# *Calculul relațional*

*prof. dr. ing. Mircea Petrescu*

Calculul relațional – două variante:
- calcul relațional pe tupluri;
- calcul relațional pe domenii.

## *Calcul relațional pe tupluri*

Expresiile în calculul relațional sunt de forma { t | Ψ(t) }, unde t este o variabilă tip tuplu (sau „variabilă-tuplu"). Prin definiție, o variabilă-tuplu este o variabilă care reprezintă un tuplu de o anumită lungime fixă. Simbolul Ψ reprezintă o formulă, construită din atomi şi din o colecție de operatori. Introducem trei tipuri de atomi:

a) R(s), unde R – nume de relație, iar s – variabilă-tuplu. Semnificația atomului R(s) este echivalent cu asertiunea „s este un tuplu în R";
b) s[i] Θ u[j], unde s şi u sunt variabile-tuplu, iar Θ operator aritmetic de comparație (>, =, etc.). Acest atom este echivalent cu asertiunea „a $i^a$ componentă din s stă în relația Θ cu a $j^a$ componentă din u". Exemplu: s[10] > u[7].
c) s[i] Θ a şi a Θ s[i], unde a este o constantă; are acelaşi sens ca în b). Exemplu: s[5]="10".

*Operatori*. Pentru a introduce operatorii, definim mai întâi noțiunile de *variabile-tuplu libere şi legate*. Dacă o variabilă-tuplu a fost introdusă prin unul din cuantificatorii ∀ (pentru toate, pentru toți) sau ∃ (există), o apariție (ipostază) a acestei variabile într-o formulă este „legată". Variabilele care sunt introduse prin ∀ şi ∃ sunt „libere".

Formulele şi aparițiile „libere" şi „legate" ale variabilelor-tuplu în aceste formule sunt definite recursiv:

1. Fiecare atom este o formulă. Toate aparițiile de variabile-tuplu în atom sunt variabile libere în această formulă.
2. Dacă $\Psi_1$ şi $\Psi_2$ sunt formule, atunci $\Psi_1 \wedge \Psi_2$, $\Psi_1 \vee \Psi_2$, $\neg\Psi_1$ sunt formule. Asertiunile care corespund acestor formule sunt: $\Psi_1$ şi $\Psi_2$ sunt ambele adevărate; $\Psi_1$ sau $\Psi_2$ sau ambele sunt adevărate; $\Psi_1$ nu este adevărată. Aparițiile variabilelor-tuplu în formulele de mai sus pot fi legate sau libere, după cum sunt legate sau libere în $\Psi_1$ sau $\Psi_2$. O variabilă tuplu s poate fi liberă în $\Psi_1$ şi legată în $\Psi_2$, etc.
3. Dacă Ψ este o formulă, atunci (∃s)(Ψ) este o formulă. Aparițiile variabilei-tuplu s, care sunt libere în Ψ, vor fi legate de (∃s) în (∃s)(Ψ). Alte apariții de variabile-tuplu în Ψ, inclusiv aparițiile posibile ale variabilei s, care sunt legate de Ψ (sau „în Ψ"), vor fi legate sau libere în (∃s)(Ψ), după cum au fost în Ψ. Asertiunea echivalentă cu (∃s)(Ψ) – există o valoare a variabilei s astfel încât atunci când aceasta este substituită tuturor aparițiilor libere ale lui s în Ψ, Ψ devine adevărată. Exemplu: (∃s)(R(s)) – echivalentă cu „R nu este vidă", deci ∃ un tuplu s în R.
4. Dacă Ψ este o formulă, atunci (∀s)(Ψ) este o formulă. Aparițiile libere ale lui s în Ψ sunt legate de (∀s) în (∀s)(Ψ). Altele, ca la 3). Asertiunea echivalentă: dacă pentru aparițiile libere ale variabilei s în Ψ se substituie orice valoare de aritate adevărată, formula Ψ devine adevărată.
5. Ordinea de precedență: operatorii aritmetici de comparație (cea mai înaltă prioritate), apoi ∃, ∀, ¬, ∧, ∨.
6. Nimic altceva (nici o altă construcție) nu este o formulă.

O *expresie* în calculul relațional pe tupluri are forma $\{ t \mid \Psi(t) \}$, unde t este singura variabilă-tuplu liberă în Ψ. Exemplu:

*Reuniunea* relațiilor R şi S este reprezentată de expresia $\{ t \mid R(t) \vee S(t) \}$. Reuniunea are sens numai dacă R şi S au aceeaşi aritate; în mod analog, formula $\{ t \mid R(t) \wedge S(t) \}$ are sens numai dacă R şi S au aceeaşi aritate, deoarece variabila-tuplu t are o lungime fixă (prin ipoteză).

*Diferența* a două relații R şi S este dată de expresia: $\{ t \mid R(t) \wedge \neg S(t) \}$.

*Produsul cartezian* al relațiilor R şi S, de arități r şi s, este reprezentat de expresia: $\{ t^{(r+s)} \mid (\exists u^{(r)}) (\exists v^{(r)}) (R(u) \wedge S(v)), \wedge t[1] = u[1] \wedge \ldots \wedge t[r] = u[r], \wedge t[r+1] = v[1] \wedge \ldots \wedge t[r+s] = v[s] \}$.

*Proiecția* $\prod_{i1, i2, \ldots, ik}$ este exprimată de $\{ t^{(k)} \mid (\exists u) (R(u) \wedge t[1] = u[i_1] \wedge \ldots t[k] = u[i_k]) \}$.

Selecția $\sigma_F(R)$ este exprimată prin expresia $\{ t \mid R(t) \wedge F` \}$, unde F` este formula F în care însă fiecare operand i, reprezentând a $i^a$ componentă, a fost înlocuit cu t[i].

Ca un ultim exemplu – dacă R este o relație de aritate 2, atunci $\{ t^{(2)} \mid (\exists u) (R(t) \wedge R(u) \wedge (t[1] \neq u[1] \vee t[2] \neq u[2])) \}$ este o expresie care reprezintă pe R dacă R are cel puțin doi membri, respectiv reprezintă o *relație vidă* dacă R este vidă sau are numai un membru.

## *Expresii sigure. Domenii de siguranță*

În principiu, în calculul relațional putem defini relații intime – ca $\{ t \mid \neg R(t) \}$ – care reprezintă toate tuplurile posibile care nu aparțin lui R, dar au lungimea pe care o asociem cu t; desigur, această lungime trebuie să fie, de asemenea, aritatea lui R, astfel încât expresia să aibă sens.

Apare însă întrebarea: cum putem, de exemplu, imprima „toate tuplurile posibile”, pe ce domeniu de valori? Evident, de aici rezultă că expresiile de acest tip trebuie evitate (de fapt, eliminate), considerând numai acele expresii $\{ t \mid \Psi(t) \}$ care sunt „sigure”.

Prin definiție, o expresie poate fi denumită „sigură”, dacă se poate demonstra că fiecare componentă a oricărui tuplu t care satisface formula Ψ trebuie să fie un membru al mulțimii (domeniului) DOM(Ψ). La rândul său, DOM(Ψ) se defineşte ca mulțimea simbolurilor care:

a) fie apar explicit în Ψ,
b) fie sunt componente ale unui tuplu oarecare al unei relații R oarecare, menționată în Ψ.

Remarcăm că DOM(Ψ) nu se determină prin simpla inspectare a forumlei Ψ, ci este o funcție de relațiile care se substituie efectiv variabilelor din Ψ. DOM(Ψ) este întotdeauna finit, deoarece presupunem că toate relațiile sunt finite. De exemplu, daca R este o relație binară şi dacă $\Psi[t] = s[t] \vee t[1] = a \vee R(t)$, atunci DOM(Ψ) este o relație unară dată de formula algebrică relațională: $\{ a \} \cup \prod_1(R) \cup \prod_2(R)$.

Prin definiție, o expresie $\{ t \mid \Psi(t) \}$ în calculul relațional este sigură, dacă sunt îndeplinite următoarele condiții:

1. Oridecâte ori t satisface Ψ, fiecare componentă din t este un membru al mulțimii DOM(Ψ);
2. Pentru fiecace subexpresie a lui Ψ, de forma $(\exists u) (\omega(u))$, dacă ω este satisfăcut de u pentru oricare valori ale celorlalte variabile libere din ω, atunci fiecare componentă din u face parte din DOM(ω);

3. Pentru fiecare subexpresie a lui $\Psi$, de forma $(\forall u)$ $(\omega(u))$, dacă vreo componentă a lui $u \notin DOM(\omega)$, atunci u satisface $\omega$ pentru toate valorile celorlalte variabile libere din $\omega$.

Această ultimă regulă poate părea neintuitivă. Observăm însă că formula $(\forall u)$ $(\omega(u))$ este echivalentă, din punct de vedere logic, cu formula $\neg(\exists u)$ $(\neg\omega(u))$. Ultima formulă este nesigură dacă şi numai dacă există un $u_0$ pentru care $\neg\omega(u_0)$ este adevărată şi $u_0$ nu face parte din domeniul formulei $\neg\omega$. Întrucât domeniile $\omega$ şi $\neg\omega$ sunt aceleaşi, regula 3 ne spune, de fapt că formula $(\forall u)$ $(\omega(u))$ este sigură exact atunci când formula $\neg(\exists u)$ $(\neg\omega(u))$ este, la rândul ei, sigură.

Vom arăta că expresiile sigure din calculul relaţional cu tupluri sunt echivalente cu algebra relaţională. Interesant este să determinăm acele expresii din calculul relaţional cu tupluri care sunt nesigure – acest exerciţiu este important, atât practic, cât şi conceptual.

Exemplu: fie $\Psi(t)$ o formulă, astfel încât nicio formulă de forma $(\exists u)$ $(\omega(u))$ sau $(\forall u)$ $(\omega(u))$ nu încalcă regulile de siguranţă. Atunci, orice expresie de forma $\{ t \mid R(t) \wedge \Psi(t) \}$ este sigură, deoarece orice tuplu t care satisface $R(t) \wedge \Psi(t)$ este în R, de unde rezultă că fiecare din componentele sale este în $DOM(R(t) \wedge \Psi(t))$. De pildă, formula pentru diferenţa a două mulţimi: $\{ t \mid R(t) \wedge \neg S(t) \}$, este de forma menţionată, cu $\Psi(t) = \neg S(t)$. Formula pentru selecţie este, de asemenea, de aceeaşi formă, cu $\Psi(t) = F`$.

Generalizare a celor de mai sus: observăm că orice formulă $\{ t \mid R_1(t) \vee R_2(t) \vee \ldots \vee R_k(t) \wedge \Psi(t) \}$ este, de asemenea, sigură; t[i] trebuie să fie un simbol care apare în a $i^a$ componentă a unui tuplu oarecare a unei relaţii oarecare $R_j$. Remarcăm, de asemenea, că formula dată pentru reuniune într-un exemplu precedent, este de această formă, dar cu $\Psi$ lipsind; cu alte cuvinte, putem lua pe $\Psi$ ca fiind o formulă întotdeauna adevărată, ca t[1] = t[1].

O altă expresie sigură: $\{ t^{(m)} \mid (\exists u_1) (\exists u_2) \ldots (\exists u_k) (R_1(u_1) \wedge R_2(u_2) \wedge \ldots \wedge R_k(u_k) \wedge t[1] = u_{i1}[j_1] \wedge t[2] = u_{i2}[j_2] \wedge \ldots \wedge t[m] = u_{im}[j_m] \wedge \Psi(t, u_1, u_2, \ldots, u_k)) \}$. Aici, componenta t[l] este constânsă la a fi un simbol care apare în cea de-a $j_l^a$ componentă a unui tuplu din $R_{jl}$. De forma dată mai sus sunt formule care corespund produsului cartezian şi proiecţiei.

### *Reducerea algebrei relaţionale la calculul relaţional cu tupluri*

Potrivit teoremei care urmează, mulţimea funcţiilor de relaţiii, exprimabile în algebra relaţională, este exact aceeaşi cu mulţimea funcţiilor exprimabile prin formule sigure în calculul relaţional cu tupluri.

*Teoremă.* Dacă E este o expresie în algebra relaţională, atunci există o expresie sigură în calculul relaţional cu tupluri, echivalentă cu E. Notă: teorema se demonstrează pentru cinci cazuri: $E = E_1 \cup E_2$, $E = E_1 - E_2$, $E = E_1 \times E_2$, $E = \prod_{i1,i2,\ldots,ik}(E_1)$, $E = \sigma_F(R)$.

*Demonstraţia* se face prin inducţie, pe numărul de apariţii ale operatorilor din E.

*Baza.* Niciun operator (zero operatori!). Atunci E este sau o relaţie constantă $\{ t_1, t_2, \ldots, t_n \}$, sau o variabilă tip relaţie (variabilă-relaţie) R. În ultimul caz E este echivalentă cu $\{ t \mid R(t) \}$, care este o expresie sigură – aşa cum am arătat în ultimul exemplu.

În primul caz, E este echivalentă cu: $\{ t \mid t = t_1 \vee t_2 \vee \ldots \vee t_n \}$, unde $t = t_i$ este o notaţie prescurtată pentru $t[1] = t_i[1] \wedge \ldots \wedge t_i[k]$; aici, k este aritatea tuplului t. Se poate vedea uşor că t[i] este unul din elementele mulţimii finite de simboluri, care apare explicit ca cea de a $i^a$ componentă a unui tuplu constant oarecare $t_j$.

*Inducția.* Să admitem că E are cel puțin un operator, şi că teorema este adevărată pentru expresii cu mai puține apariții de operatori decât are E.

*Cazul 1.* $E = E_1 \cup E_2$. Atunci $E_1$ şi $E_2$ au ambele mai puține apariții de operatori decât E şi, prin ipoteza inductivă, putem găsi expresii sigure, în calculul relațional $\{ t \mid \Psi_1(t) \}$ şi $\{ t \mid \Psi_2(t) \}$, echivalente cu $E_1$ şi $E_2$. Atunci E este echivalentă cu $\{ t \mid \Psi_1(t) \vee \Psi_2(t) \}$.

Acum, dacă t satisface $\Psi_1(t) \vee \Psi_2(t)$, atunci fiecare componentă a lui t face parte din $DOM(\Psi_1)$ sau din $DOM(\Psi_2)$. Deoarece $DOM(\Psi_1(t) \vee \Psi_2(t)) = DOM(\Psi_1) \cup DOM(\Psi_2)$, E este echivalentă cu o expresie sigură. Cu alte cuvinte, formula completă $\Psi(t) = \Psi_1(t) \vee \Psi_2(t)$ este adevărată numai atunci când $t \in DOM(\Psi)$; de asemenea, orice sub-formulă $(\exists u)\ (\omega(u))$ sau $(\forall u)\ (\omega(u))$ din $\Psi$ trebuie să $\in$ din $\Psi_1$ sau $\Psi_2$, aşa că ipoteza inductivă asigură faptul că aceste subformule nu încalcă regulile de siguranță.

Cazurile 2, 3, 4, 5 ($E = E_1 - E_2$, $E = E_1 \times E_2$, $E = \prod_{i1,i2,\dots,ik}(E_1)$, $E = \sigma_F(R)$) – demonstații similare.

*Exemplu.* Dacă R şi S sunt relații binare, compoziția lor, în sensul obişnuit al teoriei mulțimilor, este exprimată de expresia din algebra relațională $\prod_{1.4}(\sigma_{2=3}(R\times S))$. Folosind algoritmul teoremei de mai sus, construim pentru $R\times S$ expresia de calcul relațional: $\{ t \mid (\exists u)\ (\exists v)\ (R(u) \wedge S(v) \wedge t[1] = u[1] \wedge t[2] = u[2] \wedge t[3] = v[1] \wedge t[4] = v[2]) \}$.

Pentru $\sigma_{2=3}(R\times S)$ adăugăm la formula de mai sus termenul $\wedge t[2] = t[3]$. Atunci, pentru $\prod_{1.4}(\sigma_{2=3}(R\times S))$ obținem expresia: $\{ w \mid (\exists t)\ (\exists u)\ (\exists v)\ (R(u) \wedge S(v) \wedge t[1] = u[1] \wedge t[2] = u[2] \wedge t[3] = v[1] \wedge t[4] = v[2] \wedge t[2] = t[3] \wedge w[1] = t[1] \wedge w[2] = t[4]) \}$.

Observăm că expresia de mai sus nu are forma cea mai „succintă" (minimală!). Dacă vom elimina fiecare din componentele lui t prin componentele adevărate ale lui u şi v, t poate fi eliminat. Dacă realizăm această operație, obținem expresia: $\{ w \mid (\exists u)\ (\exists v)\ (R(u) \wedge S(v) \wedge u[2] = v[1] \wedge w[1] = u[1] \wedge w[2] = v[2]) \}$.

Ultima expresie poate fi recunoscută ca definiția obişnuită a compoziției, din teoria mulțimilor, exprimată în limbajul calculului relațional pe tupluri.

### *Calculul relațional pe domenii*

Se construieşte cu ajutorul aceloraşi operatori folosiți în calculul relațional cu tupluri. Principalele diferențe sunt următoarele:

1. în calculul pe ($\omega$) domenii, nu se folosesc variabile-tuplu. În schimb, componentele tuplurilor sunt reprezentate de variabile tip domeniu (variabile-domeniu).
2. Un atom poate avea una din următoarele două forme:
   a) $R(x_1x_2...x_k)$, unde R este o relație de aritate k, iar $x_i$ este o constantă sau o variabilă-domeniu.
   b) $x\ \Theta\ y$, cu x şi y constante sau variabile-domeniu, iar $\Theta$ operator aritmetic.

   Atomului $R(x_1x_2...x_k)$ îi corespunde aserțiunea „valorile acelor $x_i$ care sunt variabile trebuie alese astfel încât $x_1x_2...x_k$ să fie un tuplu în R. Iar pentru $x\ \Theta\ y$, x şi y trebuie să aibă astfel de valori încât $x\ \Theta\ y$ = adevărat".
3. Formulele, în calculul relațional pe domenii, folosesc operatorii $\wedge$, $\vee$, $\neg$ ca şi în calculul pe tupluri. Se folosesc de asemenea, $(\exists x)$ şi $(\forall x)$ pentru a forma expresii, însă x este o variabilă-domeniu, nu o variabilă-tuplu.

Variabilele-domeniu libere şi legate se definesc la fel ca în calculul relaţional pe tupluri.

O expresie în calculul relaţional pe domenii are forma: $\{ x_1x_2...x_k \mid \Psi(x_1,x_2,...,x_k) \}$, unde $\Psi$ este o formulă. Singurele variabile-domeniu libere ale formulei $\Psi$ sunt variabilele distincte $x_1,x_2,...,x_k$.

O expresie $\{ x_1x_2...x_k \mid \Psi(x_1,x_2,...,x_k) \}$ este sigură dacă:

a) $\Psi(x_1,x_2,...,x_k)$ = adevărat implică $x_i \in DOM(\Psi)$;
b) Dacă $(\exists u)\ (\omega(u))$ este o „sub-formulă” din $\Psi$, atunci $\omega(u)$ = adevărat implică $u \in DOM(\omega)$;
c) Dacă $(\forall u)\ (\omega(u))$ este o sub-formulă din $\Psi$, atunci $\omega(u)$ = fals implică $u \in DOM(\omega)$.

*Exemplu.* Fie o relaţie binară R (aritate 2); să se scrie o expresie care să fie egală cu R, dacă R are două sau mai multe elemente (tupluri), sau să fie egală cu $\varnothing$ în celălalte cazuri.

O astfel de expresie este, în calculul relaţional pe domenii: $\{ wx \mid (\exists y)\ (\exists z)\ (R(wz) \wedge R(yz) \wedge (w \neq y \vee x \neq z)) \}$.

Fie notaţia: $\Psi(w,x,y,z) = R(wx) \wedge R(yz) \wedge (w \neq y \vee x \neq z)$ şi fie R = { 12, 34} unde 12, 34 = tupluri.

Dacă facem (luăm) w = 1 şi x = 2, atunci formula $(\exists y)\ (\exists z)\ (\Psi(1, 2, y, z))$ = adevărat, deoarece putem lua y = 1 şi z = 3 pentru a face $\Psi$ = adevărat. De asemenea, dacă w = 1 şi x = 3, formula este adevărată, deoarece putem lua y = 1 şi z = 2. Aşadar, tuplurile 12 şi 34 se găsesc ambele în mulţimea reprezentată de expresia w.

Observăm încă că dacă se iau alte valori pentru x şi w, atunci $(\exists y)\ (\exists z)\ (\Psi(w, x, y, z))$ trebuie să fie falsă, deoarece clauza R(wx) din $\Psi$ este falsă. Aşadar, expresia aleasă, pentru calculul relaţional pe domenii, reprezintă o mulţime = R când R = { 12, 34 }.

Dacă R este o mulţime mono-element – de exemplu, { 12 } – atunci nicio valoare w, x nu satisface $(\exists y)\ (\exists z)\ (\Psi(w, x, y, z))$, deoarece prima clauză din $\Psi$, R(wx), este satisfăcută numai dacă w=1 şi x = 2, a doua clauză R(yz) este satisfăcută numai de y = 1 şi z = 2, şi atunci a treia clauză $(w \neq y \vee x \neq z)$ nu este satisfăcută, întrucât w = y = 1.

***Reducerea calculului relaţional pe tupluri la calculul relaţional pe domenii***

Fie o expresie $\{ t \mid \Psi(t) \}$ în calculul relaţional pe tupluri. Trecerea la o expresie echivalentă, în calculul relaţional pe domenii, este imediată.

1. Să admitem că t $\rightarrow$ aritate k. Introducem atunci k variabile-domeniu (variabile noi) $t_1, t_2, ..., t_k$ şi înlocuim expresia prin: $\{ t_1t_2…t_k \mid \Psi`(t_1, t_2, …, t_k) \}$, unde $\Psi'$ este $\Psi$, însă în care fiecare atom R(t) a fost înlocuit prin $R(t_1t_2...t_k)$, iar fiecare apariţie liberă a componentei t[i] este înlocuită tu $t_i$. Observăm că în $\Psi$ pot exista apariţii legate ale tuplului t, dacă acolo avem $(\exists u)$ sau $(\forall u)$; folosirea în acest mod a tuplului t nu este înlocuită.
2. În pasul următor, pentru fiecare cuantificator $(\exists u)$ sau $(\forall u)$, dacă u este de aritate m, se introduc m noi variabile-domeniu $u_1, u_2, ..., u_m$ şi în zonele controlate de cuantificatori, u[i] se înlocuieşte cu $u_i$, iar R(u) cu $R(u_1u_2...u_m)$. Apoi, $(\exists u)$ se înlocuieşte $(\exists u_1) ... (\exists u_m)$, iar $(\forall u)$ cu $(\forall u_1) ... (\forall u_m)$.

Rezultatul $\rightarrow$ o expresie în calculul relaţional pe domenii, care este echivalentă cu expresia iniţială, din calculul relaţional pe tupluri. Evident că valorile ce le pot lua $t_i$ sunt exact aceleaşi valori care pot fi luate de t[i] în expresia iniţială. Aşadar, dacă $\{ t \mid \Psi(t) \}$ este sigură, va fi sigură şi expresia echivalentă rezultată.

*Teoremă.* Fiecărei expresie sigure în calculul relațional pe tupluri îi corespunde o expresie sigură în calculul relațional pe domenii.

*Exemplu.* Fie expresia simplă pentru calculul compoziției a două relații binare R şi S:
$\{ w \mid (\exists u)\ (\exists v)\ (R(u) \wedge S(v) \wedge u[2] = v[1] \wedge w[1] = u[1] \wedge w[2] = v[2]) \}$.
(acest exemplu este folosit şi la trecerea de la algebra relațională la calculul relațional pe tupluri)

Acum, înlocuind w cu $w_1w_2$, u cu $u_1u_2$ şi v cu $v_1v_2$ obținem:
$\{ w \mid (\exists u_1)\ (\exists u_2)\ (\exists v_1)\ (\exists v_2)\ (R(u_1u_2) \wedge S(v_1v_2) \wedge u_2 = v_1 \wedge w_1 = u_1 \wedge w_2 = v_2) \}$.

## ***Trecerea de la calculul relațional pe domenii la algebra relațională***

*Teoremă.* Pentru fiecare expresie sigură din calculul relațional pe domenii există o expresie echivalentă în algebra relațională.