

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Departament de llenguatges i sistemes informàtics

**ACTUALITZACIÓ CONSISTENT DE
BASES DE DADES DEDUCTIVES**

Autor: Enric Mayol Sarroca
Director: Ernest Teniente i López

Barcelona, 2000

9. Conclusions i recerca futura

En aquesta tesi, hem proposat un nou mètode per a l'actualització de vistes i el manteniment de restriccions d'integritat en bases de dades deductives. Aquest mètode és una extensió del Mètode dels Esdeveniments [TO95].

El mètode proposat augmenta una base de dades deductiva amb un conjunt de regles, anomenades de transició i d'esdeveniment, que defineixen explícitament les insercions, esborrats i modificacions induïdes per l'actualització de la base de dades. Mitjançant aquestes regles i la utilització del procediment de resolució SLDNF, el nostre obté totes les solucions mínimes que permeten satisfer una petició d'actualització sense violar cap restricció d'integritat.

En la definició d'aquest nou mètode s'han considerat dos aspectes per separat. Per una banda, el mètode es defineix com una extensió del procediment de resolució SLDNF, i per l'altra, es proposa una arquitectura d'implementació en la que s'han tingut en compte explícitament diferents tècniques per a millorar l'eficiència del mètode.

9.1 Aportacions relatives a la definició del mètode

Respecte a la formalització del mètode, cal dir que s'ha demostrat la seva correctesa i completesa. Així doncs, donada una petició d'actualització, el mètode genera el conjunt de totes les solucions mínimes existents que permeten satisfer la petició d'actualització sense violar cap restricció d'integritat. En aquest sentit, podem assegurar que si el mètode no obté cap solució a una petició d'actualització, aleshores, no existeix cap manera d'actualitzar la base de dades extensional que permeti satisfer-la, sense violar la consistència de la base de dades.

Una de les principals aportacions del mètode respecte del seu precursor (i d'altres mètodes existents) és el fet de tenir definida la modificació com un nou tipus d'actualització bàsica. Aquesta aportació és especialment rellevant tenint en compte que la majoria de mètodes existents en aquesta àrea, solament tenen definits els operadors d'inserció i d'esborrat. Per altra banda, els mètodes que tenen definit l'operador de modificació, el restringeixen a modificacions de fets bàsics, en canvi, en el nostre mètode tant es poden realitzar modificacions de fets bàsics com sol·licitar modificacions de fets derivats.

Una característica addicional del nostre mètode és el fet que en la pròpia formalització del mètode s'ha incorporat el manteniment de les restriccions d'integritat de clau dels predicats definits a la base de dades. Així doncs, aquesta informació no s'ha de definir de forma explícita com a restriccions d'integritat addicionals com succeeix en altres mètodes.

En aquesta tesi també hem definit un marc per a comparar i classificar diferents mètodes d'actualització de vistes i de imposició de restriccions d'integritat. Com a resultat d'aquesta anàlisi, podem afirmar que el mètode presentat en aquesta tesi és més general que els mètodes proposats fins el moment actual, tan a nivell del problema que tracta, com del tipus de vistes i restriccions d'integritat que pot manejar. A més a més, el nostre mètode no presenta les mancances que pateixen alguns d'aquests mètodes, és a dir, obté únicament solucions vàlides i és capaç d'obtenir totes les solucions existents a una petició d'actualització.

9.2 Aportacions relatives a l'eficiència del mètode

L'aportació més rellevant del nostre mètode rau en el fet que té en compte l'eficiència del propi procés d'actualització de vistes i del manteniment de restriccions d'integritat. Concretament, en aquesta tesi proposem una arquitectura d'implementació del mètode en la que es defineixen tècniques per a la millora de l'eficiència del mètode proposat.

Per a millorar l'eficiència en el procés de manteniment de les restriccions d'integritat, es proposa la generació, en temps de compilació, del Graf de Precedències. Aquest graf estableix les interrelacions entre els violadors i reparadors potencials de les restriccions d'integritat. En temps d'execució, s'utilitza aquest graf per a determinar l'ordre en que cal comprovar (i reparar) les restriccions d'integritat. Amb aquesta tècnica s'aconsegueix reduir considerablement el nombre de vegades que cada restricció ha de ser comprovada durant tot el procés.

Per a millorar l'eficiència durant el procés d'actualització de vistes, es proposa realitzar una anàlisi de la petició d'actualització requerida per l'usuari, el contingut de la base de dades i les regles d'esdeveniment de la base de dades augmentada. Amb aquesta anàlisi s'aconsegueix, per una banda, reduir el nombre d'accessos que cal fer a la base de dades extensional per a traduir la petició d'actualització, i per l'altra, reduir el nombre d'alternatives a explorar per tal d'obtenir totes les traduccions a aquesta petició d'actualització.

També hem definit, en aquesta tesi, tres mètriques que ens han permès quantificar la millora de l'eficiència del mètode proposat respecte al Mètode dels Esdeveniments.

9.3 Recerca Futura

Aplicar el mètode a altres problemes d'actualització. El mètode definit en aquesta tesi, no té perquè limitar-se a ser aplicat únicament per a l'actualització de vistes i el manteniment de restriccions d'integritat. En realitat, aquest mètode es basa en l'avaluació de les regles d'esdeveniment de forma descendent. És a dir, donat un esdeveniment derivat el mètode determina quins esdeveniments bàsics són necessaris per a que es pugui induir aquest esdeveniment. Tal com es mostra a [TU95], existeixen diferents problemes d'actualització en bases de dades deductives que requereixen mètodes que segueixin aquest enfocament, per

tant, una primera línia de recerca futura consistirà en analitzar com aquest mètode pot ser aplicat a aquests problemes.

Aplicar el nostre mètode a bases de dades relacionals. Els sistemes de gestió de bases de dades relacionals segueixen essent els més àmpliament utilitzats en la indústria, les empreses de serveis i l'administració pública. Altres models, com l'orientat a objectes, comencen a introduir-se en aquests sectors. Però d'altres, com el deductiu, són més utilitzats en les universitats i centres de recerca, degut a que no s'han desenvolupat sistemes de gestió comercials que permetin la seva utilització de forma generalitzada en aquests sectors.

De totes formes, moltes de les extensions pròpies del model deductiu han estat incorporades darrerament al model relacional. Per tant, una segona línia de recerca futura consistirà en estudiar com el nostre mètode pot ajudar a millorar com resoldre els problemes de l'actualització de vistes i del manteniment de restriccions d'integritat en les bases de dades relacionals, i concretament, com les tècniques proposades en el nostre mètode podrien incorporar-se als sistemes de gestió de bases de dades relacionals comercialitzats en l'actualitat.

Integrar diferents enfocaments d'imposició de restriccions d'integritat. En aquesta tesi, s'ha considerat que a totes les restriccions d'integritat definides en la base de dades els hi aplicàvem una política de manteniment, és a dir, que al ser violades tenien que ser reparades. En canvi, seria interessant tenir en compte la semàntica de cada restricció d'integritat i, definir per a cada una d'elles quina política d'imposició de restriccions els hi és més adequada. Així, per exemple, podríem tenir restriccions d'integritat a mantenir i restriccions d'integritat que sols es volen comprovar. En aquest sentit, el nostre Graf de Precedències ens pot ser especialment útil. Podríem definir les condicions associades a les restriccions d'integritat a comprovar com a condicions de comprovació, mentre que les condicions de les restriccions a mantenir serien de comprovació o de generació segons sigui la seva estructura. D'aquesta forma, amb el mateix Graf de Precedències i amb el mateix mètode, podríem integrar les dues polítiques d'imposició de restriccions d'integritat.

Incorporar la informació de la transacció en l'obtenció de l'ordre de tractament de les restriccions d'integritat. La informació representada en el Graf de Precedències permet determinar l'ordre en que cal comprovar les condicions del graf (restriccions d'integritat) per tal de ser el més eficients possible. Com ja hem vist, donada una transacció T que donarà lloc a una solució S, el nombre de condicions comprovades (i reparades) és el menor possible. Però en canvi, en el cas de tenir una transacció T que no pot donar lloc a cap solució perquè viola alguna restricció d'integritat que no pot ser reparada, el nombre total de condicions considerades no sempre ha de ser el mínim. En aquest cas, la màxima eficiència s'obtidria si es pogués determinar a priori quina és la condició que no podrà ser reparada. Els criteris de selecció de nodes dels graf no poden determinar aquesta informació, ja que no depèn

únicament de l'estructura del Graf, sinó que també depèn de la pròpia transacció T. Així doncs, ens plantejem estudiar si és possible obtenir algun mecanisme que ens permeti obtenir aquesta informació, i per tant, obtenir per a una transacció T i un Graf de Precedències qualssevol l'ordre que fa que, en qualsevol cas, es comprovin el menor nombre de condicions del graf.

Simplificacions addicionals a les regles d'esdeveniment de l'A(D). En determinats casos, el nostre mètode pot obtenir solucions a una petició d'actualització que no són mínimes. Aquests casos es donen generalment quan la petició d'actualització U inclou actualitzacions de predicats derivats que contenen variables existencials en la seva definició. Per evitar aquesta situació, hem detectat que es podrien aplicar simplificacions addicionals a les regles d'esdeveniment. Les regles d'esdeveniment de l'A(D) han estat simplificades [Urp93] tenint en compte bàsicament el concepte de clau. Però no s'ha tingut en compte com aquestes regles serien avaluades en temps d'execució. Si tenim en compte que en el nostre enfocament fem una avaluació descendent d'aquestes regles, es podrien definir algunes simplificacions addicionals que es poden veure en els exemples següents:

Exemple 9.1: Consideri's les regles d'esdeveniment d'inserció I9 i I12 del predicat derivat Actiu(p) de l'exemple 3.1:

- (I9) $\iota\text{Actiu}(p) \leftarrow \text{Emp}(p,c) \wedge \neg\delta\text{Emp}(p,c) \wedge \neg\mu\text{Emp}(p,c,c1) \wedge \text{Baixa}(p) \wedge \delta\text{Baixa}(p)$
(I12) $\iota\text{Actiu}(p) \leftarrow \text{Emp}(p,c) \wedge \mu\text{Emp}(p,c,c1) \wedge c1 \neq c \wedge \text{Baixa}(p) \wedge \delta\text{Baixa}(p)$

Observi's que per a satisfer l'esdeveniment $\iota\text{Actiu}(p)$ la regla I9 solament requereix la ocurrència de $\delta\text{Baixa}(p)$. En canvi, la regla I12 requereix a més a més que ocorri l'esdeveniment $\mu\text{Emp}(p,c,c1)$. Per tant, la solució generada a partir de la regla I12 no seria mínima. Observi's també, que amb la ocurrència de l'esdeveniment $\delta\text{Baixa}(p)$, independentment de l'ocurrència de l'esdeveniment $\mu\text{Emp}(p,c,c1)$, es pot induir l'esdeveniment derivat $\iota\text{Actiu}(p)$.

Així doncs, eliminant la regla I12 i el literal $\neg\mu\text{Emp}(p,c,c1)$ de la regla I9, obtenim una regla equivalent que impedeix obtenir una solució no mínima per la mateixa petició d'actualització.

$$\iota\text{Actiu}(p) \leftarrow \text{Emp}(p,c) \wedge \neg\delta\text{Emp}(p,c) \wedge \text{Baixa}(p) \wedge \delta\text{Baixa}(p)$$

Aquesta situació es donarà sempre que un dels predicats del cos d'una regla de derivació tingui un argument no clau que no apareix en el cap de la regla ni en cap altre predicat del cos de la regla de derivació.

Una situació similar apareix en la restricció d'integritat següent:

Exemple 9.2: Consideri's les regles C12 i C15 de l'exemple 3.1.

$$(C12) \quad \neg Ic3(p,n) \leftarrow Numss(p,n) \wedge \neg \delta Numss(p,n) \wedge \neg \mu Numss(p,n,n1) \wedge Contractat(p) \wedge \delta Contractat(p)$$

$$(C15) \quad \neg Ic3(p,n) \leftarrow Numss(p,n) \wedge \mu Numss(p,n,n1) \wedge Contractat(p) \wedge \delta Contractat(p)$$

En aquesta cas, l'esdeveniment $\mu Numss(p,n,n1)$ és un reparador potencial de la condició C12 i un violador potencial de la condició C15. Però en realitat, aquest no és un reparador vàlid de la restricció Ic3. Sempre que es repari Ic3 a través de C12 amb l'esdeveniment $\mu Numss(p,n,n1)$, s'induirà una nova violació de Ic3 per C15, la qual no podrà ser reparada.

En aquest cas, podem definir una regla equivalent a les C12 i C15 de la mateixa manera que en l'exemple anterior.

$$\neg Ic3(p,n) \leftarrow Numss(p,n) \wedge \neg \delta Numss(p,n) \wedge Contractat(p) \wedge \delta Contractat(p)$$

Millorar la gestió del cas existencial. Un altre aspecte que cal estudiar en més profunditat consisteix en el mecanisme que fem servir per a assignar un valor a les variables existencials. En el nostre mètode suposem que el domini d'aquestes variables és finit, i considerem tots els valors d'aquest domini per a generar totes les solucions mínimes a una petició d'actualització. En aquest sentit, ens plantejem la possibilitat d'estudiar com aprofitar les idees exposades a [CTU97, FTU99].

En relació *aprocés de Manteniment de la Consistència* basat en la utilització del Graf de Precedències, existeixen diferents aspectes que es poden millorar substancialment.

Reduir el nombre de cicles. Un aspecte que s'hauria d'estudiar amb més detall, és com es pot reduir el nombre de cicles en el Graf de Precedències i quin és el millor criteri per a agrupar-los en nodes subgrafs. En el cas de tenir cicles que contenen algun/s node/s en comú pot ser especialment rellevant el criteri d'agrupació d'aquests nodes en subgrafs. En aquesta tesi, s'ha optat per considerar el cicle amb menor nombre de nodes com a un subgraf, i aquest subgraf pot passar a ser un node més d'algun altre del cicle. Així, en el Graf de Precedències poden aparèixer subgrafs que contenen subgrafs en la seva definició. Una alternativa a aquesta agrupació podria ser el considerar tots els cicles no disjunts agrupats en un únic node subgraf. En aquests casos, caldria analitzar amb més profunditat quines conseqüències (a nivell d'eficiència) comportaria utilitzar un criteri d'agrupació o un altre. És a dir, en quins casos una agrupació o una altre comporta visitar un menor nombre de vegades cada node. En aquesta estratègia per a poder eliminar i agrupar cicles entre sí, creiem que tècniques com les presentades a [FP97] ens poden ser de gran ajuda.

Millorar criteris selecció de nodes. En determinats casos, els criteris definits per a seleccionar el següent node a considerar del Graf de Precedències no són prou determinants. Per exemple, el primer cop que es tracta un subgraf (cicle), els nodes que componen aquest cicle poden estar tots marcats, per tant, tots tenen algun predecessor marcat i tots tenen associades condicions de generació. En aquest cas, els criteris de selecció no determinen quin

ha de ser el primer node a tractar. Segons sigui el primer node considerat, pot provocar que es comprovin més d'un cop alguna de les condicions del cicle. En aquests casos, podríem millorar aquests criteris de selecció, tenint en compte nombre de precedències en les que està involucrat cada node.

Marcatge intel·ligent de nodes. El mecanisme per a realitzar el marcatge dels nodes es podria més intel·ligentment. En aquest sentit, es podrien tenir en compte tècniques com les proposades a [LS93, LL96] per a determinar quan una consulta (o una condició del graf) no es pot veure realment afectada per una actualització (reparació). Amb l'ajut d'aquestes tècniques es podrien reduir considerablement el nombre de condicions que cal marcar després de cada reparació, i a la vegada, reduir encara més el nombre de vegades que cal comprovar cada condició durant tot el procés de manteniment de la consistència.