932

Fondements des bases de données relationnelles

I Formalisme :

Def 1 : On se donne des ensembles dénombrables distinct de variables $\boxed{VAR} = \{x, y, z, \dots\}$, d'attribut $\boxed{att}$, de domaine $\boxed{dom}$ et de noms de schéma de relation $\boxed{Rel\text{-}name}$

on considère la fonction :
sort : relname $\rightarrow \mathcal{P}_f(att)$
qui à $R \in$ relname associe un nombre fini d'attributs.

Def 2 : on dira que, pour $R \in$ relname, $|sort(R)|$ est l'arité de R

Def 3 (schéma de relation) Un élément $R \in$ relname est un schéma de relation.
on notera $R[U]$ où $U = sort(R)$

Def 4 (schéma de base de données) c'est un nombre fini de schémas de relation
on notera $\mathcal{R} = \{R_1[U_1], \dots, R_n[U_n]\}$

Ex 5 : $\mathcal{R}$ {Films[Titre, Réal, Acteur], Séances[Cinéma, Horaires, Titres]} est un schéma de base de données.

Def 6 (tuple) Un tuple sur un schéma de relation $R[U]$ est une fonction $u : U \rightarrow dom$
on notera $\langle U_1 : u(U_1), \dots, U_n : u(U_n) \rangle$

Ex 7 : ⟨Titre : Didier, Réal : A. Chabat, Acteur : A. Chabat⟩ est un tuple de FILMS
on pourra écrire ⟨Didier, A. Chabat, A. Chabat⟩

Def 8 : Une relation I sur un schéma de relation $R[U]$ est un nombre fini de tuple pour $R[U]$.

Ex 9 : FILMS est une relation.

Def 10 : Une base de donnée $\mathcal{I}$ sur $\mathcal{R}$ est un ensemble de relation, où une relation I est sur un schéma $R[U] \in \mathcal{R}$.

Ex 11 : voir annexe 1

II Requêtes conjonctives.

1) Règles conjonctives

Def 12 : Un tuple libre d'un schéma de relation $R[U]$ est un tuple pouvant avoir des variables
$u : U \rightarrow dom \cup VAR$.

Ex 13 : $\langle x$, A. Chabat, $y \rangle$ est un tuple libre.

Def 14 : Une règle conjonctive sur $\mathcal{R}$ est une expression de la forme $ans(u) \leftarrow R_1(u_1), \dots, R_n(u_n)$
où $u, u_1, \dots, u_n$ sont des tuples libres, $R_i \in \mathcal{R}$, $ans \notin \mathcal{R}$

Ex 15 : $ans(x) \leftarrow$ Films(x, A. Chabat, y) est une règle conjonctive.

Def 16 : Une valuation ν sur VAR est une fonction $\nu : VAR \rightarrow dom$
prolongée des tuples libres aux tuples ($\nu|_{dom} = id_{dom}$)

Def 17 : La sémantique d'une règle conjonctive q pour $\mathcal{I}$ sur $\mathcal{R}$ est donnée par
$$q(\mathcal{I}) = \{\nu(u) \mid \nu \text{ valuation sur } VAR(q) \text{ et } \nu(u_i) \in \mathcal{I}(R_i)\ \forall i \in \{1, \dots, n\}\}$$

Ex 18 : L'exemple 15 a pour sémantique
⟨Didier, A. Chabat, A. Chabat⟩
⟨Astérix, A. Chabat, E. Baer⟩
Cette règle répond à la question
Q_1 Quels films ont été réalisés par A. Chabat ?

Def 19 : Le domaine actif de $\mathcal{I}$, $adom(\mathcal{I})$ est l'ensemble des constantes présentes dans $\mathcal{I}$. Celui de q, $adom(q)$ est l'ensemble des constantes présentes dans q.

Prop 20 : On a $adom(q(\mathcal{I})) \subseteq adom(q) \cup adom(\mathcal{I})$

Ex 21 : $\boxed{Q_2}$ Dans quels cinémas peut-on voir un film d'A. Chabat ?

$ans(x) \leftarrow$ Séances(x, y, z), Films$(z, A.Chabat, t)$

2) Calcul conjonctif

Def 22 : Soit R, une formule du calcul conjonctif (CC) est de la forme

(i) $R(u)$ pour $R[U] \in R$ et u un tuple libre

(ii) $\varphi \wedge \psi$ où φ, ψ sont des formules du CC

(iii) $\exists x\, \varphi$ où $x \in VAR$ et φ une formule du CC

Def 23 : On définit l'ensemble des variables libre $FV(\varphi)$ comme en logique du premier ordre.

Def 24 : Une requête du CC sur R est une expression de la forme $\{e_1, \dots, e_m \mid \varphi\}$ où $\langle e_1, \dots, e_m\rangle$ tuple libre, φ formule du CC et $FV(\varphi) = \{e_1, \dots, e_m\}$

Def 25 : Pour R, φ du CC sur R, ν valuation sur $FV(\varphi)$, $\mathcal{I}$ satisfait φ sous ν, noté $\mathcal{I} \models \varphi[\nu]$, si

(i) $\varphi = R(u)$, $\nu(u) \in \mathcal{I}(R)$ ou

(ii) $\varphi = \psi \wedge \phi$, $\mathcal{I} \models \psi[\nu|_{FV(\psi)}]$ et $\mathcal{I} \models \phi[\nu|_{FV(\phi)}]$ ou

(iii) $\varphi = \exists x \psi$, $\exists c \in dom$ $\mathcal{I} \models \psi[\nu \cup (x \mapsto c)]$

Def 26 : La sémantique d'une requête q du CC est donnée par
$q(\mathcal{I}) = \{\nu(\langle e_1, \dots, e_n\rangle) \mid \mathcal{I} \models \varphi[\nu]$ et ν valuation sur $FV(\varphi)\}$

Ex 27 : on exprime $\boxed{Q_1}$ par $\{x, \exists y$ Films$(x, A.Chabat, y)\}$

et $\boxed{Q_2}$ par $\{x \mid \exists y, z, t$ Séances$(x, y, z) \wedge$ Films$(z, A.Chabat, t)\}$

Thm 28 : Le calcul conjonctif et les règles conjonctives sont équivalents.

III L'algèbre SPC

Def 29 : On définit des opérations sur $\mathcal{I} \in \mathcal{I}$ qui formeront l'algèbre SPC.

* sélection : $\sigma_{j=a}(\mathcal{I}) = \{t \in \mathcal{I}, t(j) = a\}$.

avec j attribut, $a \in dom$, t un tuple.

$\sigma_{j=k}(\mathcal{I}) = \{t \in \mathcal{I}, t(j) = t(k)\}$

avec j, k attributs, t un tuple.

* projection : $\Pi_{j_1, \dots, j_n}(\mathcal{I}) = \{\langle t(j_1), \dots, t(j_n)\rangle, t \in \mathcal{I}\}$.

avec $j_1, \dots, j_n$ des attributs de $\mathcal{I}$.

* produit cartésien : $\mathcal{I} \times \mathcal{J} = \{\langle t(1), \dots, t(n), s(1), \dots, s(m)\rangle, t \in \mathcal{I}, s \in \mathcal{J}\}$

avec $n =$ arité$(\mathcal{I})$, $m =$ arité$(\mathcal{J})$

Def 30 : La sémantique liée à ces opérations est naturelle puisque l'on effectue les opérations sur les relations.

Ex 31 : $\boxed{Q_1}$ est exprimée par $\Pi_{titre}\, \sigma_{Réal = A.Chabat}(Films)$

$\boxed{Q_2}$ est exprimée par $\Pi_{Cinéma}\, \sigma_{Réal = A.Chabat}\, \sigma_{titre_{Film} = Titre_{Séances}}(Films \times Séances)$

Prop 32 : On peut effectuer une jointure naturelle (union de deux relations où l'on fait correspondre les éléments communs) avec les trois opérations

Ex 33 : Dans l'exemple on a fait une jointure naturelle

Prop 34 : On peut exprimer aussi l'intersection grâce aux 3 opérations de SPC.

Rmq 35 : Il existe des requêtes insatisfiables.

Ex 36 : $\boxed{Q_3}$ $\Pi_A(\sigma_{A=0}(\sigma_{A=1}(table)))$ est insatisfiable.

Rmq 37 : C'est une grande différence avec les requêtes conjonctives qui sont toujours satisfiables.

Thm 38 : Il y a équivalence entre les requêtes par règles conjonctives, le calcul conjonctif et les requêtes satisfiables de l'algèbre SPC.

Limite 39 : On ne peut pas encore répondre aux questions : Quels sont les films réalisés par A. Chabat ou par Q. Dupieux ? $\boxed{Q_4}$

$\boxed{Q_5}$ Quels sont les films de Q. Dupieux dans lesquels A. Chabat ne joue pas ?

IV L'algèbre relationnelle et le calcul relationnel

1) Ajout de l'union

Def 40 : On autorise dorénavant la sélection avec $\sigma_{j=a \vee j=b}$ (avec le "ou" logique) muni de sa sémantique naturelle → Algèbre SPCU

Ex 41 : $\boxed{Q_4}$ $\pi_{titre}\,\sigma_{réal=A.Chabat \text{ ou } réal=Q.Dupieux}(Films)$

Def 42 : On ajoute le $\vee$ aux formules du CC

Problème 43 : On peut avoir des requêtes infinies

Ex 44 : $\{x,y \mid R(x) \vee R(y)\}$ si $R(x_0)=1$ est donc infini $\forall y \in dom\ R(x) \vee R(y) = 1$

Def 45 : On dit qu'une requête est sûre si $|q(I)| < +\infty$ $\forall I \in R$

Problème 46 : savoir si une requête est sûre est indécidable.

2) Ajout de la négation

Def 47 : On ajoute la différence ensembliste ce qui donne l'algèbre SPCUD.

Ex 48 : $\boxed{Q_5}$ $\pi_{titre}\left(\sigma_{réal=Q.Dupieux}(Films) \setminus \sigma_{acteur=A.Chabat}(Films)\right)$

Def 49 : On ajoute le $\neg$ aux formules pour obtenir le calcul relationnel. et $\forall$

Problème 50 : On a de nouveau des requêtes infinies

Ex 51 : $\{\langle\rangle \quad , \neg R(x_0)\}$ si le domaine est infini c'est une requête infinie.

Solution 52 : Pour une relation I sur R, on peut restreindre le domaine au domaine actif, ainsi pour une requête et une relation, la requête sera finie.

Thm 53 : Restreint au domaine actif, l'algèbre relationnelle SPCUD et le calcul relationnel (FO) sont équivalents.

Thm 54 : Le problème $\boxed{\text{REL-SAT}}$
- entrée : une requête du calcul relationnel sur R
- sortie : oui s'il existe $I \in R$ tel $q(I) \neq \emptyset$, non sinon

est indécidable. — DEV 1

Rmq 55 : Savoir si deux requêtes q, q' sont équivalentes est aussi indécidable.

V Optimisation

1) Stockage

Def 56 : (B-arbres) Un B-arbre est un arbre tel que il existe t degré minimal de T
1) chaque nœud x contient $x.n$ clés, notées $x.clé_1, \dots, x.clé_n$ stockées par ordre croissant, un booléen $x.feuille$, $n+1$ pointeurs vers ses fils $x.c_1 \dots x.c_{n+1}$
2) les clés et valeurs k_i des fils sont telles que $k_0 \leq x.clé_1 \leq k_1 \leq x.clé_2 \dots \leq x.clé_n \leq k_n$
3) toutes les feuilles ont même profondeur, h.
4) $\forall x \quad t-1 \leq x.n \leq 2t-1$

Thm 57 : Soit T un B-arbre à n éléments de degré minimal $t \geq 2$. Alors sa hauteur h vérifie $h \leq \log_t\left(\frac{n+1}{2}\right)$

Algorithme de recherche et exemple d'insertion. — DEV 2

Annexe : $R = \{FILMS, Séance\}$

FILMS

Titre	Réal	Acteur
Didier	A. Chabat	A. Chabat
Au poste	Q. Dupieux	G. Ludig
Réalité	Q. Dupieux	A. Chabat
Astérix	A. Chabat	E. Baer

Séances

Cinéma	Horaire	Titre
UGC	12h00	Au poste
Gaumont	20h50	Didier
Pathé	13h30	Réalité
Pathé	16h30	Didier

B-arbre :

$t = 2$