First Steps in Pointfree Functional Dependency Theory

Nuno Macedo

October 2010

Program Semantics, Verification, and Construction
MAPi Doctoral Program in Computer Science

Relational Calculus

- Two views on relational calculus:
  - Relational Databases (Codd, 1970's):
    - *n*-ary relations;
    - Each object is a relation of *attributes;*
  - "Classical" calculus of relations (De Morgan's, 1860's):
    - Binary relations;
    - From which the *Algebra of Programming* emerged (taught at this course!);
    - Simple, pointfree style.

Merging the Calculi

- On relational databases, binary relations are just relations with arity of 2;
- However, what if we reason about them using *pointfree* techniques?
- A pointfree representation of *functional dependency* theory will be presented.

J. N. Oliveira, First Steps in Pointfree Functional Dependency Theory, 2005

Functional Dependency

- *Functional dependency* (FD) defines the *semantic* of a scheme;
- Example: $S = \{\textsc{Pilot}, \textsc{Flight}, \textsc{Date}, \textsc{Departs}\}$
- *"A single pilot is assigned to a given flight, on a given date"*;
- attribute *PILOT* is functionally dependent on *FLIGHT* and *DATE*. $\textsc{Flight}\ \textsc{Date} \rightarrow \textsc{Pilot}$

FD Satisfiability

- For attributes *x* and *y*, the FD satisfiability is defined by:

$$\langle \forall\, t, t' \;:\; t, t' \in R : \; t[x] = t'[x] \Rightarrow t[y] = t'[y] \,\rangle$$

- The inference rules are derived from the *Armstrong axioms.*

Recalling the Lectures

- *Kernel* and *Image* of relations:

$$\text{ker}\, R = R^\circ \cdot R$$
$$\text{img}\, R = R \cdot R^\circ$$

- Relation classification:

	Reflexive	*Coreflexive*
ker R	entire R	injective R
img R	surjective R	simple R

- Functions are *entire* and *simple* relations.

Pointfree FD

- Attributes are (projection) functions of *n*-ary tuples;
- *n*-ary relations are seen as coreflexives;
- Converting the satisfiability formula to PF:

$$\mathsf{img}\,(y \cdot [\![R]\!] \cdot x^{\circ}) \subseteq id$$

- Which actually defines $y \cdot [\![R]\!] \cdot x^{\circ}$ as simple.

Generalizing PF FD

- Generalizing to any relation and functions we will denote:

$$\pi_{g,f} R \stackrel{\text{def}}{=} g \cdot R \cdot f^\circ$$

- May be defined as a *Galois Connection*, benefiting from a variety of properties;

$$\pi_{g,f} R \subseteq S \equiv R \subseteq g^\circ \cdot S \cdot f$$

- Easier formula reasoning and manipulation.

Ordering by Injectivity

- Can be shown that FD depends on the "level of injectivity" of f and g;
- A new ordering on relations is defined:

$$R \leq S \equiv \ker S \subseteq \ker R$$

- FD can now be defined by: $f \stackrel{R}{\rightharpoonup} g \;\equiv\; g \leq f \cdot R^\circ$
- The order is very rich on properties, making this definition more amenable for calculation.

Proving the Armstrong Axioms

- Example: using the PF style, all Armstrong axioms are proven;
- For instance, *Augmentation*:

$$x \xrightarrow{T} y \Rightarrow xz \xrightarrow{T} yz$$

- Proofs are simpler than the original.

$$\begin{array}{l}
\quad xz \xrightarrow{T} yz \\
\equiv \\
\quad xz \xrightarrow{T} y \ \wedge \ xz \xrightarrow{T} z \\
\equiv \\
\quad xz \xrightarrow{T} y \\
\Leftarrow \\
\quad x \xrightarrow{T} y
\end{array}$$

Multi-valued Dependency

- More general dependency: multi-valued (MVD);
- If *(a,b,c)* and *(a,d,e)* in *R*, then *(a,b,e)* also in *R*:

$$
\begin{array}{ll}
\langle \forall\, t, t' \;:\; t, t' \in T : & t[x] = t'[x] \qquad\qquad \rangle \\
& \Downarrow \\
\langle \exists\, t'' \;:\; t'' \in T : & t[xy] = t''[xy] \,\wedge \qquad \rangle \\
& t''[S - xy] = t'[S - xy]
\end{array}
$$

- Represented in pointfree:

$$x \xrightarrow{R} y \stackrel{\mathrm{def}}{=} \mathsf{ker}\,(x \cdot R^{\circ}) \subseteq (\mathsf{ker}\,xy) \cdot R \cdot \mathsf{ker}\,\overline{xy}$$

Conclusions

- The FD theory becomes simpler and more general, and thus easier to be reasoned about;
- As such, proofs are easier to perform and understand;
- Generalizable to MVD;
- Does not extend or improve the FD theory (yet), only introduces a new way to reason about it.