

Унификация посредством поиска путей с контекстно-свободными ограничениями в графе

Source-tracking unification

Екатерина Вербицкая

Лаборатория языковых инструментов JetBrains

16 ноября 2020

Задачу унификации можно свести к
поиску путей с КС ограничениями в графе¹

Путь — доказательство (не)унифицируемости термов

¹Choppella, V., and Haynes, C. T. (2005). Source-tracking unification.

- Что такое унификация
- Как задача унификации представима в виде графа
- Какой язык будем использовать в качестве ограничений
- Почему это работает
- Какую дополнительную информацию можно получить из пути

Даны два терма t, s

Задача: найти подстановку на свободных переменных термов (унификатор) θ , такую что

$$t\theta = s\theta$$

Терм: $\mathcal{T} :: \mathcal{V} \mid \mathcal{F}^n \mathcal{T}_1 \dots \mathcal{T}_n$

Подстановка: $\theta :: \mathcal{V} \rightarrow \mathcal{T}$

Применение подстановки $t\{x_1 \mapsto t_1, \dots, x_k \mapsto t_k\}$:
одновременно заменить свободные переменные x_i терма t на t_i

$$(f\ x\ a\ (g\ z)\ y)\{x \mapsto h\ a\ y, z \mapsto y\} = f\ (h\ a\ y)\ a\ (g\ y)\ y$$

Применение унификации

```
apply :: (a -> b) -> a -> b  
apply f x = f x
```

```
f :: Int -> Int  
f x = x + 1
```

```
apply_f :: ?  
apply_f = apply f
```

Унифицируем $a \rightarrow b$ и $\text{Int} \rightarrow \text{Int}$, получаем $a == \text{Int}$, $b == \text{Int}$

```
apply_f :: Int -> Int
```

Простой алгоритм унификации

Будем искать подстановку как множество уравнений $\mathcal{E} = \{t_i = s_i\}$

- Упрощение термов: $(f\ t_1 \dots t_n = g\ s_1 \dots s_m) \in \mathcal{E}$
 - ▶ Если f, g — различные константы, то $\mathcal{E} = \perp$
 - ▶ Иначе заменяем уравнение в \mathcal{E} на множество $t_1 = s_1, \dots, t_n = s_n$
- Переориентация: $(t = x) \in \mathcal{E}$
 - ▶ Если t — терм, x — переменная, заменяем в \mathcal{E} уравнение на $x = t$
- Элиминация переменных: $(x = t) \in \mathcal{E}$, x входит в какое-то уравнение
 - ▶ Если x входит в t , $t \equiv x$, то удаляем уравнение из \mathcal{E}
 - ▶ Иначе, если x входит в t , то $\mathcal{E} = \perp$
 - ▶ Иначе, подставляем t вместо x во всех уравнениях в \mathcal{E}

Унификация: пример

$$\{node\ El\ T\ T = node\ 1\ (node\ 2\ emp\ emp)\ (node\ 2\ emp\ emp)\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, T = node\ 2\ emp\ emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, node\ 2\ emp\ emp = node\ 2\ emp\ emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, 2 = 2, emp = emp, emp = emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp\}$$

Унификация: пример

$$\{node\ El\ T\ T = node\ 1\ (node\ 2\ emp\ emp)\ (node\ 3\ emp\ emp)\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, T = node\ 3\ emp\ emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, node\ 2\ emp\ emp = node\ 3\ emp\ emp\}$$

$$\{El = 1, T = node\ 2\ emp\ emp, 2 = 3, emp = emp, emp = emp\}$$

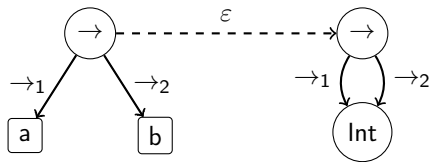
⊥

Чем плох простой алгоритм

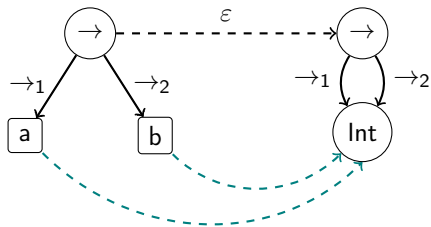
- Не очень эффективный
- Не говорит, почему унификация не завершилась успехом

Граф унификации

$$a \rightarrow b \stackrel{?}{=} \text{Int} \rightarrow \text{Int}$$

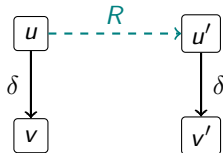


$$a \rightarrow b \stackrel{?}{=} \text{Int} \rightarrow \text{Int}$$



Отношение эквивалентности на вершинах

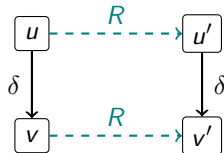
Отношение на вершинах R замкнуто вниз, если для любой метки на ребре δ и двух вершин в отношении uRu' с ребрами $u \xrightarrow{\delta} v$ и $u' \xrightarrow{\delta} v'$ верно vRv'



Замыкание унификации отношения R это наименьшее замкнутое вниз отношение на вершинах, содержащее R

Отношение эквивалентности на вершинах

Отношение на вершинах R замкнуто вниз, если для любой метки на ребре δ и двух вершин в отношении uRu' с ребрами $u \xrightarrow{\delta} v$ и $u' \xrightarrow{\delta} v'$ верно vRv'



Замыкание унификации отношения R это наименьшее замкнутое вниз отношение на вершинах, содержащее R

Факторграф унификации

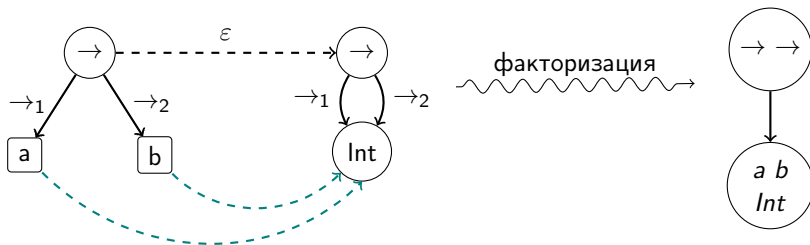
Вершины *равны*, если связаны ε -ребром

Факторизуем граф унификации по отношению эквивалентности на вершинах, которое построено как замыкание унификации отношения равенства вершин

Факторграф унификации

Вершины *равны*, если связаны ε -ребром

Факторизуем граф унификации по отношению эквивалентности на вершинах, которое построено как замыкание унификации отношения равенства вершин



Унификация невозможна, тогда и только тогда, когда в факторграфе есть цикл или вершина с разными функциональными символами

Большой пример

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$t_2 \stackrel{?}{=} t_4$$

$$t_3 \stackrel{?}{=} \textit{bool}$$

$$t_4 \stackrel{?}{=} t_5$$

$$t_3 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

$$t_5 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} \textit{int} \rightarrow \textit{int}$$

$$t_7 \stackrel{?}{=} t_1$$

Большой пример

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$t_2 \stackrel{?}{=} t_4$$

$$t_3 \stackrel{?}{=} \textit{bool}$$

$$t_4 \stackrel{?}{=} t_5$$

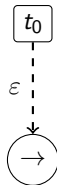
$$t_3 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

$$t_5 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} \textit{int} \rightarrow \textit{int}$$

$$t_7 \stackrel{?}{=} t_1$$



Большой пример

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$t_2 \stackrel{?}{=} t_4$$

$$t_3 \stackrel{?}{=} \text{bool}$$

$$t_4 \stackrel{?}{=} t_5$$

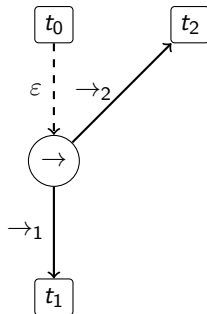
$$t_3 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

$$t_5 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} \text{int} \rightarrow \text{int}$$

$$t_7 \stackrel{?}{=} t_1$$



Большой пример

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$t_2 \stackrel{?}{=} t_4$$

$$t_3 \stackrel{?}{=} \text{bool}$$

$$t_4 \stackrel{?}{=} t_5$$

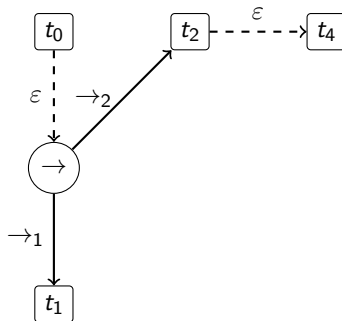
$$t_3 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

$$t_5 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} \text{int} \rightarrow \text{int}$$

$$t_7 \stackrel{?}{=} t_1$$



Большой пример

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$t_2 \stackrel{?}{=} t_4$$

$$t_3 \stackrel{?}{=} \text{bool}$$

$$t_4 \stackrel{?}{=} t_5$$

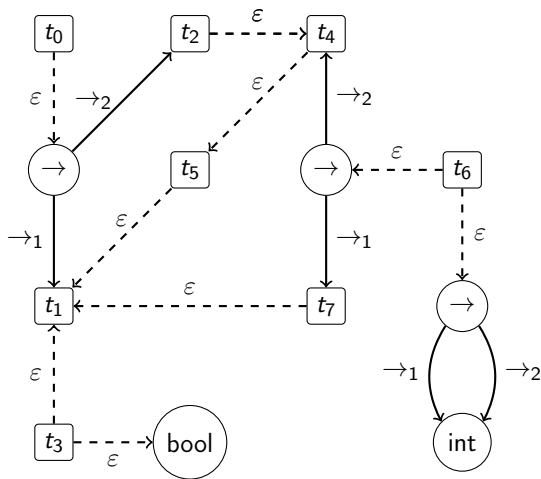
$$t_3 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

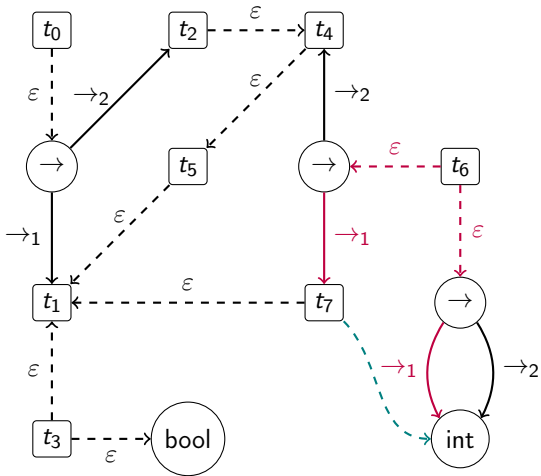
$$t_5 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} \text{int} \rightarrow \text{int}$$

$$t_7 \stackrel{?}{=} t_1$$



Большой пример: $t_7 = int$

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$
$$t_2 \stackrel{?}{=} t_4$$
$$t_3 \stackrel{?}{=} \text{bool}$$
$$t_4 \stackrel{?}{=} t_5$$
$$t_3 \stackrel{?}{=} t_1$$
$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$
$$t_5 \stackrel{?}{=} t_1$$
$$t_6 \stackrel{?}{=} int \rightarrow int$$
$$t_7 \stackrel{?}{=} t_1$$


Большой пример: $t_4 = int$

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$t_2 \stackrel{?}{=} t_4$$

$$t_3 \stackrel{?}{=} bool$$

$$t_4 \stackrel{?}{=} t_5$$

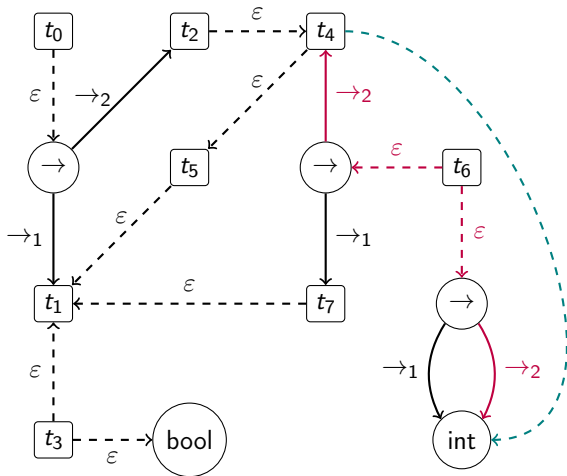
$$t_3 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

$$t_5 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} int \rightarrow int$$

$$t_7 \stackrel{?}{=} t_1$$



Большой пример: $t_7 = \text{bool}$

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$t_2 \stackrel{?}{=} t_4$$

$$t_3 \stackrel{?}{=} \text{bool}$$

$$t_4 \stackrel{?}{=} t_5$$

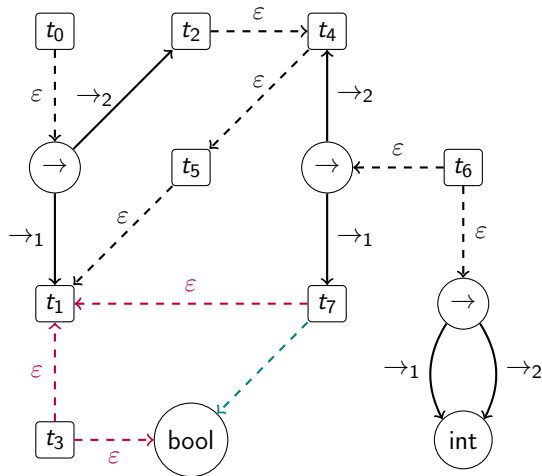
$$t_3 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

$$t_5 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} \text{int} \rightarrow \text{int}$$

$$t_7 \stackrel{?}{=} t_1$$



Большой пример: $t_4 = \text{bool}$

$$t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$t_2 \stackrel{?}{=} t_4$$

$$t_3 \stackrel{?}{=} \text{bool}$$

$$t_4 \stackrel{?}{=} t_5$$

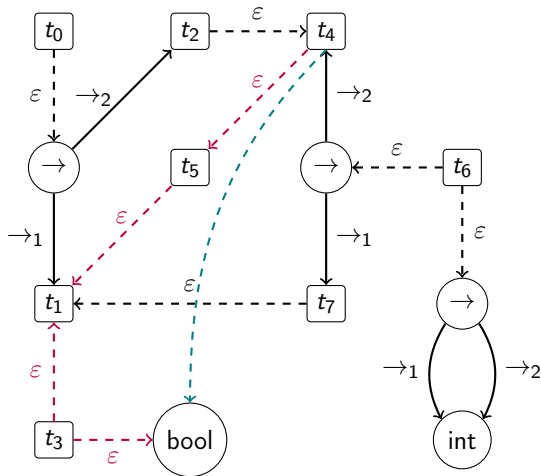
$$t_3 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

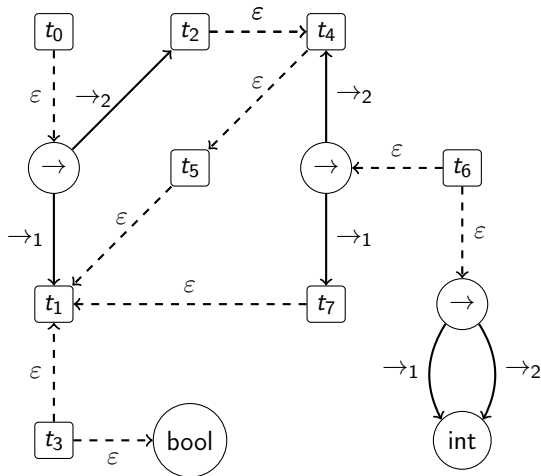
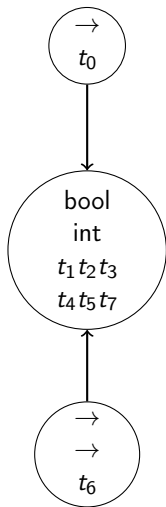
$$t_5 \stackrel{?}{=} t_1$$

$$t_6 \stackrel{?}{=} \text{int} \rightarrow \text{int}$$

$$t_7 \stackrel{?}{=} t_1$$



Факторграф для примера: унификация невозможна



$$\Sigma = \{\delta_1, \dots, \delta_n\}$$

$$\Sigma^{-1} = \{\delta_1^{-1}, \dots, \delta_n^{-1}\}$$

$$\Sigma \cap \Sigma^{-1} = \emptyset$$

$$\mathbf{S}(\Sigma) = \{\delta^{-1}\delta \approx \varepsilon \mid \delta \in \Sigma\}$$

$u \stackrel{\mathbf{S}(\Sigma)}{\approx} v$, если u можно получить из v применением равенств из $\mathbf{S}(\Sigma)$

$\mu_{\mathbf{S}(\Sigma)}(l)$ — нормальная форма l относительно равенств из $\mathbf{S}(\Sigma)$

$$S(\Sigma, L) \stackrel{\text{def}}{=} \{l \in (\Sigma \cup \Sigma^{-1})^* \mid \mu_{\mathbf{S}(\Sigma)}(l) \in L\}$$

$$S^0(\Sigma) = S(\Sigma, \varepsilon)$$

$$S^+(\Sigma) = S(\Sigma, \Sigma^+)$$

$$S^*(\Sigma) = S(\Sigma, \Sigma^*)$$

$S^0(\Sigma)$ — полудиков язык (semi-Dyck set)

²Определение полудикова языка может показаться неожиданным

Примеры строк полудикова языка

$$\Sigma = \{ (, [\}$$
$$\Sigma^{-1} = \{),] \}$$

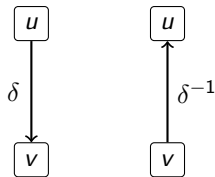
$$\varepsilon,))((,]][[,)[((\in S^0(\Sigma)$$

$$(,)) (,])([\notin S^0 \Sigma$$

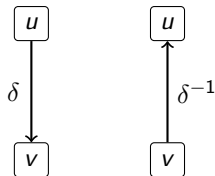
$$(,)[(,)(([\in S^*(\Sigma) \setminus S^0(\Sigma)$$

$$),)() \notin S^*(\Sigma)$$

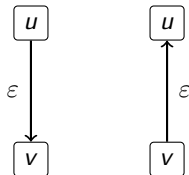
Пути унификации: обращение ребер



Пути унификации: обращение ребер



$$\varepsilon^{-1} = \varepsilon$$



Путь унификации — такой путь p в графе $G \cup G^{-1}$, что нормальная форма его метки $\mu_{S(\Sigma)}(l(p)) \in \Sigma^*$

G — граф унификации, G^{-1} — граф с обращенными ребрами

Путь унификации: пример

$$\mu_{\mathbf{S}(\Sigma)}(l(p^{-1}f^{-1}hr)) = \mu_{\mathbf{S}(\Sigma)}((\rightarrow_1)^{-1} \rightarrow_1) = \varepsilon \in \Sigma^*$$

$$a : t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$b : t_2 \stackrel{?}{=} t_4$$

$$c : t_3 \stackrel{?}{=} \text{bool}$$

$$d : t_4 \stackrel{?}{=} t_5$$

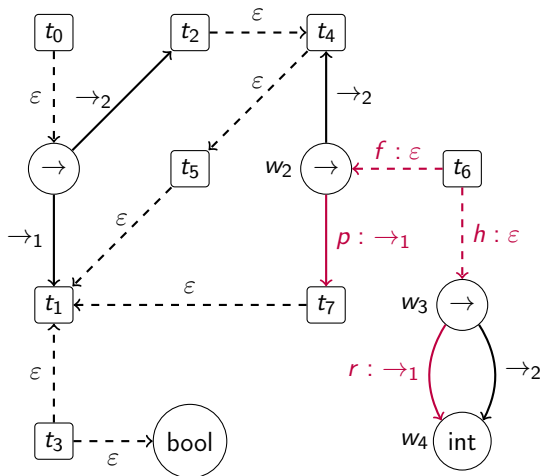
$$e : t_3 \stackrel{?}{=} t_1$$

$$f : t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

$$g : t_5 \stackrel{?}{=} t_1$$

$$h : t_6 \stackrel{?}{=} \text{int} \rightarrow \text{int}$$

$$i : t_7 \stackrel{?}{=} t_1$$



Путь унификации: пример

$$\mu_{\mathbf{S}(\Sigma)}(l(s^{-1}f^{-1}hq)) = \mu_{\mathbf{S}(\Sigma)}((\rightarrow_2)^{-1} \rightarrow_2) = \varepsilon \in \Sigma^*$$

$$a : t_0 \stackrel{?}{=} t_1 \rightarrow t_2$$

$$b : t_2 \stackrel{?}{=} t_4$$

$$c : t_3 \stackrel{?}{=} \text{bool}$$

$$d : t_4 \stackrel{?}{=} t_5$$

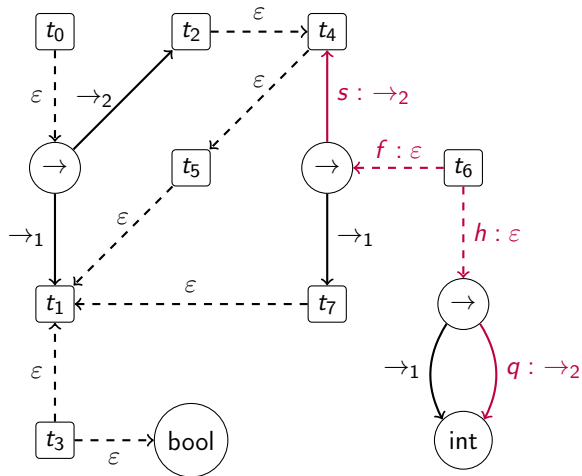
$$e : t_3 \stackrel{?}{=} t_1$$

$$f : t_6 \stackrel{?}{=} t_7 \rightarrow t_4$$

$$g : t_5 \stackrel{?}{=} t_1$$

$$h : t_6 \stackrel{?}{=} \text{int} \rightarrow \text{int}$$

$$i : t_7 \stackrel{?}{=} t_1$$



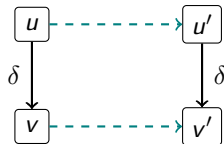
Путь унификации в факторграфе

Theorem

Если в графе унификации существует путь p из вершины u в вершину v , то в факторграфе существует путь из класса эквивалентности $[u]$ в класс эквивалентности $[v]$ с меткой $\mu_{S(\Sigma)}(l(p))$

Theorem

Если в факторграфе унификации существует путь p' из класса эквивалентности $[u]$ в класс эквивалентности $[v]$, то в графе унификации существует путь p из u в вершину v с меткой, чья нормальная форма $\mu_{S(\Sigma)}(l(p)) = p'$



Критерий унифицируемости термов

Два терма унифицируются тогда и только тогда, когда между ними есть путь унификации, нормальная форма которого — ε

Два терма унифицируются тогда и только тогда, когда между ними есть путь унификации, метка которого в полудиковом языке $S^0(\Sigma)$

$$\frac{}{G \vdash c : u \xrightarrow{\eta} v}, c : u \xrightarrow{\eta} v \in G$$

$$\frac{}{G \vdash \varepsilon : u \xrightarrow{\varepsilon} v}, u \in G$$

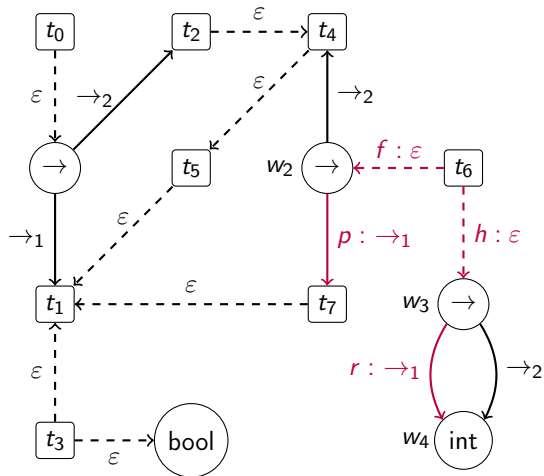
$$\frac{G \vdash p : v \xrightarrow{\varepsilon} u}{G \vdash p^{-1} : u \xrightarrow{\varepsilon} v}$$

$$\frac{G \vdash p : u \xrightarrow{l} v' \quad G \vdash q : v' \xrightarrow{l''} u}{G \vdash pq : u \xrightarrow{l''} v}$$

$$\frac{G \vdash p : u' \xrightarrow{\varepsilon} v'}{G \vdash b^{-1}pb' : u \xrightarrow{\varepsilon} v}, b : u' \xrightarrow{\delta} u \in G, b' : v' \xrightarrow{\delta} v \in G$$

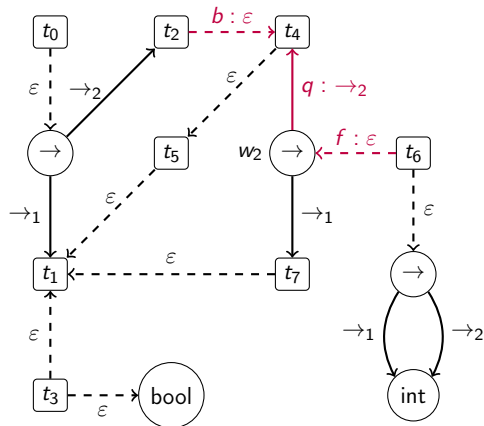
Вывод успешного пути

$$\begin{array}{c}
 \frac{G \vdash f : t_6 \xrightarrow{\varepsilon} w_2}{G \vdash f^{-1} : w_2 \xrightarrow{\varepsilon} t_6} \quad G \vdash h : t_6 \xrightarrow{\varepsilon} w_3 \\
 \hline
 G \vdash f^{-1}h : w_2 \xrightarrow{\varepsilon} w_3 \\
 \hline
 G \vdash p^{-1}f^{-1}hr : t_7 \xrightarrow{\varepsilon} w_4
 \end{array}$$



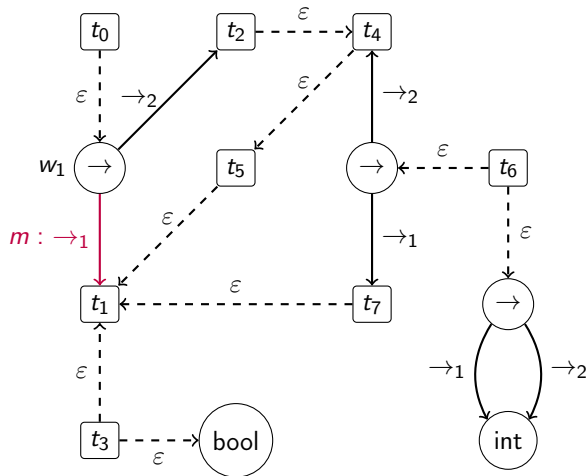
Вывод успешного пути

$$\begin{array}{c}
 \frac{G \vdash f : t_6 \xrightarrow{\varepsilon} w_2 \quad \frac{G \vdash q : w_2 \xrightarrow{2} t_4 \quad G \vdash b^{-1} : t_4 \xrightarrow{\varepsilon} t_2}{G \vdash qb^{-1} : w_2 \xrightarrow{2} t_2}}{G \vdash fqb^{-1} : t_6 \xrightarrow{2} t_2}
 \end{array}$$



Вывод неуспешного пути

$$\begin{array}{c}
 \frac{G \vdash m : w_1 \xrightarrow{\rightarrow_1} t_1}{G \vdash m : w_1 \xrightarrow{\rightarrow_1} t_1} \quad \frac{G \vdash m : w_1 \xrightarrow{\rightarrow_1} t_1}{G \vdash m : w_1 \xrightarrow{\rightarrow_1} t_1} \\
 \frac{G \vdash m : w_1 \xrightarrow{\rightarrow_1} t_1 \quad G \vdash m^{-1} : t_1 \xrightarrow{\rightarrow_1^{-1}} w_1}{G \vdash mm^{-1} : w_1 \xrightarrow{\rightarrow_1 \rightarrow_1^{-1}} w_1}
 \end{array}$$



Алгоритмы унификации через поиск путей

- Один алгоритм ищет кратчайшие пути, работает за $O(n^3)$
- Второй алгоритм интегрируется в стандартный алгоритм унификации, работает на константу хуже и строит не самые маленькие доказательства унифицируемости (но и не самые плохие)