L3I COMBINATOIRE — 5 : RELATIONS ET ALGEBRE RELATIONNELLE

Christian RONSE, ICube, Département d'Informatique de l'Université de Strasbourg

Soient $D_1, \ldots, D_n$ des ensembles ($n \geq 1$) ; une *relation sur* $D_1 \times \cdots \times D_n$ (ou *entre* $D_1, \ldots, D_n$) est une partie de $D_1 \times \cdots \times D_n$. Pour $(d_1, \ldots, d_n) \in R$, on écrira $R(d_1, \ldots, d_n)$ et on dira que $d_1, \ldots, d_n$ sont en relation par R. Le nombre n est appelé l'*arité* de R ; ainsi pour $n = 1, 2, 3$ on dira que R est respectivement *unaire, binaire, ternaire.* En particulier, tout ensemble D est une relation unaire. Pour $i = 1, \ldots, n$ et $(d_1, \ldots, d_n) \in R$, d_i est le i-ème champ de $(d_1, \ldots, d_n)$. L'ensemble $D_1 \times \cdots \times D_n$ est appelé le *domaine de* R, tandis que pour $i = 1, \ldots, n$, D_i est le *domaine de champ* i *de* R.

Les relations constituent le modèle mathématique des *bases de données relationnelles.* Ici chaque champ peut recevoir un nom, de même que la relation. Par exemple une relation 5-aire "Bacheliers" aura comme champs "Nom", "Prénom", "Numéro INE", "Série", "Mention", elle sera un ensemble de 5-uples, chacun correspondant à un bachelier, et les différents champs d'un tel 5-uple (nom, prénom, etc.) seront des chaînes de caractères ; donc chaque domaine de champ D_i sera une partie de l'ensemble de toutes les chaînes de caractères.

Un intérêt des relations est qu'il est possible de faire de nombreuses opérations sur celles-ci. La structure formée par les opérations sur les relations est appelée *algèbre relationnelle.* Un système de gestion de bases de données relationnelles doit pouvoir effectuer de telles opérations (par exemple au moyen de requêtes en SQL).

La *négation* d'une relation R est la relation $\neg R$ définie par

$$\neg R(d_1, \ldots, d_n) \text{ vrai} \quad \Leftrightarrow \quad R(d_1, \ldots, d_n) \text{ faux} \quad .$$

La *conjonction* de deux relations R et S de même arité est la relation $R \wedge S$ définie par

$$(R \wedge S)(d_1, \ldots, d_n) \;\Leftrightarrow\; \big[R(d_1, \ldots, d_n) \text{ et } S(d_1, \ldots, d_n)\big] \;,$$

tandis que leur *disjonction* $R \vee S$ se définit par

$$(R \vee S)(d_1, \ldots, d_n) \;\Leftrightarrow\; \big[R(d_1, \ldots, d_n) \text{ ou } S(d_1, \ldots, d_n)\big] \;.$$

Du point de vue ensembliste, $\neg R = R^c = (D_1 \times \cdots \times D_n) \setminus R$, $R \wedge S = R \cap S$ et $R \vee S = R \cup S$. La conjonction, la disjonction, la négation et les combinaisons logiques de celles-ci constituent les *opérations booléennes.*

Le *produit cartésien* $R \times S$ de deux relations quelconques se définit au sens ensembliste, donc pour $R \subseteq D_1 \times \cdots \times D_n$ et $S \subseteq D'_1 \times \cdots \times D'_m$, on a

$$R \times S = \big\{(d_1, \ldots, d_n, d'_1, \ldots, d'_m) \mid (d_1, \ldots, d_n) \in R,\ (d'_1, \ldots, d'_m) \in S\big\} \;,$$

ainsi $R \times S \subseteq D_1 \times \cdots \times D_n \times D'_1 \times \cdots \times D'_m$.

Etant donnés une relation n-aire R et k champs $i_1, \ldots, i_k$ ($k \leq n$), la *projection* de R sur ces champs est la relation k-aire

$$\pi_{i_1,\ldots,i_k}(R) = \big\{(d_{i_1}, \ldots, d_{i_k}) \mid (d_1, \ldots, d_n) \in R\big\} \;.$$

Par exemple pour la relation "Bacheliers" donnée plus haut, la projection sur les 2 premiers champs "Nom" et "Prénom" donnera la liste des couples (nom, prénom) de tous les bacheliers.

Soit F un prédicat logique, dont les variables sont les champs, formé par combinaison logique (avec les connecteurs "et", "ou" et "non") de comparaisons, chacune liant par une des relations

$=$, $\neq$, $<$, $>$, $\leq$ ou $\geq$, deux champs ou un champs et une constante. Par exemple pour 6 champs numériques,

$$F(d_1, \ldots, d_6) \equiv \Big((d_1 \geq 3) \text{ et } \big[(d_2 = d_3) \text{ ou } (d_2 \neq 5)\big] \Big) \ .$$

La *sélection* de la relation R par F est la relation

$$\sigma_F(R) = \big\{(d_1, \ldots, d_n) \in R \mid F(d_1, \ldots, d_n)\big\} \ .$$

Par exemple pour la relation "Bacheliers", on peut sélectionner tous les 5-uples où le 5ème champ "Mention" a la valeur "AB" (tous les bacheliers ayant la mention assez-bien). Notons que F peut être vue comme une relation de même domaine que R, et dans ce cas $\sigma_F(R) = R \wedge F$.

Soient R et S deux relations ayant certains champs de même domaine, c.-à-d. $R \subseteq D_1 \times \cdots \times D_n \times D'_1 \times \cdots \times D'_p$ et $S \subseteq D'_1 \times \cdots \times D'_p \times D''_1 \times \cdots \times D''_q$. La *jointure* de R et S est une variante du produit cartésien, où les champs communs ne sont pas dupliqués :

$$R \bowtie S = \big\{(d_1, \ldots, d_n, d'_1, \ldots, d'_p, d''_1, \ldots, d''_q) \mid \\ (d_1, \ldots, d_n, d'_1, \ldots, d'_p) \in R,\ (d'_1, \ldots, d'_p, d''_1, \ldots, d''_q) \in S\big\} \ .$$

Considérons par exemple trois ensembles E d'étudiants, S de sandwichs et C de cafés (en fait, formellement ce sont des ensembles de noms, c.-à-d. de chaînes de caractères) ; définissons les relations $A \subseteq E \times S$ et $D \subseteq S \times C$, avec $(e, s) \in A$ ssi l'étudiant e aime le sandwich s, et $(s, c) \in D$ ssi le sandwich s est disponible dans le café c. La jointure $A \bowtie D$ comprend tous les triplets (e, s, c) tels que e aime s qui est disponible dans d.

Notons que les opérations de produit cartésien et de jointure sont toutes deux associatives.

Soient R et S deux relations telles que le domaine de S soit le produit de certains domaines de champs de R, c.-à-d. $S \subseteq D'_1 \times \cdots \times D'_m$ et $R \subseteq D_1 \times \cdots \times D_n \times D'_1 \times \cdots \times D'_m$. Alors le *quotient* de R par S est la relation $R \div S$ dont le domaine est le produit des domaines de champs restants de R, définie par

$$R \div S = \big\{(d_1, \ldots, d_n) \in D_1 \times \cdots \times D_n \mid \forall (d'_1, \ldots, d'_m) \in S,\ (d_1, \ldots, d_n, d'_1, \ldots, d'_m) \in R\big\} \ .$$

Dans l'exemple précédent, A/S est l'ensemble des étudiants qui aiment tous les sandwichs, et D/C est celui des sandwichs disponibles dans tous les cafés.