

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

1. Introducció

La tecnologia de les bases de dades ha sofert en els darrers anys una ràpida evolució. L'aparició del model relacional va determinar dos períodes clarament diferenciats. Un primer període previ a l'aparició d'aquest model i un període post-relacional encara vigent en l'actualitat.

En aquest segon període es poden identificar dues estratègies complementàries en l'evolució de la tecnologia de les bases de dades. Una primera tendència consisteix en la definició de nous models de bases de dades per a millorar algun aspecte que el model relacional no pot representar o manegar amb facilitat. En aquest sentit, han aparegut diversos models de bases de dades que estenen el model relacional en diferents aspectes. Per exemple, les bases de dades orientades a objectes, que integren en la mateixa base de dades, les dades i els tractament necessaris per a gestionar-les; les bases de dades actives, que afegeixen a les bases de dades la capacitat reactiva davant de certs canvis en el seu contingut; o les bases de dades deductives, que incorporen el coneixement dels usuaris en la definició de la base de dades.

La segona tendència en l'evolució consisteix en estendre el propi model relacional incorporant algunes de les característiques pròpies dels models anteriors. Per exemple, incorporant la capacitat reactiva; estenent el llenguatge de definició de vistes; permetent emmagatzemar procediments en la pròpia base de dades; o evolucionant cap a un model relacional orientat a objectes.

Les bases de dades deductives estenen les bases de dades relacionals perquè incorporen una part del coneixement de l'usuari en el propi esquema de la base de dades. Aquest coneixement es representa mitjançant les regles deductives i les restriccions d'integritat. Les regles deductives especifiquen, en forma declarativa, el coneixement que es pot deduir a partir dels fets bàsics que estan explícitament emmagatzemats a la base de dades. Aquest nous fets derivats es calculen o deriven mitjançant aquestes regles. Les restriccions d'integritat defineixen condicions que tot estat de la base de dades ha de satisfer. Encara que el model relacional permet definir algun tipus de vista i de restricció d'integritat, aquestes estan limitades a casos molt concrets i específics. El model deductiu permet definir-les molt més àmpliament i sense imposar tantes restriccions.

Aquest capítol introductori està dividit en tres seccions. A la primera secció, es descriu el problema d'actualització de vistes i d'imposició de restriccions d'integritat. A la segona, es fa una anàlisi i classificació de diferents mètodes, apareguts darrerament, orientats a resoldre aquests problemes. A la darrera secció, es fa una descripció general del mètode proposat en aquesta tesi.

1.1 Descripció del problema

Tot sistema de gestió de bases de dades deductiva, a part de permetre definir i realitzar consultes sobre el contingut d'una base de dades deductiva, ha d'incloure algun mecanisme que permeti actualitzar el contingut d'aquesta base de dades. Aquest mecanisme ha de permetre diferents tipus d'actualitzacions, com per exemple actualitzacions de fets bàsics o actualitzacions de fets derivats. En el procés d'actualització d'una base de dades deductiva, poden aparèixer diferents tipus de problemes. Una descripció i classificació d'aquests problemes es pot trobar a [TU95]. En particular, en aquesta tesi es consideren dos d'aquests problemes: la *imposició de restriccions d'integritat* i l'*actualització de vistes*.

La *imposició de restriccions d'integritat* consisteix en assegurar que, un cop aplicada una actualització del contingut de la base de dades, les restriccions d'integritat segueixen essent satisfetes. Quan un usuari fa una requesta per actualitzar la base de dades deductiva, les restriccions d'integritat d'aquesta base de dades poden ser violades. És a dir, les actualitzacions necessàries per a satisfer la petició de l'usuari junt amb el contingut de la base de dades poden provocar que alguna restricció d'integritat esdevingui falsa. Aleshores, cal aplicar alguna mesura correctora per tal d'assegurar la consistència de la base de dades. En aquest sentit, existeixen dos enfocaments alternatius per tal d'assegurar aquesta consistència.

L'enfocament més clàssic consisteix en la *comprovació de restriccions d'integritat*. En aquest enfocament, quan es detecta que una restricció esdevé violada per alguna actualització, es rebutja aquesta actualització i el contingut de la base de dades queda tal com estava en l'estat previ a l'aplicació d'aquesta actualització. El principal esforç dels treballs que segueixen aquest enfocament [GL90, Oli91, GCMD94, Sel95, CHM95] ha estat orientat a definir mecanismes, el més eficients possibles, per a detectar quan una restricció esdevé violada per alguna actualització. Però el principal inconvenient d'aquest enfocament rau en el fet que l'usuari no pot conèixer quins canvis addicionals hauria d'aplicar a la base de dades per tal de satisfer la seva requesta i que totes les restriccions d'integritat se satisfessin.

El segon enfocament és el *manteniment de restriccions d'integritat*. En aquest enfocament, al detectar-se que una restricció d'integritat esdevé violada per alguna actualització, s'intenta reparar aquesta violació realitzant actualitzacions addicionals de forma que la restricció d'integritat se satisfaci de nou. Solament en el cas de no ser possible aquesta reparació, es rebutja l'actualització. Així doncs, en qualsevol cas, es garanteix que en l'estat final de la base de dades (un cop actualitzada) se satisfan totes les restriccions d'integritat i se satisfà la requesta d'actualització de l'usuari. Aquest enfocament resol el problema de l'enfocament anterior. El principal esforç dels treballs que segueixen aquest enfocament [KM90, ML91, Wüt93, CST95, TO95, Dec97, LT97, Maa98, Sch98] ha estat orientat a definir mètodes efectius que obtinguin les diferents formes de reparar les restriccions d'integritat. En canvi, pocs són els treballs

[CFPT94, Ger94, FP97] que hagin considerat la definició de mètodes eficients per a resoldre aquest mateix problema.

El segon problema considerat en aquesta tesi és el de l'*actualització de vistes*. En una base de dades deductiva, els fets derivats poden no estar explícitament emmagatzemats a la base de dades, sinó que són deduïts a partir dels fets bàsics emmagatzemats i les regles de derivació. Així doncs, les peticions d'actualització de fets derivats (o vistes) han de ser traduïdes, de forma correcte, en actualitzacions de fets dels predicats bàsics que defineixen la vista o predicat derivat. Diversos treballs [GL90, LPS93, Wüt93, CHM95, CST95, TO95, Dec97, LT97] orientats a resoldre aquest problema defineixen mètodes per a obtenir diferents traduccions a aquestes peticions, però en la majoria de casos, no es considera explícitament l'eficiència en el procés de traducció d'aquests peticions d'actualització de vista.

Existeix una interrelació molt estreta entre la imposició de restriccions d'integritat i l'actualització de bases de dades deductives. En general, les restriccions d'integritat solament poden esdevenir violades per l'aplicació d'alguna actualització sobre la base de dades. Per exemple, les actualitzacions de fets bàsics resultants de la traducció d'una actualització d'una vista poden ser invalidades perquè violen alguna restricció d'integritat. Per altra banda, un mecanisme que restableixi la consistència d'una base de dades pot requerir resoldre el problema d'actualització de vistes en el cas que alguna de les restriccions d'integritat estigui definida en termes de predicats derivats.

Degut a l'existència d'aquesta interrelació entre aquests dos problemes, creiem que és necessari definir mètodes que combinin l'actualització de vistes i la imposició de restriccions d'integritat. A [TO95] ja es mostrava la necessitat de tractar aquests problemes de forma integrada. Mentre que els problemes de l'actualització de vistes i de la comprovació de restriccions d'integritat es poden resoldre en dues etapes separades, els problemes del manteniment de restriccions d'integritat i de l'actualització de vistes tenen que tractar-se de forma integrada.

Per altra banda, per tal de que aquests mètodes pugin tenir una aplicació real, cal que resolguin aquests problemes de forma correcta i completa, i a més a més, que en la implementació d'aquests mètodes es tinguin en compte aspectes d'eficiència.

En aquest sentit, en aquesta tesi, proposem un nou mètode efectiu (correcte i complet) que resol els problemes de l'actualització de vistes i el manteniment de restriccions d'integritat, de forma integrada i tenint en compte aspectes d'eficiència. És a dir, donada una petició d'actualització que inclou actualitzacions de vistes, el mètode proposat obté de forma eficient totes les possibles solucions (conjunts d'actualitzacions bàsiques) que satisfan la petició d'actualització i que no violen cap restricció d'integritat.

1.2 Estat de l'Art

En aquesta secció fem una anàlisi de la recerca realitzada en els darrers anys referent a l'actualització de vistes i el manteniment de restriccions d'integritat en bases de dades. Concretament, analitzem els diferents aspectes que cal tenir en compte durant el procés d'actualització de vistes i del manteniment de restriccions d'integritat. Tenint en compte aquests aspectes, definim un marc per a comparar i classificar els mètodes més rellevants proposats per a resoldre aquests dos problemes. Una anàlisi detallada de cada mètode i una comparació de cada un d'ells respecte al mètode proposat en aquesta tesi es pot trobar al capítol 5 d'aquest mateix document.

En aquesta anàlisi de l'estat de l'art, sols hem considerat en detall els mètodes apareguts a partir de 1993. Altres treballs publicats amb anterioritat a aquesta data poden trobar-se analitzats amb detall a [Ten92, TO95, FP93].

En aquesta anàlisi de l'estat de l'art, no hem considerat cap mètode definit completament en temps de compilació, sinó que solament hem considerat mètodes definits en temps d'execució. De totes formes, existeixen diferents treballs que segueixen aquest enfocament [Sch96, Pas97, Qui93]. Aquests mètodes estan basats en la generació en temps de compilació de programes predefinitos per a actualitzar vistes i/o mantenir restriccions d'integritat. En temps d'execució, l'usuari selecciona el programa a executar i assigna valors reals als paràmetres del programa predefinit segons sigui l'actualització que vol aplicar a la base de dades. La generació d'aquests programes predefinitos en temps de compilació comporta dificultats addicionals que no presenten els mètodes que segueixen un enfocament basat en temps d'execució. Aquestes dificultats vénen determinades bàsicament pel fet de no tenir disponible la informació referent al contingut de la base de dades extensional, ni de la petició real d'actualització.

Aquests dos enfocaments no són alternatius, sinó que són complementaris. En realitat, existeixen mètodes que segueixen un enfocament mixt en el sentit que, la seva definició inclou aspectes considerats en temps de compilació i aspectes considerats en temps d'execució. Per una banda, hi ha mètodes [CFPT94, Ger94, CHM95, Maa98, Sch98] que en temps de compilació defineixen i generen un conjunt de regles específiques (actives) per a traduir vistes i/o mantenir restriccions d'integritat. En temps d'execució, aquestes regles s'executen, però en certs casos, poden requerir un procés de monitorització d'aquesta execució. Per altra banda, existeixen mètodes [TO95, MT00] que temps d'execució defineixen com traduir l'actualització d'una vista i/o per mantenir una restricció d'integritat. Però aquests mètodes es basen en la utilització de regles generades en temps de compilació. Per tant, en els dos casos, considerem que aquests mètodes no estan completament definits en temps de compilació ni en temps d'execució.

En la definició d'aquest marc s'han considerat cinc dimensions a tenir en compte per tal de classificar i comparar els diferents mètodes analitzats. Aquestes dimensions són: El *problema*

considerat per cada mètode, el model de *base de dades* definit, el tipus de *petició d'actualització* que el mètode pot tractar, el *mecanisme* utilitzat per a resoldre el problema considerat i, finalment, les característiques de les *solucions* obtingudes.

En la figura següent (figura 1.1) es mostren les cinc dimensions que defineixen el nostre marc per a classificar i comparar l'estat de l'art en l'àrea de l'actualització de vistes i el manteniment de restriccions d'integritat.

Així doncs, tenint en compte aquestes cinc dimensions, tot mètode es pot descriure com l'aplicació d'un determinat mecanisme per tal de donar solució a un determinat problema considerant una base de dades i una petició d'actualització concreta.

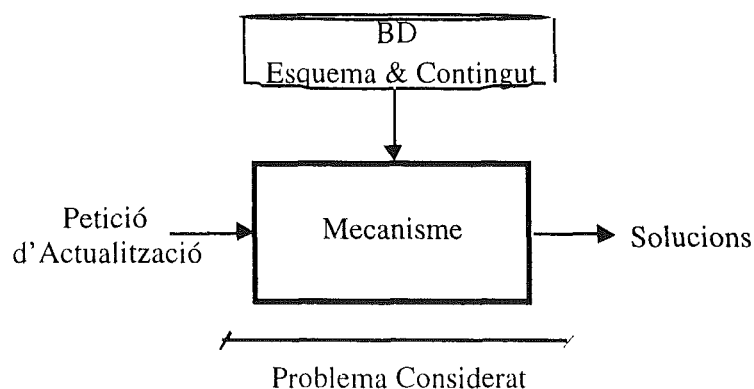


Figura 1.1 Factors considerats en l'actualització de vistes i el manteniment de restriccions

A continuació analitzarem en detall cadascuna d'aquestes cinc dimensions. Per a cada una d'elles, definirem alguns aspectes específics a considerar dins de cada dimensió, que ens permetran classificar i analitzar en més detall els mètodes considerats. A més a més, en cada una d'aquestes dimensions comentarem algunes de les limitacions més comuns que presenten els mètodes analitzats.

Al final d'aquesta secció es troba la taula 1.1 on es mostra l'anàlisi dels diversos mètodes estudiats.

1.2.1 Problema tractat

No tots els mètodes analitzats consideren de la mateixa manera els problemes de l'actualització de vistes i la imposició de les restriccions d'integritat. Per tant, en aquesta dimensió hem identificat tres aspectes a analitzar de cada mètode.

- Si el mètode gestiona l'actualització de vistes. En aquest cas a la taula 1.1 ho indiquem amb un 'Si' o 'No' segons sigui el cas.
- L'estratègia d'imposició de les restriccions d'integritat que aplica cada mètode. A la taula 1.1 solament considerem els casos de Comprovació ('Comp') i Manteniment ('Mant').

- L'enfocament que segueix aquest mètode. És a dir, si és un mètode definit totalment en temps d'execució ('Exec') o si en la seva definició és necessari, a més a més, un treball preparatori en temps de compilació ('Comp/Exec').

Per una banda tenim un primer grup de mètodes que consideren els problemes de l'actualització de vistes i del manteniment de restriccions d'integritat de forma integrada [KM90, Wüt93, CST95, TO95, Dec97, LT97, MT00]. Existeixen dos mètodes que també consideren els dos problemes de forma conjunta, però aplicant una estratègia de comprovació de les restriccions d'integritat [GL90, CHM95]. Existeix un únic mètode definit per a l'actualització de vistes [LPS93]. En canvi, ja són més els mètodes que apliquen una estratègia de manteniment de les restriccions d'integritat sense permetre l'actualització de vistes [ML91, CFPT94, Ger94, Maa98, Sch98].

Respecte a l'enfocament seguit pels mètodes analitzats, podem distingir dos grups de mètodes. Els mètodes [GL90, KM90, ML91, LPS93, Wüt93, CST95, Dec97, LT97] estan definits completament en temps d'execució, mentre que la resta de mètodes [CFPT94, Ger94, CHM95, TO95, Maa98, Sch98, MT00] requereixen un procés previ en temps de compilació. Concretament, aquest darrer grup de mètodes generen en temps de compilació un conjunt de regles (actives, d'esdeveniment, ...) que són necessàries en temps d'execució per a traduir les actualitzacions de vistes en actualitzacions de fets bàsics, i per a comprovar i reparar les restriccions d'integritat. En temps d'execució es defineix com aquestes regles s'utilitzen o s'executen.

1.2.2 Base de Dades

En aquesta dimensió del marc s'ha considerat bàsicament el poder expressiu del llenguatge de definició de base de dades emprat independentment del model de dades utilitzat per cada mètode. Els aspectes considerats en aquesta dimensió són els següents:

- El llenguatge de definició de la base de dades. Bàsicament, els llenguatges utilitzats són el relacional ('Rel'), extensions de la lògica de primer ordre ('Lògica') o un llenguatge orientat a objectes ('OO').
- La possibilitat o no de definir predicats derivats o vistes a l'esquema de la base de dades ('Si, No').
- Si cada mètode permet definir restriccions d'integritat sense limitacions ('Sí') o si el mètode imposa certes limitacions a la definició de restriccions d'integritat ('Limitada'). En particular, una limitació comú a diversos mètodes, és el fet de restringir la definició de les restriccions d'integritat a predicats bàsics i avaluables, és a dir, que no poden estar definides en termes de vistes. Aquesta limitació s'indica a la taula 1.1 amb el indicador 'Plana'.

- El tipus de restricció d'integritat que pot manegar cada un dels mètodes. 'Estàtica' es refereix a restriccions d'integritat que han de satisfer-se en tot estat de la base de dades, mentre que l'indicador 'Dinàmica' es refereix a les restriccions d'integritat que involucren diferents estats a la vegada.

El llenguatge utilitzat majoritàriament en la definició de les restriccions d'integritat és el llenguatge basat en la lògica de primer ordre. En canvi, els mètodes [CFPT94, Ger94, Sch98] el combinen amb el llenguatge relacional en la definició de l'esquema de la base de dades. El mètode [CHM95] utilitza el llenguatge de definició propi d'un model orientat objectes.

La principal limitació de la majoria dels mètodes analitzats rau en els tipus de restriccions d'integritat que poden tractar. Respecte el tipus de restricció d'integritat que pot mantenir cada mètode, tant sols els mètodes [Ger94, TO95, Maa98, MT00] poden mantenir restriccions d'integritat estàtiques i dinàmiques. La resta solament pot tractar restriccions d'integritat estàtiques. A més a més, molts dels mètodes analitzats imposen limitacions addicionals a la definició de restriccions d'integritat estàtiques. Les diferents limitacions que imposa cada mètode en la definició de les restriccions d'integritat estan descrites junt amb l'anàlisi detallada de cada mètode al capítol 5 d'aquesta tesi.

De tota manera, la limitació més generalitzada en els mètodes estudiats, en concret en els mètodes [Ger94, CST95, LT97, Maa98, Sch98], és el fet de restringir la definició de restriccions d'integritat tan sols a predicats bàsics i predicats avaluable. Aquesta limitació restringeix considerablement el poder expressiu del llenguatge de definició de les restriccions d'integritat ja que existeixen molts casos de restriccions d'integritat que no es podran definir ni mantenir pels mètodes anteriors.

Les restriccions d'integritat de l'exemple 1.1 són dos exemples clars de restriccions d'integritat escrites en forma de denegació, que no poden ésser definides utilitzant únicament predicats bàsics.

Exemple 1.1: La restricció d'integritat amb l'etiqueta (1) requereix un predicat derivat per assegurar que la definició de la restricció d'integritat sigui *permissible*. La segona restricció (2) no pot ser expressada per la majoria de mètodes sense la utilització del predicat derivat $Q(x)$:

$$(1) \quad \leftarrow P(x) \wedge \neg R(x, y)$$

$$(2) \quad \leftarrow P(x) \wedge Q(x)$$

$$Q(x) \leftarrow S(x)$$

$$Q(x) \leftarrow T(x)$$

□

Existeixen altres aspectes que podien ser considerats en aquesta mateixa dimensió. Per exemple, la possibilitat de tenir literals negats en el cos de les regles, o la possibilitat de definir predicats derivats recursius. Aquests aspectes no han estat contemplats explícitament, ja que no

ens aportaven criteris nous per a diferenciar uns mètodes d'altres. Respecte a la negació, tots els mètodes considerats la contemplen d'una forma més o menys explícita. En canvi, pocs són els mètodes que permeten i defineixen un mecanisme especial per a gestionar correctament l'actualització de vistes recursives.

1.2.3 Petició d'actualització

Els aspectes considerats en aquesta dimensió són:

- Si la petició d'actualització pot ser múltiple o si ha de contenir una sola actualització ('Si', 'No').
- Quin tipus d'actualització es poden sol·licitar: insercions, esborrats o modificacions. A la taula 1.1 indiquem aquests tipus d'actualització amb els indicadors 'ι', 'δ' i 'μ', respectivament.

Respecte a la petició d'actualització sol·licitada per l'usuari, tots els mètodes analitzats, a excepció de [GL90], permeten que aquesta sigui múltiple, és a dir, que estigui composta per més d'una actualització i que combini diferents tipus d'actualitzacions.

Tots els mètodes analitzats tenen definits els operadors d'actualització bàsics d'inserció i d'esborrat. Únicament els mètodes [CFPT94, Ger94, MT00] tenen definit l'operador bàsic de modificació. A més a més, cal destacar que els mètodes [CFPT94, Ger94] solament permeten aplicar l'operador de modificació als fets bàsics de la base de dades, mentre que el mètode [MT00] és l'únic que pot aplicar l'operador de modificació a predicats derivats o vistes.

El fet de no considerar els mateixos tipus d'operadors d'actualització en tots els mètodes comporta una diferència semàntica entre una inserció (o esborrat) en els mètodes que contemplen l'operador de modificació respecte als mètodes que no el contemplen. Aquesta diferència es podrà observar en detall en la comparació del Mètode dels Esdeveniments [TO95] amb el mètode presentat en aquesta tesi a la secció 5.4.

1.2.4 Mecanisme de traducció

Respecte al mecanisme utilitzat per cada mètode per a realitzar la traducció de peticions d'actualització de vistes i per a assegurar la satisfacció de les restriccions d'integritat, hem considerat tres aspectes:

- La tècnica base utilitzada en el procés de traducció aplicat per cada mètode. Les tècniques utilitzades consisteixen en un procés de Desplegament ('Despleg.') de la petició d'actualització; o en aplicar una extensió del procediment de resolució 'SLDNF'; o en l'execució de regles 'Actives'.
- Si en el procés de traducció, es tenen en compte els fets emmagatzemats a la base de dades. A la taula 1.1 ho indiquem amb l'etiqueta 'Si' o 'No' segons sigui el cas.

- Si durant el procés de traducció és necessària la participació de l'usuari. També ho indiquem amb l'etiqueta 'Si' o 'No'.

Tenint en compte la tècnica utilitzada per a l'obtenció de solucions a la petició d'actualització, podem classificar els mètodes analitzats en tres grups. El primer grup el componen els mètodes [Wüt93, CST95, LT97] que apliquen un procediment de desplegament sobre la petició d'actualització tenint en compte les vistes i restriccions d'integritat definides a la base de dades, obtenint al final una fórmula que representa les solucions a la petició inicial. El segon grup de mètodes el componen els mètodes ([GL90, KM90, TO95, Dec97, MT00]) que defineixen una extensió del procediment de resolució SLDNF. El tercer grup el componen mètodes, com per exemple [CFPT94, Ger94, CHM95, Maa98, Sch98]. Aquests mètodes segueixen un enfocament actiu que consisteix bàsicament en la generació d'un conjunt de regles (actives) que executaran per a mantenir les restriccions d'integritat. Solament en el cas de [CHM95] aquestes regles servirán per a actualitzar vistes. La resta de mètodes defineixen tècniques específiques que no són comuns a altres mètodes.

Durant el procés de generació de solucions a una petició d'actualització, els mètodes definits en temps d'execució (totalment o en part), tenen accessible la informació referent al contingut de la base de dades, és a dir, els fets emmagatzemats a la base de dades. Aquesta informació és d'especial rellevància per tal d'obtenir solucions que solament incloguin les actualitzacions realment necessàries per a satisfer la petició d'actualització, i que aquestes actualitzacions realment es puguin aplicar a la base de dades. Alguns dels mètodes analitzats [GL90, KM90, LPS93, Wüt93, Dec97, LT97], no tenen en compte en tot moment aquesta informació, i per tant, en alguns casos poden generar solucions que incloguin actualitzacions innecessàries (solucions mínimes), o fins i tot, no poden obtenir certes solucions que altres mètodes sí obtenen. Casos concrets d'aquestes situacions es mostren al capítol 5, on s'analitzen en més detall cada un dels mètodes.

El darrer aspecte considerat fa referència a la necessitat de requerir la participació de l'usuari o dissenyador durant el procés de generació de solucions a una petició d'actualització. En aquest cas, no hem considerat com a participació de l'usuari el fet de proporcionar valors a variables existencials, ni el fet de triar una solució a aplicar a la base de dades, en cas de que el mètode generi més d'una alternativa. Aquesta participació generalment és necessària en mètodes basats en l'execució de regles actives, per a que el dissenyador defineixi o simplifiqui el conjunt final de regles actives a considerar [CFPT94, Ger94]. Altre mètodes com [Wüt93, LT97] requereixen a l'usuari final per a que guiï el procés de traducció davant de dues o més alternatives. La necessitat d'aquesta participació indica, en la majoria de casos, que el mètode no és totalment automàtic, . que l'obtenció o no d'una determinada solució depèn de les decisions que prengui el dissenyador o usuari en cada moment.

1.2.5 Solucions generades

Donades diverses peticions d'actualització associades a diferents bases de dades, i per a cada un dels mètodes analitzats, s'han contrastat les diferents solucions obtingudes.

En qualsevol cas, s'espera que tota solució obtinguda per un mètode satisfaci la petició inicial d'actualització i no violi cap restricció d'integritat. Però, hem detectat que existeixen mètodes que obtenen solucions que no són correctes, és a dir, que no satisfan la petició inicial d'actualització encara que assegurin la no violació de cap restricció d'integritat. A més a més, per a una petició d'actualització i una base de dades concreta, poden existir més d'una solució que satisfacin aquesta petició sense violar cap restricció d'integritat. És doncs també d'esperar que qualsevol mètode sigui capaç d'obtenir totes les solucions alternatives a una petició d'actualització. Però de la mateixa manera que en el cas anterior, hem detectat mètodes que en determinats casos no poden obtenir solucions que altres mètodes sí obtenen, és a dir, mètodes que no són complets.

Així doncs, els aspectes considerats relacionats amb les solucions obtingudes són els següents:

- Si el mètode és correcte, és a dir, que tota solució obtinguda pel mètode satisfà la petició inicial d'actualització i no viola cap restricció d'integritat.
- Si el mètode és complet, és a dir, si és capaç de generar totes les possibles solucions que satisfacin la petició inicial d'actualització sense violar cap restricció d'integritat.

A la taula 1.1 i per als dos aspectes anteriors, hem indicat amb el indicador 'Si' el cas en que està demostrada la correctesa / completesa del mètode. La etiqueta 'No' indica que el mètode no és correcte / complet, i a més a més, hem detectat exemples en que es pot veure aquesta limitació. En els casos en que no està demostrada la correctesa / completesa del mètode i no hem trobat cap exemple en que es mostri aquesta limitació, ho indicarem a la taula 1.1 amb el indicador 'No provat'.

Els mètodes que tenen demostrada la seva correctesa són [CHM95, CST95, TO95, MT00] i els que tenen demostrada la seva completesa són [TO95, MT00].

Els mètodes que no tenen demostrada la correctesa i als que no els hem detectat cap exemple en que no siguin correctes són [GL90, ML91, LPS93, Wüt93, Dec97, Sch98] i els que no tenen demostrada la completesa i als que tampoc hem detectat cap exemple en que no poden generar alguna solució existent són [ML91, LPS93, CST95, LT97].

Per a la resta de mètodes, hem identificat exemples específics on es mostra que obtenen solucions no correctes i/o que no poden obtenir solucions que altres mètodes sí que obtenen. Aquestes exemples es troben descrits amb detall al capítol 5 on s'analitzen les principals limitacions dels mètodes analitzats.

1.2.6 Síntesi dels mètodes actuals

A la taula 1.1. es mostra el resultat de l'anàlisi realitzat als mètodes considerats en aquest estudi de l'estat de l'art en l'àmbit de l'actualització de vistes i el manteniment de les restriccions d'integritat. En aquesta taula, es pot veure com ha estat classificat cada mètode tenint en compte els diferents aspectes considerats en cada una de les 5 dimensions que defineixen el marc comparatiu.

Aquesta taula pot servir de punt de partida base per a fer una classificació i comparació dels diferents mètodes actuals per a l'actualització de vistes i el manteniment de restriccions d'integritat.

D'una anàlisi exhaustiva de la taula 1.1, se'n pot deduir, sense necessitat d'entrar en detall amb la comparació específica de cada mètode, que el mètode presentat en aquesta tesi és més general que tots els mètodes proposats fins el moment actual per a tractar els problemes d'actualització de vistes i de manteniment de restriccions d'integritat en bases de dades deductives. El nostre mètode és més general tan a nivell del problema que tracta, com del tipus de vistes i restriccions d'integritat que pot manejar. A més a més, no presenta cap de les possibles mancances que presenten la resta de mètodes, és a dir, obté únicament solucions vàlides i és capaç d'obtenir totes les solucions existents, és per tant correcte i complet.

Cal destacar que l'aportació més important del nostre treball en relació als treballs analitzats en aquest marc rau en la incorporació de tècniques específiques per a millorar l'eficiència del procés de traducció. De tots els mètodes analitzats en aquesta secció, sols hi ha els mètodes [CFPT94, Ger94] i el treball [FP97] que posen especial èmfasi en la definició de mètodes eficients. En aquests mètodes, es proposen tècniques específiques per a la millora de l'eficiència en el procés d'obtenció de solucions a una petició d'actualització. Mentre que els mètodes [CFPT94, Ger94, FP97] es centren en la millora de l'eficiència en el procés de manteniment de les restriccions d'integritat, el mètode proposat en aquesta tesi presenta tècniques per a la millora en la eficiència tant en el procés de manteniment de restriccions d'integritat, com en el procés d'actualització de vistes. La descripció d'aquestes tècniques i la comparació detallada a nivell d'eficiència entre aquests mètodes es pot trobar al capítol 7 d'aquesta tesi.

Mètode	Problema			Base de Dades				Ref. Actualitz.		Mecanisme			Solucions	
	Actualitz. Vistes	Imposició R.L.	Exec / Comp	Llenguat. Definició	Vistes	Definició R.L.	Tipus R.L.	Múltiple	Operad. Actualitz.	Tècnica	Fets BD	Particip. Usuari	Completa	Coberta
GL90	Sí	Comp	Exec	Lògica	Sí	Sí	Estàtica	No	1 δ	SLDNF	No	No	No	No Provat
KM90	Sí	Mant	Exec	Lògica	Sí	Sí	Estàtica	Sí	1 δ	SLDNF	No	No	No	No
ML91	No	Mant	Exec	Lògica	Sí	Sí	Estàtica	Sí	1 δ	---	Sí	No	No Provat	No Provat
LPS93	Sí	No	Exec	Lògica	Sí	No	No	Sí	1 δ	---	No	No	No Provat	No Provat
Wit93	Sí	Mant	Exec	Lògica	Sí	Sí	Estàtica	Sí	1 δ	Despleg.	No	Sí	No Provat	No Provat
CRPT94	No	Mant	Comp Exec	Rel. Lògica	Sí	Limitada	Estàtica	Sí	1 δ μ	Actives	Sí	No	No	No
Ger94	No	Mant	Comp Exec	Rel. Lògica	No	Plana Limitada	Estàtica Dinàmica	Sí	1 δ μ	Actives	Sí	No	No	No
CHM95	Sí	Comp Mant	Comp Exec	OO	Classe Atribut	Limitada	Estàtica	Sí	1 δ	Actives	Sí	No	No	Sí
CST95	Sí	Mant	Exec	Lògica	Sí	Plana Limitada	Estàtica	Sí	1 δ	Despleg.	Sí	No	No Provat	Sí
TO95	Sí	Mant	Comp Exec	Lògica	Sí	Sí	Estàtica Dinàmica	Sí	1 δ	SLDNF	Sí	No	Sí	Sí
Dec97	Sí	Mant	Exec	Lògica	Sí	Sí	Estàtica	Sí	1 δ	SLDNF	No	No	No	No Provat
LT97	Sí	Mant	Exec	Lògica	Sí	Plana Limitada	Estàtica	Sí	1 δ	Despleg.	No	Sí	No Provat	No
Maz98	No	Mant Restaura	Comp Exec	Lògica	No	Plana Limitada	Estàtica Dinàmica	Sí	1 δ	Actives	Sí	No	No	No
Sch98	No	Mant	Comp Exec	Rel.	No	Plana Limitada	Estàtica	Sí	1 δ	Actives	Sí	No	No	No Provat
MT00	Sí	Mant	Comp Exec	Lògica	Sí	Sí	Estàtica Dinàmica	Sí	1 δ μ	SLDNF	Sí	No	Sí	Sí

Taula 1.1 Taula resum de la anàlisi de diferents mètodes

1.3 Descripció del mètode

En aquesta tesi, proposem un nou mètode que és una extensió del Mètode dels Esdeveniments [TO95]. Donada una petició d'actualització, aquest mètode tradueix de forma automàtica aquesta petició en el conjunt de totes les possibles formes d'actualitzar la base de dades extensional de forma que la petició sigui satisfeta i que no es violi cap restricció d'integritat.

De la mateixa manera que el nostre precursor [TO95], el nostre mètode està basat en un conjunt de regles que defineixen la diferència entre dos estats consecutius de la base de dades. Aquesta diferència es determina definint explícitament les insercions, esborrats i, en el nostre cas, també les modificacions que es poden induir com a conseqüència de l'aplicació d'una actualització a la base de dades. Aquestes regles, conjuntament amb la base de dades original, componen el que anomenem base de dades augmentada $A(D)$ [Urp93].

El nostre mètode està basat en una extensió del procediment de resolució SLDNF [Llo87]. Sigui D una base de dades deductiva, $A(D)$ la base de dades augmentada associada, U una petició inicial d'actualització i T un conjunt d'actualitzacions de fets bàsics. Direm que el conjunt T satisfà la petició d'actualització U i no viola cap restricció d'integritat de D si, utilitzant la resolució SLDNF, l'objectiu $\leftarrow U \wedge \neg \text{IC}^1$ té èxit amb el conjunt d'entrada $A(D) \cup T$. Així doncs, el mètode consistirà en fer tenir èxit a les derivacions SLDNF fracassades. Per a fer-ho, s'inclouran al conjunt T aquelles actualitzacions de fets bàsics que cal realitzar per tal de que la derivació assoleixi l'èxit. Les diferents formes com es pot assolir aquest èxit es corresponen a les diferents solucions a la petició d'actualització U .

El nostre mètode segueix un enfocament mixt en el sentit que, encara que la definició és en temps d'execució, en temps de compilació s'ha realitzat un treball preparatori. En concret, cal que en temps de definició s'hagi generat la base de dades augmentada $A(D)$, entre altres coses.

El mètode proposat es demostrarà que és correcte i complet. En aquest sentit, assegurem que donada una petició d'actualització U , el mètode obté totes les possibles formes de satisfer aquesta petició i que, a la vegada, es satisfacin les restriccions d'integritat definides a la base de dades.

A diferència del nostre precursor, el nostre mètode gestiona les modificacions de fets com un nou tipus d'actualització bàsic. Aquest nou tipus d'actualització, junt amb la demostració de correctesa i completesa, és una de les principals aportacions del nostre mètode respecte als mètodes considerats en la secció anterior. La segona gran aportació del nostre mètode és el fet d'utilitzar de forma explícita tècniques per a millorar l'eficiència del procés de traducció de vistes i del procés de manteniment de restriccions d'integritat.

¹ El literal $\neg \text{IC}$ s'utilitza per a assegurar que les restriccions d'integritat no es violen.

Per a millorar l'eficiència del procés de manteniment de restriccions d'integritat, proposem una tècnica per a determinar l'ordre en que cal comprovar les restriccions d'integritat. Aquesta tècnica està basada en la generació en temps de compilació del anomenat Graf de Precedències, el qual estableix les relacions entre violadors i reparadors potencials d'aquestes restriccions. Aquest Graf és utilitzat en temps d'execució per a determinar l'ordre en què es comproven i reparen les restriccions d'integritat. Aquest ordre redueix el nombre de vegades que cada restricció d'integritat ha de ser comprovada (i reparada) després de reparar qualsevol altre restricció.

Per a millorar l'eficiència del procés d'actualització de vistes, proposem fer una anàlisi de la petició d'actualització, del contingut de la base de dades i de les regles de la base de dades augmentada abans d'iniciar la traducció de la petició d'actualització U. Aquesta anàlisi té com a objectiu el minimitzar el nombre d'accessos al contingut de base de dades que cal realitzar per a traduir la petició d'actualització, i per altra banda, aquesta anàlisi també ha de permetre determinar quines alternatives no podran donar lloc a una traducció vàlida a la petició U, permetent així, considerar únicament aquelles alternatives que sí proporcionaran una traducció vàlida a U.

La definició detallada del mètode proposat s'ha dividit en dues parts clarament diferenciades. En la primera part de la tesi es presenta la definició i formalització del mètode. En la segona part es descriu una arquitectura d'implementació del mètode i es presenten les diferents tècniques per a incrementar l'eficiència del mètode.

La primera part de la tesi està composta pels capítols 2 al 5. El capítol 2 repassa el concepte de base de dades deductiva. Al capítol 3 s'introdueix el concepte de base de dades augmentada $A(D)$ (segons [Urp93]) i es presenten un conjunt de reescriptures i simplificacions addicionals a les realitzades a [Urp93]. Al capítol 4, es presenta la formalització del mètode, es mostre mitjançant diferents exemples l'aplicació del mètode i es demostra la completesa i correctesa del mateix. Al capítol 5 es fa una comparació detallada, a nivell de definició, del mètode proposat respecte a diferents mètodes que tracten el mateixos problemes que el nostre i que han estat inclosos en l'estat de l'art de la secció anterior.

La segona part de la tesi inclou la resta de seccions. Al capítol 6 presenta una arquitectura d'implementació del mètode, descrivint en detall les diferents tècniques per a millorar l'eficiència del mètode. Al capítol 7 es compara, a nivell d'eficiència, el nostre mètode amb altres mètodes que consideren aspectes d'eficiència en la definició dels seus mètodes. Al capítol 8, es fa una breu descripció de la implementació realitzada del mètode. Finalment, al capítol 9 es presenten les conclusions i el treball futur.

Alguns dels materials presentats en aquesta tesi han aparegut a [MT93, MT95, MT96, MT97, MT99a, MT99b, MT00]. La realització d'aquesta tesi està emmarcada en els projectes CICYT PRONTIC TIC94-0512 i TIC-97-1157.