

# TD: Contraintes & poursuite

Pierre Senellart ([pierre.senellart@ens.fr](mailto:pierre.senellart@ens.fr))

8 mars 2017

Le but de ce TD est d'étudier plus en détail les notions de dépendances et l'algorithme de poursuite.

## Rendu

Le rendu de ce TD doit être fait par mél, avant 23:59 le 8 mars, à [pierre.senellart@ens.fr](mailto:pierre.senellart@ens.fr) en fournissant, en attachement du mél, un fichier PDF contenant une version électronique, des scans ou des photographies *lisibles* de vos réponses au TD. Des pénalités seront comptées en cas de rendu tardif (0,5 point par heure de retard). Les exercices sont par difficulté croissante, il n'est pas nécessaire de finir les quatre exercices de la fiche de TD pour avoir une bonne note.

## 1 Poursuite de FD et MVD

Considérons la relation  $R = \{A, B, C, D, E, F\}$  avec l'ensemble de dépendances :

$$\Sigma = \{\{C\} \rightarrow \{B\}, \{D, E\} \rightarrow \{F\}, \{A\} \twoheadrightarrow \{C, F\}\}.$$

- 1a. Donner une expression de chacune de ces dépendances comme une EGD ou TGD (pas besoin de faire figurer les quantificateurs universels).
- 1b. En utilisant la poursuite, montrer que la décomposition de  $R$  en  $R_1 = \{A, B, C\}$  et  $R_2 = \{A, D, E, F\}$  est sans perte.

## 2 FD et MVD sur des données réelles

`UnicodeData.txt` est un fichier standard décrivant les propriétés de base de l'ensemble des caractères Unicode. Il est téléchargeable depuis <ftp://ftp.unicode.org/Public/UNIDATA/UnicodeData.txt> et son format est décrit sur <http://www.unicode.org/reports/tr44/#UnicodeData.txt>.

- 2a. En étudiant le format du fichier, indiquer l'ensemble des FD et MVD dont vous pensez qu'elles sont naturellement imposées par le format. Inutile de faire figurer une dépendance impliquée par une autre dépendance.
- 2b. Charger le fichier `UnicodeData.txt` dans une table que vous aurez préalablement créée sur votre base PostgreSQL, et vérifier en effectuant des requêtes que les dépendances proposées à la question précédente sont effectivement satisfaites, ou donner des contre-exemples quand ce n'est pas le cas.

## 3 Faible acyclicité

Le *graphe* d'un ensemble  $\Sigma$  de TGD est le graphe orienté dans lequel :

- les sommets sont les positions des relations du schéma relationnel (c'est-à-dire, les couples formés d'un nom de relation et d'un nom d'attribut de cette relation) ;

- il existe une arête *normale* d'une position  $R[A]$  à une autre  $S[B]$  s'il existe un TGD dans  $\Sigma$ , avec  $R$  en partie gauche du TGD, tel que la variable en position  $A$  de  $R$  apparaît en partie droite en position  $B$  d'un atome  $S$ ;
- il existe une arête *spéciale* d'une position  $R[A]$  à une autre  $S[B]$  s'il existe un TGD dans  $\Sigma$ , avec  $R$  en partie gauche du TGD, tel qu'une variable quantifiée existentiellement apparaît en partie droite en position  $B$  d'un atome  $S$ .

Un ensemble  $\Sigma$  de TGD est dit *faiblement acyclique* s'il n'existe aucun cycle *comportant une arête spéciale* dans le graphe de  $\Sigma$ .

- 3a. Montrer que si aucun TGD d'un ensemble  $\Sigma$  n'a de variables existentiellement quantifiée en partie droite, alors  $\Sigma$  est faiblement acyclique.
- 3b. Montrer que la poursuite par un ensemble de TGD faiblement acyclique et un ensemble arbitraire d'EGD termine en un temps fini.