.php/teaching`

Construction par l'exemple

- Représentation *« par l'exemple »* d'ensemble de contraintes : on manipule des valeurs, visualisation plus simple des éventuels conflits, incohérences, mauvaise conception, ...
- Une *« bonne »* relation exemple par rapport à des contraintes :
  - doit *respecter* ces contraintes ... et *seulement* celles-ci !
  - on l'appelle une *relation d'Armstrong*.
  - pour nous, focus (encore) sur les *dépendances fonctionnelles*.
- Deux principaux domaines d'application :
  - Conception par l'exemple
  - Échantillonnage, compréhension des bases de données existantes
- Exemple introductif : les nuages !
  - Attributs, $R = \{Forme, Structure, Nom, Couleur\}$,
  - des DFs : Forme, Structure $\rightarrow$ Nom ; Nom $\rightarrow$ Forme ; Couleur $\rightarrow$ Structure.

Relation exemple : ni pas assez

Forme	Structure	Nom	Couleur
couche	cristaux	cirrostratus	translucide
couche	c + g	altostratus	gris
mouton	c + g	cumulonimbus	gris
mouton	gouttelettes	cumulonimbus	blanc
enclume	gouttelettes	cumulonimbus	blanc
enclume	gouttelettes	cumulonimbus	noir

- $R = \{Forme, Structure, Nom, Couleur\}$, DFs : Forme, Structure $\rightarrow$ Nom ; Nom $\rightarrow$ Forme ; Couleur $\rightarrow$ Structure.

Relation exemple : ni pas assez

Forme	Structure	Nom	Couleur
couche	cristaux	cirrostratus	translucide
couche	c + g	altostratus	gris
mouton	c + g	cumulonimbus	gris
mouton	gouttelettes	cumulonimbus	blanc
enclume	gouttelettes	cumulonimbus	blanc
enclume	gouttelettes	cumulonimbus	noir

- $R = \{Forme, Structure, Nom, Couleur\}$, DFs : Forme, Structure $\longrightarrow$ Nom ; Nom $\longrightarrow$ Forme ; Couleur $\longrightarrow$ Structure.

✗ La DF Nom $\longrightarrow$ Forme n'est pas respectée !

Relation exemple : ni trop

Forme	Structure	Nom	Couleur
couche	cristaux	cirrostratus	translucide
couche	c + g	altostratus	gris
mouton	c + g	cumulonimbus	gris
mouton	gouttelettes	cumulonimbus	blanc
rouleaux	gouttelettes	altocumulus	b + g

- $R = \{Forme, Structure, Nom, Couleur\}$, DFs : Forme, Structure $\rightarrow$ Nom ; Nom $\rightarrow$ Forme ; Couleur $\rightarrow$ Structure.

Relation exemple : ni trop

Forme	Structure	Nom	Couleur
couche	cristaux	cirrostratus	translucide
couche	c + g	altostratus	gris
mouton	c + g	cumulonimbus	gris
mouton	gouttelettes	cumulonimbus	blanc
rouleaux	gouttelettes	altocumulus	b + g

- R = $\{Forme, Structure, Nom, Couleur\}$, DFs : Forme, Structure $\rightarrow$ Nom ; Nom $\rightarrow$ Forme ; Couleur $\rightarrow$ Structure.
- ✓ La relation satisfait les DFs

Relation exemple : ni trop

Forme	Structure	Nom	Couleur
couche	cristaux	cirrostratus	translucide
couche	c + g	altostratus	gris
mouton	c + g	cumulonimbus	gris
mouton	gouttelettes	cumulonimbus	blanc
rouleaux	gouttelettes	altocumulus	b + g

- R = {*Forme*, *Structure*, *Nom*, *Couleur*}, DFs : Forme, Structure → Nom ; Nom → Forme ; Couleur → Structure.
- ✓ La relation satisfait les DFs
- ✗ mais Forme, Structure → Nom, Couleur est aussi valide (Forme, Structure est une clé) !

Relation exemple : juste ce qu'il faut

Forme	Structure	Nom	Couleur
couche	cristaux	cirrostratus	translucide
couche	c + g	altostratus	gris
mouton	c + g	cumulonimbus	gris
mouton	gouttelettes	cumulonimbus	blanc
mouton	gouttelettes	cumulonimbus	noir

- R = $\{Forme, Structure, Nom, Couleur\}$, DFs : Forme, Structure $\rightarrow$ Nom ; Nom $\rightarrow$ Forme ; Couleur $\rightarrow$ Structure.

Relation exemple : juste ce qu'il faut

Forme	Structure	Nom	Couleur
couche	cristaux	cirrostratus	translucide
couche	c + g	altostratus	gris
mouton	c + g	cumulonimbus	gris
mouton	gouttelettes	cumulonimbus	blanc
mouton	gouttelettes	cumulonimbus	noir

- $R = \{Forme, Structure, Nom, Couleur\}$, DFs : Forme, Structure $\rightarrow$ Nom ; Nom $\rightarrow$ Forme ; Couleur $\rightarrow$ Structure.

✓ La relation satisfait les DFs

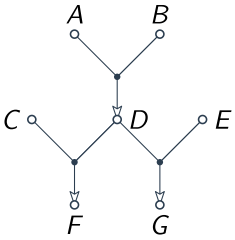
Relation exemple : juste ce qu'il faut

Forme	Structure	Nom	Couleur
couche	cristaux	cirrostratus	translucide
couche	c + g	altostratus	gris
mouton	c + g	cumulonimbus	gris
mouton	gouttelettes	cumulonimbus	blanc
mouton	gouttelettes	cumulonimbus	noir

- $R = \{Forme, Structure, Nom, Couleur\}$, DFs : Forme, Structure $\rightarrow$ Nom ; Nom $\rightarrow$ Forme ; Couleur $\rightarrow$ Structure.
- ✓ La relation satisfait les DFs
- ✓ et aucune autre qui ne découle pas de celles de départ

Dans l'épisode précédent

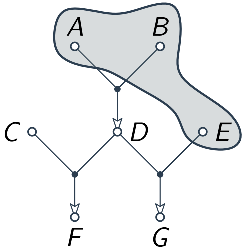
- Les *contraintes d'intégrité* définissent ce qu'est une bonne base de données,
- Σ ensemble de *dépendances fonctionnelles* (DFs) $X \rightarrow Y$ sur un *schéma de relation* R,
- $\Sigma \models X \rightarrow Y$ est l'*implication logique* : « on déduit $X \rightarrow Y$ à partir de Σ ».
- $X^\Sigma = \{A \in \mathrm{R} \mid \Sigma \models X \rightarrow A\}$, pour tout $X \subseteq \mathrm{R}$ (algo de *fermeture*).



- $\Sigma = \{AB \rightarrow D, CD \rightarrow F, DE \rightarrow G\}$,
- $ABE^\Sigma = ABDEG$.

Dans l'épisode précédent

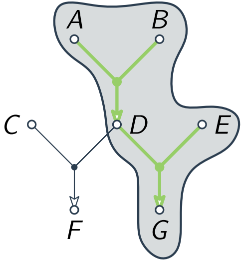
- Les *contraintes d'intégrité* définissent ce qu'est une bonne base de données,
- Σ ensemble de *dépendances fonctionnelles* (DFs) $X \to Y$ sur un *schéma de relation* R,
- $\Sigma \models X \to Y$ est l'*implication logique* : « on déduit $X \to Y$ à partir de Σ ».
- $X^\Sigma = \{A \in \mathrm{R} \mid \Sigma \models X \to A\}$, pour tout $X \subseteq \mathrm{R}$ (algo de *fermeture*).



- $\Sigma = \{AB \to D, CD \to F, DE \to G\}$,
- $ABE^\Sigma = ABDEG$.

Dans l'épisode précédent

- Les *contraintes d'intégrité* définissent ce qu'est une bonne base de données,
- Σ ensemble de *dépendances fonctionnelles* (DFs) $X \rightarrow Y$ sur un *schéma de relation* R,
- $\Sigma \models X \rightarrow Y$ est l'*implication logique* : « on déduit $X \rightarrow Y$ à partir de Σ ».
- $X^{\Sigma} = \{A \in \mathrm{R} \mid \Sigma \models X \rightarrow A\}$, pour tout $X \subseteq \mathrm{R}$ (algo de *fermeture*).



- $\Sigma = \{AB \rightarrow D, CD \rightarrow F, DE \rightarrow G\}$,
- $ABE^{\Sigma} = ABDEG$.

Relation d'Armstrong

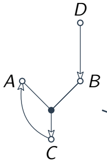
Définition - Relation d'Armstrong

Soit Σ un ensemble de DFs sur R. Une relation r sur R est une *relation d'Armstrong* (relation exemple) pour Σ si pour toute DF $X \rightarrow Y$:

$$\Sigma \models X \rightarrow Y \text{ si et seulement si } r \models X \rightarrow Y$$

- Une relation d'Armstrong pour Σ satisfait Σ et *rien d'autre*.
- Les relations d'Armstrong *existent aussi* pour les autres types de contraintes !

Exemple



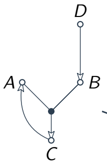
r_1	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	5	c	z	4
t_5	5	c	z	6

r_2	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

r_3	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	5	c	t	6

- Où est la relation d'Armstrong de $\Sigma = \{D \rightarrow B, AB \rightarrow C, C \rightarrow A\}$?

Exemple



r_1	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	5	c	z	4
t_5	5	c	z	6

✗

r_2	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

✓

r_3	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	5	c	t	6

✗

- Où est la relation d'Armstrong de $\Sigma = \{D \rightarrow B, AB \rightarrow C, C \rightarrow A\}$?

Existence

- **Question :** peut-on trouver une relation d'Armstrong pour n'importe quel Σ ?

Existence

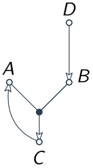
- **Question :** peut-on trouver une relation d'Armstrong pour n'importe quel Σ ?

OUI !!!

- Il existe même une procédure « *algorithmique* » pour le faire !
- Hypothèse : pas de DFs $\emptyset \rightarrow X$, domaines infinis.

Théorème (Armstrong 1974) : Soit Σ un ensemble de DFs sur R. Il existe une relation d'Armstrong r pour Σ.

Armstrong wars : Épisode I



?

?

?

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

Opérateur de fermeture

Définition - Opérateur de fermeture

Soit ϕ: $2^{\mathsf{R}} \to 2^{\mathsf{R}}$ une application. On dit que ϕ est un *opérateur de fermeture* (ou simplement une *fermeture*) si pour tout $X, Y \subseteq \mathsf{R}$:

- $X \subseteq \phi(X)$ (*extensive*),
- si $X \subseteq Y$ alors $\phi(X) \subseteq \phi(Y)$ (*monotone*),
- $\phi(\phi(X)) = \phi(X)$ (*idempotente*).

- **Exercice :** les applications suivantes sont-elles des fermetures ?
  - Pour tout $X \subseteq \mathsf{R}$, $\phi(X) = \mathsf{R}$.
  - Soit $Y \subseteq \mathsf{R}$. Pour tout $X \subseteq \mathsf{R}$, $\phi(X) = X \cap Y$.
  - Soit $Y \subseteq \mathsf{R}$. Pour tout $X \subseteq \mathsf{R}$, $\phi(X) = X \cup Y$.
  - Soit $G = (V, E)$ un graphe. Pour tout $X \subseteq V$, $\phi(X) = X \cup \{u \in V \mid \exists x \in X \text{ tel que } (x, u) \in E\}$.

Système de fermeture

Définition - Système de fermeture

Soit $\mathcal{F} \subseteq 2^{\mathsf{R}}$ une *famille d'ensembles sur* R. On dit que $\mathcal{F}$ est un *système de fermeture*, et ses éléments sont des *fermés*, si :

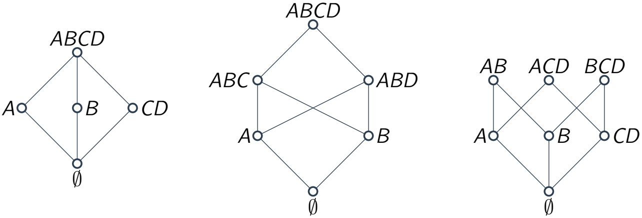
- $\mathsf{R} \in \mathcal{F}$,
- $F_1, F_2 \in \mathcal{F}$ implique que $F_1 \cap F_2 \in \mathcal{F}$.

- Les systèmes de fermeture sont partout !
  - logique, data mining (FCA), base de données
  - géométrie, optimisation, graphes, ...

Exemple (fil rouge)

- Parmi ces exemples, un est un système de fermeture :
  - $\mathcal{F}_1 = \{\emptyset, A, B, CD, ABCD\}$,
  - $\mathcal{F}_2 = \{\emptyset, A, B, ABC, ABD, ABCD\}$,
  - $\mathcal{F}_3 = \{\emptyset, A, B, C, AB, ACD, BCD\}$

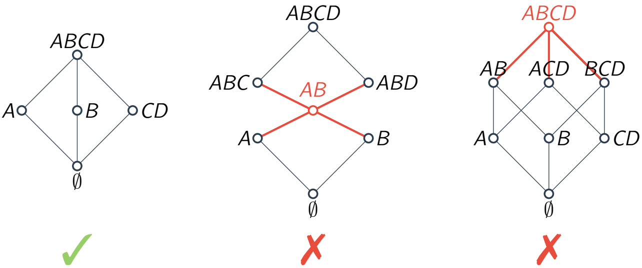
$(\mathcal{F}_1, \subseteq)$ $(\mathcal{F}_2, \subseteq)$ $(\mathcal{F}_3, \subseteq)$



Exemple (fil rouge)

- Parmi ces exemples, un est un système de fermeture :
  - $\mathcal{F}_1 = \{\emptyset, A, B, CD, ABCD\}$,
  - $\mathcal{F}_2 = \{\emptyset, A, B, ABC, ABD, ABCD\}$,
  - $\mathcal{F}_3 = \{\emptyset, A, B, C, AB, ACD, BCD\}$

$(\mathcal{F}_1, \subseteq)$ $(\mathcal{F}_2, \subseteq)$ $(\mathcal{F}_3, \subseteq)$



Propriétés

- Une fermeture est associé à un *unique* système de fermeture.
- Un système de fermeture est associé à un *unique* opérateur de fermeture.
- Autrement dit : ce sont deux *représentations* d'une *même information* !

Propriété : Soit ϕ une fermeture sur R. Alors la collection $\mathfrak{F}_\phi = \{X \subseteq \mathrm{R} \mid \phi(X) = X\}$ est un système de fermeture.

Propriété : Soit $\mathfrak{F}$ un système de fermeture sur R. Alors l'opérateur $\phi_\mathfrak{F}$ telle que $\phi_\mathfrak{F}(X) = \bigcap\{F \in \mathfrak{F} \mid X \subseteq F\}$ est une fermeture.

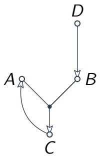
Le rapport avec nos DFs

- L'opérateur $(\cdot)^{\Sigma}$ associé à un Σ est une *fermeture* !
- Donc, la collection des fermés de Σ

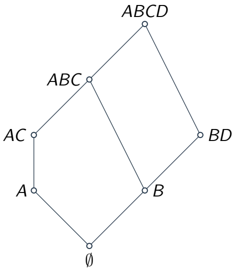
$$\mathcal{F}_{\Sigma} = \{X^{\Sigma} \mid X \subseteq \mathrm{R}\} = \{X \subseteq \mathrm{R} \mid X = X^{\Sigma}\}$$

est un *système de fermeture* !

Propriété : Soit Σ un ensemble de DFs. La famille $\mathcal{F}_{\Sigma}$ est un système de fermeture.

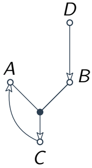


Σ



$(\mathcal{F}, \subseteq)$

Armstrong wars : Épisode II



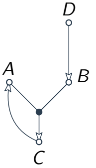
?

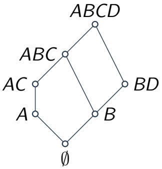
?

?

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

Armstrong wars : Épisode II





?

?

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

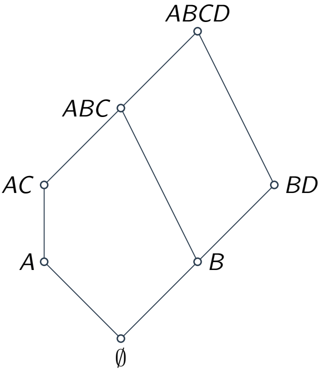
Représentation d'un système

Définition - Infs-irréductibles, MAX-SETS

Soit $\mathcal{F}$ un système de fermeture sur R, et $M \in \mathcal{F}$, $M \neq$ R. On dit que M est un *inf-irréductible* (ou *MAX-SET*) si quelque soit $F_1, F_2 \in \mathcal{F}$, $M = F_1 \cap F_2$ implique que $M = F_1$ ou $M = F_2$. Les inf-irréductibles de $\mathcal{F}$ sont notés $\mathcal{M}$.

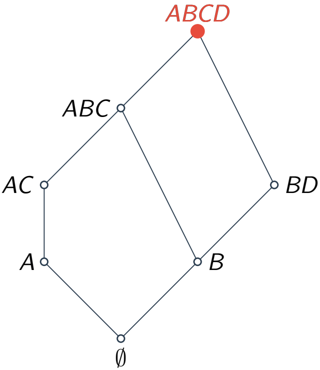
- Dans $\mathcal{F}$, certains éléments sont obtenus par l'*intersection* d'autres, ils sont *redondants* et donc *réductibles*.
- Quand on enlève tous ces éléments redondants, il reste ceux qu'on ne peut pas obtenir par intersection : ce sont les *infs-irréductibles* !
- Autrement dit : $\mathcal{M}$ est la partie « génératrice » de $\mathcal{F}$, on peut *reconstruire* tout $\mathcal{F}$ depuis $\mathcal{M}$ (par $\cap$).
- $\mathcal{M}$ concentre toute la « connaissance » de $\mathcal{F}$.

Exemple (fil rouge)



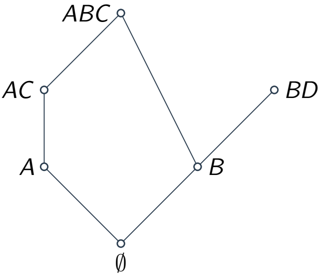
- $\mathcal{F} = \{\emptyset, A, B, AC, ABC, BD, ABCD\}$,
-

Exemple (fil rouge)



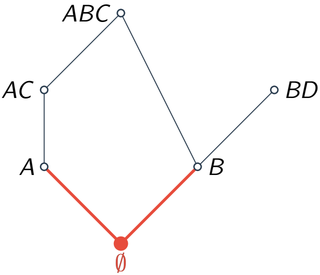
- $\mathcal{F} = \{\emptyset, A, B, AC, ABC, BD, ABCD\}$,
- $ABCD$ n'est *pas* inf-irréductible par définition

Exemple (fil rouge)



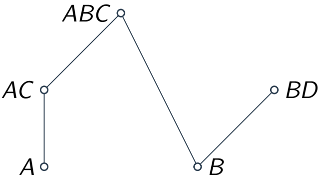
- $\mathcal{F} = \{\emptyset, A, B, AC, ABC, BD, ABCD\}$,
- $ABCD$ n'est *pas* inf-irréductible par définition

Exemple (fil rouge)



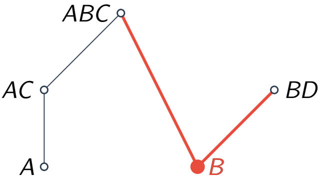
- $\mathcal{F} = \{\emptyset, A, B, AC, ABC, BD, ABCD\}$,
- $\emptyset$ est l'intersection de A et B, il n'est *pas* inf-irréductible

Exemple (fil rouge)



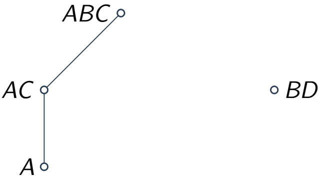
- $\mathcal{F} = \{\emptyset, A, B, AC, ABC, BD, ABCD\}$,
- $\emptyset$ est l'intersection de A et B, il n'est *pas* inf-irréductible

Exemple (fil rouge)



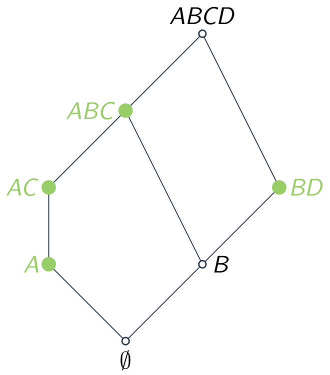
- $\mathcal{F} = \{\emptyset, A, B, AC, ABC, BD, ABCD\}$,
- B est l'intersection de ABC et BD, il n'est *pas* inf-irréductible

Exemple (fil rouge)



- $\mathcal{F} = \{\emptyset, A, B, AC, ABC, BD, ABCD\}$,
- B est l'intersection de ABC et BD, il n'est *pas* inf-irréductible

Exemple (fil rouge)



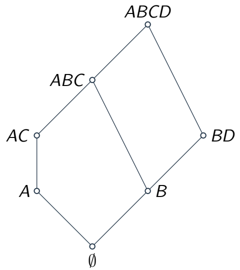
- $\mathcal{F} = \{\emptyset, A, B, AC, ABC, BD, ABCD\}$,
- Au final, les infs-irréductibles sont A, AC, ABC, BD.

Propriétés

- Soit $\mathcal{F}$ un système de fermeture sur R, et $A \in$ R.
- On définit $MAX(A) = max_{\subseteq}\{F \in \mathcal{F} \mid A \notin F\}$.
- En d'autre termes : M est dans $MAX(A)$ s'il *ne contient pas* A et si quelque soit le fermé M' « au dessus » de M ($M \subseteq M'$), M' contient A.

Propriété : Pour tout $A \in$ R, $MAX(A) \subseteq \mathcal{M}$. De plus, $\mathcal{M} = \bigcup_{A \in \text{R}} MAX(A)$.

- Considérons C, son fermé est AC.
-
-

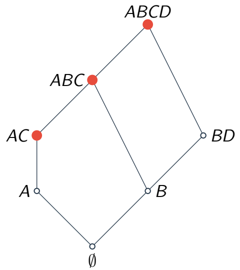


Propriétés

- Soit $\mathcal{F}$ un système de fermeture sur R, et $A \in$ R.
- On définit $MAX(A) = max_{\subseteq}\{F \in \mathcal{F} \mid A \notin F\}$.
- En d'autre termes : M est dans $MAX(A)$ s'il *ne contient pas* A et si quelque soit le fermé M' « au dessus » de M ($M \subseteq M'$), M' contient A.

Propriété : Pour tout $A \in$ R, $MAX(A) \subseteq \mathcal{M}$. De plus, $\mathcal{M} = \bigcup_{A \in R} MAX(A)$.

- Considérons C, son fermé est AC.
- $AC, ABC, ABCD$ ne *sont pas intéressants* (ils ont C),
-

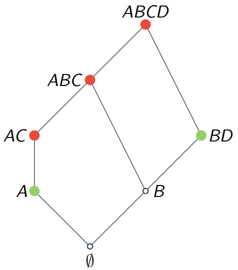


Propriétés

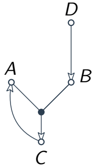
- Soit $\mathcal{F}$ un système de fermeture sur R, et $A \in$ R.
- On définit $MAX(A) = max_{\subseteq}\{F \in \mathcal{F} \mid A \notin F\}$.
- En d'autre termes : M est dans $MAX(A)$ s'il *ne contient pas* A et si quelque soit le fermé M' « au dessus » de M ($M \subseteq M'$), M' contient A.

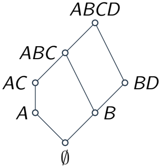
Propriété : Pour tout $A \in$ R, $MAX(A) \subseteq \mathcal{M}$. De plus, $\mathcal{M} = \bigcup_{A \in R} MAX(A)$.

- Considérons C, son fermé est AC.
- $AC, ABC, ABCD$ ne *sont pas intéressants* (ils ont C),
- Les *maximaux restants* sont A et BD



Armstrong wars : Épisode III



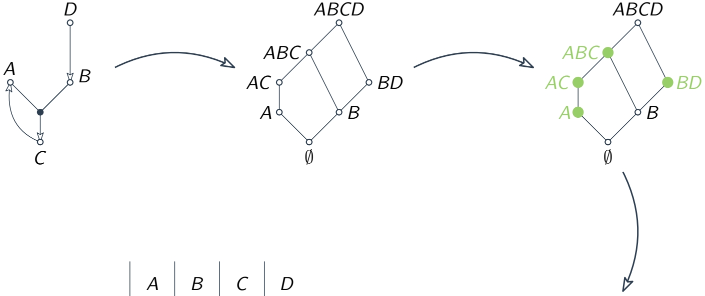


?

?

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

Armstrong wars : Épisode III



?

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

Agrées

- **Idée :** deux tuples d'une relation r respectent forcément les DFs de r. Donc, si on compare ces tuples et que l'on note les attributs sur lesquels ils sont en *en accord* (égaux), on obtient un *fermé* des DFs de r !

Définition - Agree set

Soit r une relation sur R et $t, t' \in r$. L'*agree set* de t, t', noté $ag(t, t')$ est l'ensemble des attributs de R sur lesquels t et t' sont égaux, c.à.d, $ag(t, t') = \{A \in \mathrm{R} \mid t[A] = t'[A]\}$. L'ensemble des agree sets de r est noté $ag(r)$.

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

Agrées

- **Idée :** deux tuples d'une relation r respectent forcément les DFs de r. Donc, si on compare ces tuples et que l'on note les attributs sur lesquels ils sont en *en accord* (égaux), on obtient un *fermé* des DFs de r !

Définition - Agree set

Soit r une relation sur R et $t, t' \in r$. L'*agree set* de t, t', noté $ag(t, t')$ est l'ensemble des attributs de R sur lesquels t et t' sont égaux, c.à.d, $ag(t, t') = \{A \in \mathrm{R} \mid t[A] = t'[A]\}$. L'ensemble des agree sets de r est noté $ag(r)$.

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

Agrées

- **Idée :** deux tuples d'une relation r respectent forcément les DFs de r. Donc, si on compare ces tuples et que l'on note les attributs sur lesquels ils sont en *en accord* (égaux), on obtient un *fermé* des DFs de r !

Définition - Agree set

Soit r une relation sur R et $t, t' \in r$. L'*agree set* de t, t', noté $ag(t, t')$ est l'ensemble des attributs de R sur lesquels t et t' sont égaux, c.à.d, $ag(t, t') = \{A \in \mathrm{R} \mid t[A] = t'[A]\}$. L'ensemble des agree sets de r est noté $ag(r)$.

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

$ag(t_2, t_4) = AC$

Le super théorème

Théorème : Soit Σ un ensemble de DFs et r une relation sur R. Alors, r est une relation d'Armstrong pour Σ si et seulement si $\mathcal{M} \subseteq ag(r) \subseteq \mathcal{F}$.

- Première idée de construction :
  - pour chaque $M \in \mathcal{M}$, on crée deux tuples t_M^1, t_M^2 qui codent exactement M (agree set) ;
  - pour tout $M' \neq M$, on garantit que $ag(t_M^i, t_{M'}^j) = \emptyset$ en faisant en sorte que les valeurs utilisées pour coder M soient disjointes de celles utilisées pour coder M' ;
  - on obtient une relation avec $2 \times |\mathcal{M}|$ tuples.
- On peut faire mieux : en $|\mathcal{M}| + 1$ tuples.

Algo relation d'Armstrong

Algorithme ARMSTRONG($\mathcal{M}$)

Data: $\mathcal{M}$ un ensemble d'infs sur R
Result: r, une relation d'Armstrong

```
// Pour tout A ∈ R, on fixe dom(A) = ℕ
foreach A ∈ R do t_0[A] := 0 end
r := {t_0} ;
i := 1 ;
foreach M ∈ 𝓜 do
    foreach A ∈ R do
        if A ∈ M then t_i[A] := t_{i-1}[A] ;
        else t_i[A] := t_{i-1}[A] + 1 ;
    end
    r := r ∪ t_i ;
    i := i + 1 ;
end
return r ;
```

Trace sur un exemple (fil rouge)

$\mathcal{M}$		A	B	C	D
	t_0	0	0	0	0
A					
BD					
AC					
ABC					

Trace sur un exemple (fil rouge)

$\mathcal{M}$		A	B	C	D
	t_0	0	0	0	0
A	t_1	0	1	1	1
BD					
AC					
ABC					

+1 +1 +1

Trace sur un exemple (fil rouge)

$\mathcal{M}$		A	B	C	D
	t_0	0	0	0	0
	t_1	0	1	1	1
BD					
AC					
ABC					

$ag(t_0, t_1) = A$

Trace sur un exemple (fil rouge)

$\mathcal{M}$		A	B	C	D
	t_0	0	0	0	0
	t_1	0	1	1	1
BD	t_2	1	1	2	1
AC					
ABC					

Trace sur un exemple (fil rouge)

$\mathcal{M}$

	A	B	C	D
t_0	0	0	0	0
t_1	0	1	1	1
t_2	1	1	2	1

$ag(t_1, t_2) = BD$

AC

ABC

Trace sur un exemple (fil rouge)

$\mathcal{M}$		A	B	C	D
	t_0	0	0	0	0
	t_1	0	1	1	1
	t_2	1	1	2	1
AC	t_3	1	$\overset{+1}{2}$	2	$\overset{+1}{2}$
ABC					

Trace sur un exemple (fil rouge)

$\mathcal{M}$

	A	B	C	D
t_0	0	0	0	0
t_1	0	1	1	1
t_2	1	1	2	1
t_3	1	2	2	2

$ag(t_2, t_3) = AC$

ABC

Trace sur un exemple (fil rouge)

$\mathcal{M}$		A	B	C	D
	t_0	0	0	0	0
	t_1	0	1	1	1
	t_2	1	1	2	1
	t_3	1	2	2	2
ABC	t_4	1	2	2	3 $+1$

Trace sur un exemple (fil rouge)

$\mathcal{M}$

	A	B	C	D
t_0	0	0	0	0
t_1	0	1	1	1
t_2	1	1	2	1
t_3	1	2	2	2
t_4	1	2	2	3

$ag(t_3, t_4) = ABC$

Trace sur un exemple (fil rouge)

	A	B	C	D
t_0	0	0	0	0
t_1	0	1	1	1
t_2	1	1	2	1
t_3	1	2	2	2
t_4	1	2	2	3

Trace sur un exemple (fil rouge)

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

✓

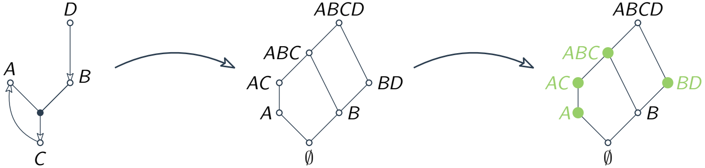
Trace sur un exemple (fil rouge)

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

✓

Théorème : Soit Σ un ensemble de DFs et $\mathcal{M}$ ses infs-irréductibles. L'algorithme $\mathrm{ARMSTRONG}(\mathcal{M})$ calcule une relation d'Armstrong pour Σ.

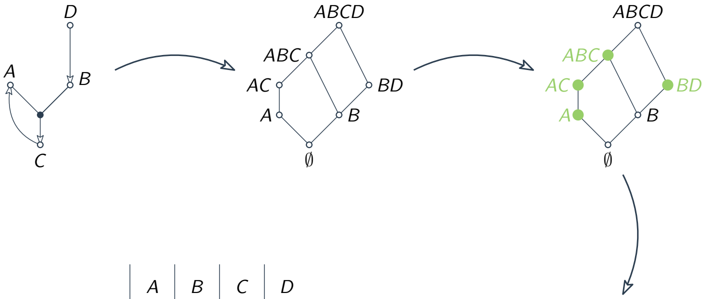
Armstrong wars : Épisode IV



?

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

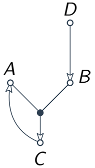
Armstrong wars : Épisode IV

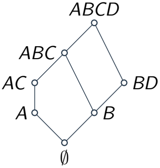


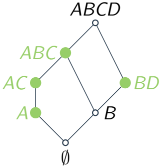
Algo + Thm

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

Armstrong wars : Épisode IV







On peut le faire à l'envers !

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	6

$\boxed{\text{Algo + Thm}}$

Formellement

- On part d'un ensemble de DFs Σ,
- on calcule la collection de ses fermés $\mathcal{F}$,
- on en déduit les infs-irréductibles $\mathcal{M}$,
- en utilisant $\mathcal{M}$, théorème des agrées et l'algorithme $\mathrm{ARMSTRONG}$, on peut calculer une relation d'Armstrong r pour Σ.

La suite : correction d'une relation

- Ce qu'on a vu : comment *construire* une relation exemple qui satisfait un ensemble de DFs et *seulement* celles-ci.
- Ce qu'on va voir : comment *corriger* une relation pour qu'elle satisfasse des DFs, quitte à en satisfaire d'autres.
- Soit r une relation, Σ des DFs et $r \not\models \Sigma$. Principe :
  - r doit contenir un *contre-exemple* à Σ, c.à.d, une paire de tuples t, t' telle que $ag(t, t') \notin \mathcal{F}$.
  - donc, on va *corriger* ce contre-exemple en changeant les valeurs, et répéter cette opération jusqu'à ce qu'aucun contre-exemple ne subsiste.

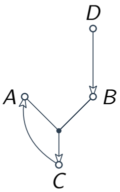
Algorithme CHASE(Σ, r)

Data: Σ un ensemble de DF sur R, r une relation
Result: une relation r' telle que $r' \models \Sigma$

Tant que c'est possible, appliquer la règle suivante :
Soit $X \rightarrow Y \in \Sigma$ et $t, t' \in r$
if $t[X] = t'[X]$ et il existe $A \in Y$ tel que $t[A] \neq t'[A]$ **then**
$t[A] := t'[A] := min(t[A], t'[A])$;

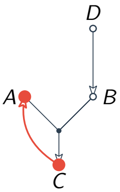
Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	5	c	z	4
t_5	5	c	z	6



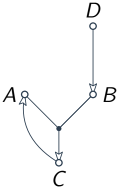
Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	5	c	z	4
t_5	5	c	z	6



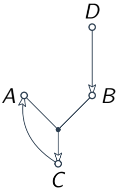
Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	4
t_5	5	c	z	6



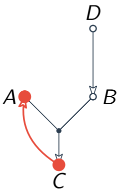
Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	4
t_5	5	c	z	6



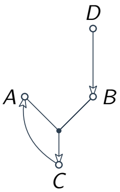
Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	4
t_5	5	c	z	6



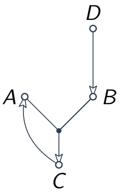
Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	4
t_5	3	c	z	6



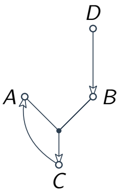
Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_4	3	c	z	4
t_5	3	c	z	6



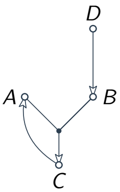
Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_5	3	c	z	6



Exemple

	A	B	C	D
t_0	1	a	x	0
t_1	1	b	y	2
t_2	3	b	z	2
t_3	3	c	z	4
t_5	3	c	z	6



Théorème : L'algorithme $\mathrm{CHASE}(\Sigma, r)$ s'arrête, et la relation résultat r' satisfait Σ.

Que peut-on faire d'autre avec Chase ?

- Tester en partie la « qualité » d'une décomposition de R :
  - ✓ Décomposition sans perte de *données* (jointures)
  - ✗ Décomposition sans perte de *dépendances fonctionnelles* (projection des DFs)
- Calculer la fermeture d'un ensemble d'attributs par rapport à Σ,
- Tester l'implication logique (tableaux).

Algorithme de calcul de fermeture

Algorithme FERMETURE(X, Σ)

Data: $X \subseteq \mathrm{R}$; Σ un ensemble de DF sur R

Result: X^{Σ}, la fermeture de X par rapport à Σ

$X^{\Sigma} := X$;

$fini := false$;

On construit une relation $r = t, t'$ avec $t[X] = t'[X]$ et $t[A] \neq t'[A]$ pour tout $A \notin X$;

$r' = \mathrm{CHASE}(r, \Sigma)$;

if $|r'| = 2$ **then**

$X^{\Sigma} = ag(t, t')$

else

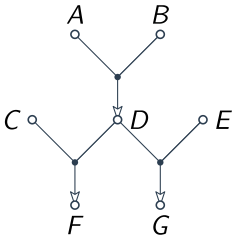
$X^{\Sigma} = \mathrm{R}$

end

return X^{Σ} ;

Trace sur un exemple

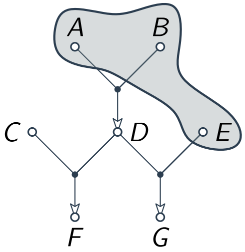
	A	B	C	D	E	F	G
t							
t'							



- Soit $\Sigma = \{AB \rightarrow D, CD \rightarrow F, DE \rightarrow G\}$. On cherche ABE^{Σ}.
-

Trace sur un exemple

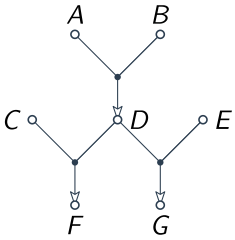
	A	B	C	D	E	F	G
t	1	1	1	1	1	1	1
t'	1	1	0	0	1	0	0



- Soit $\Sigma = \{AB \to D, CD \to F, DE \to G\}$. On cherche ABE^{Σ}.
-

Trace sur un exemple

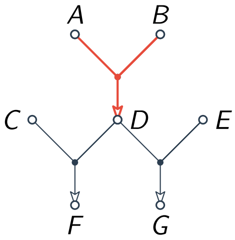
	A	B	C	D	E	F	G
t	1	1	1	1	1	1	1
t'	1	1	0	0	1	0	0



- Soit $\Sigma = \{AB \rightarrow D, CD \rightarrow F, DE \rightarrow G\}$. On cherche ABE^{Σ}.
-

Trace sur un exemple

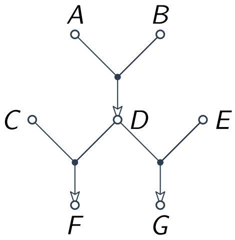
	A	B	C	D	E	F	G
t	1	1	1	1	1	1	1
t'	1	1	0	0	1	0	0



- Soit $\Sigma = \{AB \to D, CD \to F, DE \to G\}$. On cherche ABE^{Σ}.
-

Trace sur un exemple

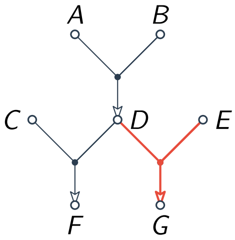
	A	B	C	D	E	F	G
t	1	1	1	0	1	1	1
t'	1	1	0	0	1	0	0



- Soit $\Sigma = \{AB \to D, CD \to F, DE \to G\}$. On cherche ABE^{Σ}.
-

Trace sur un exemple

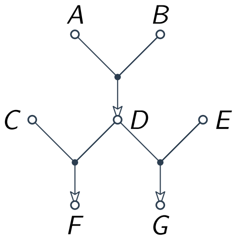
	A	B	C	D	E	F	G
t	1	1	1	0	1	1	1
t'	1	1	0	0	1	0	0



- Soit $\Sigma = \{AB \to D, CD \to F, DE \to G\}$. On cherche ABE^{Σ}.
-

Trace sur un exemple

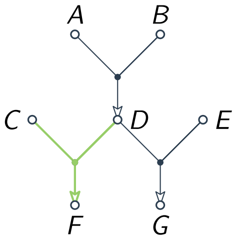
	A	B	C	D	E	F	G
t	1	1	1	0	1	1	0
t'	1	1	0	0	1	0	0



- Soit $\Sigma = \{AB \to D, CD \to F, DE \to G\}$. On cherche ABE^{Σ}.
-

Trace sur un exemple

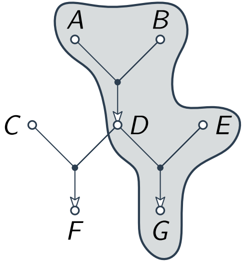
	A	B	C	D	E	F	G
t	1	1	1	0	1	1	0
t'	1	1	0	0	1	0	0



- Soit $\Sigma = \{AB \rightarrow D, CD \rightarrow F, DE \rightarrow G\}$. On cherche ABE^{Σ}.
-

Trace sur un exemple

	A	B	C	D	E	F	G
t	1	1	1	0	1	1	0
t'	1	1	0	0	1	0	0



- Soit $\Sigma = \{AB \to D, CD \to F, DE \to G\}$. On cherche ABE^{Σ}.
- On obtient $ABE^{\Sigma} = ABDEG$.

Pour finir

- Soit $R = \{A, B, C, D, E\}$ et $\Sigma = \{AB \rightarrow C, AC \rightarrow D, BC \rightarrow E\}$:
  1. Construire une relation d'Armstrong r pour Σ,
  2. Corriger la relation r pour que $r \models \Sigma \cup \{A \rightarrow E, B \rightarrow D\}$
  3. **Bonus :** remplir r avec des données « réalistes » (films, livres, types de pantoufles, pizzas, ...)