CoLogNET Workshop on Logic-Based Methods for Information Integration - 23 August 2003

Integrity checking for combined databases

Davide Martinenghi

Computer Science Roskilde University

Computer Science, building 42.1
Roskilde University
Universitetsvej 1
P.O. Box 260
DK-4000 Roskilde
Denmark
Phone: +45 4674 2000
Fax: +45 4674 3072
www.dat.ruc.dk

Content

- Description of the problem
- A simplification framework
- GaV and LaV mappings
- Application to data integration
- Examples

Description of the problem

- ICs are properties of the DB that must always hold
- The integrity must be checked wrt the ICs after every update (typically tested in an *ad hoc* way at the application level)
- In a data integration system, it's the same
- Idea: generate specialized versions of the ICs to be automatically executed
  - For expected kinds of updates
  - Assuming the integrity before the update
  - Generalize this technique to data integration systems

A simplification framework

1. Produce a weakest precondition

   Ex: $\varphi = \leftarrow p(x)$ $U = p(a)$

   $After^U(\varphi) = \leftarrow (p(x) \vee x=a)$

A condition about the updated state that can be checked in the present state

2. Use the fact that φ was known to hold before the update (Cond. Weak. Prec.).
3. Take the weakest CWP.

DEF: $Simp^U(\varphi) = Weaken_\varphi(After^U(\varphi))$

A simplification framework - Example

$\varphi = \leftarrow m(x, y) \land m(x, z) \land y \neq z$

checked by posing it as a query against the DB and expecting an empty answer

$U = m(Bob, Alice)$

$simp^U(\varphi) = \leftarrow m(Bob, y) \land y \neq Alice$

Mappings

- Mapping = a way to associate n local DBs to a global DB
- GaV mapping = the global DB is expressed as a set of views over the local sources.
- LaV mapping = the local DBs are expressed as a set of views over the global DB.
- We assume:
  - sound mappings (the views produce only but not necessarily all correct information)
  - no existential quantifier in LaV mappings
  - $\Rightarrow$ LaV mappings can be rewritten as GaV mappings without skolemization

Mappings - example

LaV mapping L =

{ $m_1(x, y) \rightarrow m(x, y) \wedge n(x, it)$,

$m_2(x, y) \rightarrow m(x, y) \wedge n(x, dk)$ }

GaV mapping M_L =

{ $m(x, y) \leftarrow m_1(x, y)$,

$m(x, y) \leftarrow m_2(x, y)$,

$n(x, y) \leftarrow m_1(x, z) \wedge y{=}it$,

$n(x, y) \leftarrow m_2(x, z) \wedge y{=}dk$ }

Application to data integration

- $After^{M}(\varphi)$ is a weakest precondition
  M is a GaV mapping
- $Simp^{O}_{\Delta}(\varphi) = Weaken_{\Delta}(After^{O}(\varphi))$

A condition about the global DB that can be checked on the local DBs

Conditions known to hold locally

Conditions to check globally

Example 1

$\varphi = \leftarrow m(x, y) \wedge m(x, z) \wedge y \neq z$

$\varphi_1 = \leftarrow m_1(x, y) \wedge m_1(x, z) \wedge y \neq z$

$\varphi_2 = \leftarrow m_2(x, y) \wedge m_2(x, z) \wedge y \neq z$

Check:

$\{ \leftarrow m_1(x, y) \wedge m_1(x, z) \wedge y \neq z,$

$\leftarrow m_1(x, y) \wedge m_2(x, z) \wedge y \neq z,$

$\leftarrow m_2(x, y) \wedge m_1(x, z) \wedge y \neq z,$

$\leftarrow m_2(x, y) \wedge m_2(x, z) \wedge y \neq z \}$

...

Example 1

$\varphi = \leftarrow m(x, y) \wedge m(x, z) \wedge y \neq z$

$\varphi_1 = \leftarrow m_1(x, y) \wedge m_1(x, z) \wedge y \neq z$

$\varphi_2 = \leftarrow m_2(x, y) \wedge m_2(x, z) \wedge y \neq z$

Check:

{ ~~$\leftarrow m_1(x, y) \wedge m_1(x, z) \wedge y \neq z$~~,

$\leftarrow m_1(x, y) \wedge m_2(x, z) \wedge y \neq z$,

~~$\leftarrow m_2(x, y) \wedge m_1(x, z) \wedge y \neq z$~~,

~~$\leftarrow m_2(x, y) \wedge m_2(x, z) \wedge y \neq z$~~ }

$= \mathrm{Simp}^M_{\varphi 1 \wedge \varphi 2}(\varphi)$

…

Example 1

$\varphi = \leftarrow m(x, y) \wedge m(x, z) \wedge y \neq z$

$\varphi_1 = \leftarrow m_1(x, y) \wedge m_1 (x, z) \wedge y \neq z$

$\varphi_2 = \leftarrow m_2(x, y) \wedge m_2 (x, z) \wedge y \neq z$

Check:

{ ~~$\leftarrow m_1(x, y) \wedge m_1 (x, z) \wedge y \neq z$~~,

~~$\leftarrow m_1(x, y) \wedge m_2 (x, z) \wedge y \neq z$~~,

~~$\leftarrow m_2(x, y) \wedge m_1 (x, z) \wedge y \neq z$~~,

~~$\leftarrow m_2(x, y) \wedge m_2 (x, z) \wedge y \neq z$~~ }

$= \mathrm{Simp}^M_{\varphi 1 \wedge \varphi 2 \wedge \varphi 1,2}(\varphi)$

If we knew $\varphi_{1,2} = \leftarrow m_1(x, y) \wedge m_2 (x, z)$

Example 2

$M = \{\ f(i,t,r) \leftarrow m(i,t,y) \wedge r(i,r)\ \}$

$\varphi_1 = \{\leftarrow m(i,t_1,y_1) \wedge m(i,t_2,y_2) \wedge t_1 \neq t_2,$

$\leftarrow m(i,t_1,y_1) \wedge m(i,t_2,y_2) \wedge y_1 \neq y_2\}$

$\varphi_{1,2} = \leftarrow r(i, r) \wedge \neg m(i,t,y)$

$\varphi = \{ \leftarrow f(i,t_1,r_1) \wedge f(i,t_2,r_2) \wedge t_1 \neq t_2,$

$\leftarrow f(i,t_1,r_1) \wedge f(i,t_2,r_2) \wedge r_1 \neq r_2\}$

Global check:

$\{ \leftarrow m(i,t_1,y_1) \wedge r(i, r_1) \wedge m(i,t_2,y_2) \wedge r(i, r_2) \wedge t_1 \neq t_2,$

$\leftarrow m(i,t_1,y_1) \wedge r(i, r_1) \wedge m(i,t_2,y_2) \wedge r(i, r_2) \wedge r_1 \neq r_2\}$

…

Example 2

$M = \{\ f(i,t,r) \leftarrow m(i,t,y) \wedge r(i,r)\ \}$

$\varphi_1 = \{ \leftarrow m(i,t_1,y_1) \wedge m(i,t_2,y_2) \wedge t_1 \neq t_2,$

$\quad \leftarrow m(i,t_1,y_1) \wedge m(i,t_2,y_2) \wedge y_1 \neq y_2 \}$

$\varphi_{1,2} = \leftarrow r(i, r) \wedge \neg m(i,t,y)$

$\varphi = \{ \leftarrow f(i,t_1,r_1) \wedge f(i,t_2,r_2) \wedge t_1 \neq t_2,$

$\quad \leftarrow f(i,t_1,r_1) \wedge f(i,t_2,r_2) \wedge r_1 \neq r_2 \}$

Global check:

{ ~~$\leftarrow m(i,t_1,y_1) \wedge r(i, r_1) \wedge m(i,t_2,y_2) \wedge r(i, r_2) \wedge t_1 \neq t_2$~~,

$\leftarrow$ ~~$m(i,t_1,y_1) \wedge$~~ $r(i, r_1) \wedge$ ~~$m(i,t_2,y_2) \wedge$~~ $r(i, r_2) \wedge r_1 \neq r_2$ }

$= \mathrm{Simp}^{M}_{\varphi 1 \wedge \varphi 1,2}(\varphi)$

Summary

- Express the data integration in terms of a GaV-mapping
- Reformulate the condition to check in terms of the sources by calculating a weakest precondition wrt the mapping
- Remove from it all conditions known to hold locally (plus possible cross-conditions)

References

1. H. Christiansen, D. Martinenghi

   Simplification of integrity constraints for data integration

   Submitted to FoIKS '04

2. H. Christiansen, D. Martinenghi

   Simplification of database integrity constraints revisited: A transformational approach

   Accepted for presentation at LOPSTR '03

http://www.dat.ruc.dk/~dm/publications