Typed Embedding of a Relational Language in OCaml

Dmitrii Kosarev, Dmitrii Boulytchev

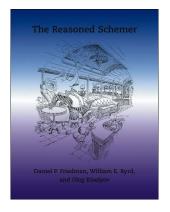
Saint-Petersburg State University JetBrains Research

> ML Family Workshop September 22, 2016 Nara, Japan

Relational Programming in miniKanren

From programs as *functions* to programs as *relations*:

$$f: X \to Y \leadsto f^o \subseteq X \times Y$$



- Daniel P. Friedman, William Byrd and Oleg Kiselyov. The Reasoned Schemer, The MIT Press, Cambridge, MA, 2005
- A DSL for Scheme/Racket with rather simple minimal implementation
- A family of languages (μKanren, α-Kanren, cKanren etc.)
- Implemented as DSL for a wide range of host languages (including OCaml, Haskell, Scala etc.)

```
\text{append: } \alpha \operatorname{list} \to \alpha \operatorname{list} \to \alpha \operatorname{list} \quad \operatorname{append}^o \subseteq \alpha \operatorname{list} \times \alpha \operatorname{list} \times \alpha \operatorname{list}
```

```
append: \alpha list \rightarrow \alpha list \rightarrow \alpha list let rec append rec append
```

```
append: \alpha list \rightarrow \alpha list \rightarrow \alpha list append \alpha list \alpha list
```

```
append: \alpha list \rightarrow \alpha list \rightarrow \alpha list append \alpha list \alpha list
```

```
append: \alpha list \rightarrow \alpha list \rightarrow \alpha list append \alpha \subseteq \alpha list \alpha
```

```
append: \alpha list \rightarrow \alpha list \rightarrow \alpha list append \alpha \subseteq \alpha list \alpha
```

```
append: \alpha list \rightarrow \alpha list \rightarrow \alpha list | append^o \subseteq \alpha list \times \alpha list \times \alpha list |

let rec append xs ys = 
match xs with | [] \rightarrow ys | h::tl \rightarrow h :: (append tl ys) | (xys \equiv h % tys) | (append^o t ys tys) | (append^o t ys tys) | (append^o t ys tys) | (xys \equiv h % tys) | (append^o t ys tys) |
```

```
append: \alpha list \rightarrow \alpha list \rightarrow \alpha list append \alpha \subseteq \alpha list \alpha list \alpha list append \alpha \subseteq \alpha list \alpha list
```

In the original miniKanren:

$$X = \{x_1, x_2, \dots\}$$

```
Logic variables X = \{x_1, x_2, \dots\} Symbols (constructors) S = \{s_1, s_2, \dots\}
```

Logic variables
$$X = \{x_1, x_2, \dots\}$$
 Symbols (constructors)
$$S = \{s_1, s_2, \dots\}$$
 Terms
$$T = X \cup \{s \ (t_1, \dots, t_k) \mid s \in S, \ t_i \in T\}$$

```
 \begin{array}{ll} \text{Logic variables} & X = \{x_1, x_2, \dots\} \\ \text{Symbols (constructors)} & S = \{s_1, s_2, \dots\} \\ \text{Terms} & T = X \cup \{s \ (t_1, \dots, t_k) \mid s \in S, \ t_i \in T\} \\ \text{Substitutions} & \Sigma = T^X \\ \end{array}
```

| Logic variables | $X = \{x_1, x_2, \dots\}$ |
|------------------------|--|
| Symbols (constructors) | $S = \{s_1, s_2, \dots\}$ |
| Terms | $T = X \cup \{s (t_1, \ldots, t_k) \mid s \in S, t_i \in T\}$ |
| Substitutions | $\Sigma = T^X$ |
| Unification | $(\equiv)\colon \Sigma \mathop{ ightarrow} T \to T \to \Sigma_\perp$ |

Jason Hemann, Daniel P. Friedman. μ Kanren: A Minimal Functional Core for Relational Programming // Scheme'13:

Logic variables $X = \{x_1, x_2, \dots\}$ Symbols (constructors) $S = \{s_1, s_2, \dots\}$ Terms $T = X \cup \{s \ (t_1, \dots, t_k) \mid s \in S, \ t_i \in T\}$ Substitutions $\Sigma = T^X$ Unification $(\equiv) : \Sigma \to T \to T \to \Sigma_\perp$ State (a substitution + some info to create fresh variables)

of states)

Jason Hemann, Daniel P. Friedman. μ Kanren: A Minimal Functional Core for Relational Programming // Scheme'13:

 $\begin{array}{lll} \mbox{Logic variables} & X = \{x_1, x_2, \dots\} \\ \mbox{Symbols (constructors)} & S = \{s_1, s_2, \dots\} \\ \mbox{Terms} & T = X \cup \{s \; (t_1, \dots, t_k) \; | \; s \in S, \; t_i \in T\} \\ \mbox{Substitutions} & \Sigma = T^X \\ \mbox{Unification} & (\equiv) \colon \Sigma \to T \to T \to \Sigma_\perp \\ \mbox{State (a substitution + some info to create fresh variables)} \\ \mbox{Goal (a function from a state to a stream} & g : \sigma \to \sigma \; \text{stream} \\ \end{array}$

| Logic variables | $X = \{x_1, x_2, \dots\}$ |
|---|---|
| Symbols (constructors) | $S = \{s_1, s_2, \dots\}$ |
| Terms | $T = X \cup \{s (t_1, \ldots, t_k) \mid s \in S, t_i \in T\}$ |
| Substitutions | $\Sigma = T^X$ |
| Unification | $(\equiv)\colon \Sigma 	o T 	o T 	o \Sigma_\perp$ |
| State (a substitution + some info to cre- | σ |
| ate fresh variables) | |
| Goal (a function from a state to a stream | $g: {f \sigma} ightarrow {f \sigma}$ stream |
| of states) | |
| Conjunction $g \wedge g$ | "bind" |

| Logic variables | $X = \{x_1, x_2, \dots\}$ |
|---|---|
| Symbols (constructors) | $S = \{s_1, s_2, \dots\}$ |
| Terms | $T = X \cup \{s (t_1, \ldots, t_k) \mid s \in S, t_i \in T\}$ |
| Substitutions | $\Sigma = T^X$ |
| Unification | $(\equiv)\colon \Sigma 	o T 	o T 	o \Sigma_\perp$ |
| State (a substitution + some info to cre- | σ |
| ate fresh variables) | |
| Goal (a function from a state to a stream | $g: {f \sigma} ightarrow {f \sigma}$ stream |
| of states) | |
| Conjunction $g \wedge g$ | "bind" |
| Disjunction $g \lor g$ | "mplus" |

Jason Hemann, Daniel P. Friedman. μKanren: A Minimal Functional Core for Relational Programming // Scheme'13:

Logic variables
Symbols (constructors)
Terms
Substitutions
Unification
State (a substitution + some info to create fresh variables)
Goal (a function from a state to a stream of states)
Conjunction $g \land g$ Disjunction $g \lor g$

Refinement of answers

$$X = \{x_1, x_2, \dots\}$$
 $S = \{s_1, s_2, \dots\}$
 $T = X \cup \{s \ (t_1, \dots, t_k) \mid s \in S, \ t_i \in T\}$
 $\Sigma = T^X$
 $(\equiv) \colon \Sigma \to T \to T \to \Sigma_{\perp}$
 σ
 $g \colon \sigma \to \sigma \text{ stream}$
"bind"
"mplus"
 $refine \colon \sigma \to X \to T$

Jason Hemann, Daniel P. Friedman. μKanren: A Minimal Functional Core for Relational Programming // Scheme'13:

```
Logic variables
                                                                X = \{x_1, x_2, \dots\}
                                                                S = \{s_1, s_2, \dots\}
Symbols (constructors)
                                                    T = X \cup \{s (t_1, ..., t_k) \mid s \in S, t_i \in T\}
Terms
                                                                   \nabla - T^X
Substitutions
                                                            (\equiv): \Sigma \to T \to T \to \Sigma_{\perp}
Unification
State (a substitution + some info to cre-
                                                                          σ
ate fresh variables)
Goal (a function from a state to a stream
                                                               g: \sigma \to \sigma stream
of states)
Conjunction g \wedge g
                                                                       "bind"
Disjunction g \lor g
                                                                      "mplus"
                                                              refine: \sigma \to X \to T
Refinement of answers
```

Unification and refinement are virtually the main things to implement

- Non-solution:
 - implement unification for a fixed term representation;
 - convert user-type data to- and from that universal representation.

- Non-solution:
 - implement unification for a fixed term representation;
 - convert user-type data to- and from that universal representation.

Does not help in detecting the mistakes statically

- Non-solution:
 - implement unification for a fixed term representation;
 - convert user-type data to- and from that universal representation.

Does not help in detecting the mistakes statically

 Bad solution: generate boilerplate unification code for each user type (there is no direct support for ad – hoc polymorphism in OCaml yet).

- Non-solution:
 - implement unification for a fixed term representation;
 - convert user-type data to- and from that universal representation.

Does not help in detecting the mistakes statically

• Bad solution: generate boilerplate unification code for each user type (there is no direct support for ad-hoc polymorphism in OCaml yet). Boilerplatish

- Non-solution:
 - implement unification for a fixed term representation;
 - convert user-type data to- and from that universal representation.

Does not help in detecting the mistakes statically

 Bad solution: generate boilerplate unification code for each user type (there is no direct support for ad – hoc polymorphism in OCaml yet).
 Boilerplatish

Users would need to adjust their types to represent logical variables

- Non-solution:
 - implement unification for a fixed term representation;
 - convert user-type data to- and from that universal representation.

Does not help in detecting the mistakes statically

• Bad solution: generate boilerplate unification code for each user type (there is no direct support for ad-hoc polymorphism in OCaml yet). Boilerplatish

Users would need to adjust their types to represent logical variables

Polymorphic unification:

$$\equiv : \Sigma \mathop{\rightarrow} \alpha \mathop{\rightarrow} \alpha \mathop{\rightarrow} \Sigma_{\bot}$$

- Non-solution:
 - implement unification for a fixed term representation;
 - convert user-type data to- and from that universal representation.

Does not help in detecting the mistakes statically

 Bad solution: generate boilerplate unification code for each user type (there is no direct support for ad – hoc polymorphism in OCaml yet).
 Boilerplatish

Users would need to adjust their types to represent logical variables

Polymorphic unification:

$$\equiv : \Sigma \rightarrow \alpha \rightarrow \alpha \rightarrow \Sigma_{\perp}$$

Has to be implemented in an untyped manner

- Non-solution:
 - implement unification for a fixed term representation;
 - convert user-type data to- and from that universal representation.

Does not help in detecting the mistakes statically

 Bad solution: generate boilerplate unification code for each user type (there is no direct support for ad – hoc polymorphism in OCaml yet).
 Boilerplatish

Users would need to adjust their types to represent logical variables

Polymorphic unification:

$$\equiv : \Sigma \rightarrow \alpha \rightarrow \alpha \rightarrow \Sigma_{\perp}$$

Has to be implemented in an untyped manner Might be a good solution (lightweight, efficient), if type safety is justified

Works for all *logic* types α^o :

$$\equiv : \Sigma \mathop{\rightarrow} \alpha^{\mathit{o}} \mathop{\rightarrow} \alpha^{\mathit{o}} \mathop{\rightarrow} \Sigma_{\bot}$$

Works for all *logic* types α^o :

$$\equiv : \Sigma \rightarrow \alpha^o \rightarrow \alpha^o \rightarrow \Sigma_{\perp}$$

Is implemented as the standard algorithm with triangular substitution and occurs check by traversing runtime representation, using unsafe interface Obj.

Works for all *logic* types α^o :

$$\equiv : \Sigma \rightarrow \alpha^o \rightarrow \alpha^o \rightarrow \Sigma_{\perp}$$

Is implemented as the standard algorithm with triangular substitution and occurs check by traversing runtime representation, using unsafe interface Obj.

Pitfalls:

 compiler loses the track of types after the results of unification are stored in a substitution → refinement has to be implemented untyped as well;

Works for all *logic* types α^o :

$$\equiv : \Sigma \rightarrow \alpha^o \rightarrow \alpha^o \rightarrow \Sigma_{\perp}$$

Is implemented as the standard algorithm with triangular substitution and occurs check by traversing runtime representation, using unsafe interface Obj.

Pitfalls:

- compiler loses the track of types after the results of unification are stored in a substitution → refinement has to be implemented untyped as well;
- the safety of unification/refinement implementation has to be justified separately;

Works for all *logic* types α^o :

$$\equiv : \Sigma \rightarrow \alpha^o \rightarrow \alpha^o \rightarrow \Sigma_{\perp}$$

Is implemented as the standard algorithm with triangular substitution and occurs check by traversing runtime representation, using unsafe interface Obj.

Pitfalls:

- compiler loses the track of types after the results of unification are stored in a substitution → refinement has to be implemented untyped as well;
- the safety of unification/refinement implementation has to be justified separately;
- states must not escape their scope (otherwise the coherence between variable types and terms, stored in states, can be lost).

Properties of Polymorphic Unification

It can be shown, that for our concrete implementation:

- variables in a substitution are always associated with the terms of the same type;
- all variables preserve their types, assigned by the compiler;
- all variables occur in terms only in a "type-safe" positions:

 $t[x] \iff$ the type of x corresponds to the type of the hole of t

Properties of Polymorphic Unification

It can be shown, that for our concrete implementation:

- variables in a substitution are always associated with the terms of the same type;
- all variables preserve their types, assigned by the compiler;
- all variables occur in terms only in a "type-safe" positions:

 $t[x] \iff$ the type of x corresponds to the type of the hole of t

 \sim

the refinement is type-safe, if a variable is refined in a state, which is an inheritor of the state that variable was created in.

States and refinement function are hidden and can not be accessed directly.

States and refinement function are hidden and can not be accessed directly.

The refinement is performed transparently as the top-level running primitive is invoked:

States and refinement function are hidden and can not be accessed directly.

The refinement is performed transparently as the top-level running primitive is invoked:

```
run \bar{n} (fun q_1 q_2 \dots q_n \to g) (fun a_1 a_2 \dots a_n \to h)
```

States and refinement function are hidden and can not be accessed directly.

The refinement is performed transparently as the top-level running primitive is invoked:

```
run \bar{n} (fun q_1 q_2 \dots q_n \rightarrow g) (fun a_1 a_2 \dots a_n \rightarrow h)
```

Here:

- run the only way to run goals;
- n̄ a numeral, describing the number of fresh variables, available for running the goal g; numerals can be manufactured quantum satis using the successor function, which is provided as well;
- $q_1, q_2 \dots q_n$ these fresh variables;
- a₁, a₂...a_n the streams of *refined* answers for the variables q₁, q₂...q_n respectively;
- h a handler, which can make use of refined answers.

The framework guarantees, that variables are refined only in correct states.

Injecting a user-type into logic domain and projecting the logical results back:

$$\uparrow_t : t \to t^o
\downarrow_t : t^o \to t$$

Injecting a user-type into logic domain and projecting the logical results back:

$$\uparrow_t : t \to t^o
\downarrow_t : t^o \to t$$

- \bullet "\$\frac{1}{\psi}\$", "\$\frac{1}{\psi}\$" are polymorphic shallow injection/projection;
- $\bullet\,$ for the deep case, make the type a functor and use ${\it fmap}.$

Injecting a user-type into logic domain and projecting the logical results back:

$$\uparrow_t : t \to t^o
\downarrow_t : t^o \to t$$

- " \uparrow_{\forall} ", " \downarrow_{\forall} " are polymorphic shallow injection/projection;
- for the deep case, make the type a functor and use *fmap*.

```
type tree = Leaf of int | Node of tree * tree
```

Injecting a user-type into logic domain and projecting the logical results back:

$$\uparrow_t : t \to t^o
\downarrow_t : t^o \to t$$

- " \uparrow_{\forall} ", " \downarrow_{\forall} " are polymorphic shallow injection/projection;
- for the deep case, make the type a functor and use *fmap*.

Injecting a user-type into logic domain and projecting the logical results back:

$$\uparrow_t : t \to t^o
\downarrow_t : t^o \to t$$

- \bullet "\$\dagger\$", "\$\dagger\$" are polymorphic shallow injection/projection;
- for the deep case, make the type a functor and use *fmap*.

```
type tree = Leaf of int | Node of tree * tree \rightsquigarrow type ('int, 'tree) tree<sub>f</sub> = Leaf of 'int | Node of 'tree * 'tree type tree = (int, tree) tree<sub>f</sub>
```

Injecting a user-type into logic domain and projecting the logical results back:

$$\uparrow_t : t \to t^o
\downarrow_t : t^o \to t$$

- " \uparrow_{\forall} ", " \downarrow_{\forall} " are polymorphic shallow injection/projection;
- for the deep case, make the type a functor and use *fmap*.

```
type tree = Leaf of int | Node of tree * tree \rightsquigarrow type ('int, 'tree) tree<sub>f</sub> = Leaf of 'int | Node of 'tree * 'tree type tree = (int, tree) tree<sub>f</sub> type tree<sup>o</sup> = ((int<sup>o</sup>, tree<sup>o</sup>) tree<sub>f</sub>)<sup>o</sup>
```

Injecting a user-type into logic domain and projecting the logical results back:

$$\uparrow_t : t \to t^o
\downarrow_t : t^o \to t$$

- " \uparrow_{\forall} ", " \downarrow_{\forall} " are polymorphic shallow injection/projection;
- for the deep case, make the type a functor and use *fmap*.

```
type tree = Leaf of int | Node of tree * tree \sim type ('int, 'tree) tree<sub>f</sub> = Leaf of 'int | Node of 'tree * 'tree type tree = (int, tree) tree<sub>f</sub> type tree<sup>o</sup> = ((int<sup>o</sup>, tree<sup>o</sup>) tree<sub>f</sub>)<sup>o</sup>

let rec (\uparrowtree) t = \uparrow_{\forall} ( fmaptree<sub>f</sub> (\uparrow_{\forall}) (\uparrowtree) t)
```

Injecting a user-type into logic domain and projecting the logical results back:

$$\uparrow_t : t \to t^o
\downarrow_t : t^o \to t$$

- "↑∀", "↓∀" are polymorphic shallow injection/projection;
- for the deep case, make the type a functor and use *fmap*.

```
type tree = Leaf of int | Node of tree * tree \sim

type ('int, 'tree) tree<sub>f</sub> = Leaf of 'int | Node of 'tree * 'tree type tree = (int, tree) tree<sub>f</sub>
type tree<sup>o</sup> = ((int<sup>o</sup>, tree<sup>o</sup>) tree<sub>f</sub>)<sup>o</sup>

let rec (\(\earrow\)tree) t = \(\earrow\)\(\psi\) (fmaptree<sub>f</sub> (\(\earrow\)\)\(\earrow\)tree) t)

let rec (\(\psi\)tree) 1 = fmaptree<sub>f</sub> (\(\psi\)\)\(\psi\) (\(\psi\)tree) (\(\psi\)\)
```

Example

.

Current Implementation

- Repository: https://github.com/dboulytchev/OCanren
- Implements μKanren + disequality constraints
- Passes most of the original tests
- Outperforms μ Kanren on long queries