Implication and Axiomatization of Functional Constraints

Jelle Hellings[1], Marc Gyssens[1], Jan Paredaens[2], Yuqing Wu[3]

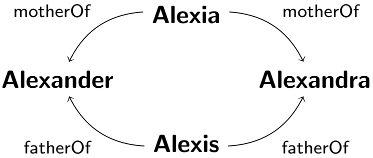
[1]Hasselt University and Transnational University of Limburg
[2]University of Antwerp
[3]Indiana University

Overview

Functional Constraints in RDF (Akhtar et al.)

RDF data

Alexia	motherOf	Alexandra
Alexia	motherOf	Alexander
Alexis	fatherOf	Alexandra
Alexis	fatherOf	Alexander



Functional Constraint

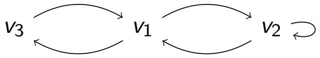
$$(P, L \rightarrow R)$$

Example

- $(\{(\$p, \texttt{motherOf}, \$c)\}, \$c \rightarrow \$p)$
- $(\{(\$p, \$t, \$c)\}, \$c \rightarrow \$p)$

Functional Constraints on structural data

Graph data



Example

- Conditions on self-loops: $(\{(\$n, \$n)\}, \emptyset \rightarrow \$n)$
- Conditions on cycles: $(\{(\$n, \$m), (\$m, \$n)\}, \$n \rightarrow \$m)$

Overview

Patterns and relations

Definition (patterns and relations)

A *n-ary pattern* is a set of *n-ary tuples* over *terms*. Thereby, terms are *constants* from $\mathcal{U}$ or *variables* from $\mathcal{V}$. A *n-ary relation* is a variable-free pattern.

Example

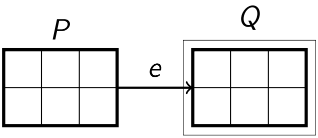
$$P = \{(\$p, motherOf, \$c)\}, \quad \mathcal{V}_P = \{\$p, \$c\}$$

$$\mathcal{R} = \begin{array}{llll} \{ & (\text{Alexia}, & \text{motherOf}, & \text{Alexandra}), \\ & (\text{Alexia}, & \text{motherOf}, & \text{Alexander}), \\ & (\text{Alexis}, & \text{fatherOf}, & \text{Alexandra}), \\ & (\text{Alexis}, & \text{fatherOf}, & \text{Alexander}) \quad \} \end{array}, \quad \mathcal{V}_{\mathcal{R}} = \emptyset$$

Embeddings

Definition

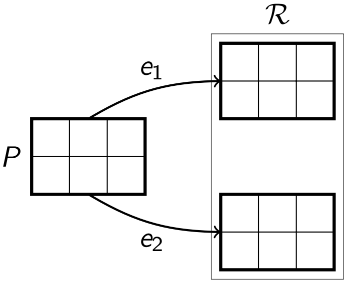
A function $e : \mathcal{V}_P \cup \mathcal{U} \to \mathcal{V} \cup \mathcal{U}$ is an embedding of pattern P into pattern Q if $e|_\mathcal{U} = \mathrm{id}_\mathcal{U}$ and $e(P) \subseteq Q$.



Functional Constraints

Definition (notation)

A *functional constraint* is a pair $(P, L \to R)$, where P is a nonempty pattern and $L, R \subseteq \mathcal{V}_P$.

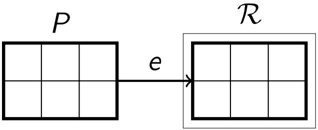


if $e_1 =_L e_2$ then $e_1 =_R e_2$

Equality-generating constraint

Definition (notation)

An *equality-generating constraint* is a pair (P, E), where P is a nonempty pattern and E is a set of equalities of the form $t_1 = t_2$ with $t_1, t_2 \in \mathcal{V}_P \cup \mathcal{U}$.



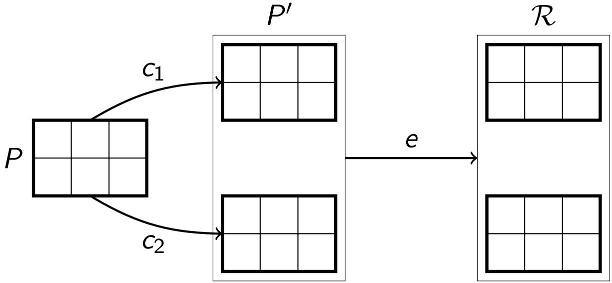
$$e(t_1) = e(t_2) \text{ for all } t_1 = t_2 \in E$$

Example

$$(\{(\$p, \$t, \$c)\}, \$t = \text{`motherOf'})$$

Functional constraints and equality-generating constraints

Let $(P, L \to R)$ be a functional constraint



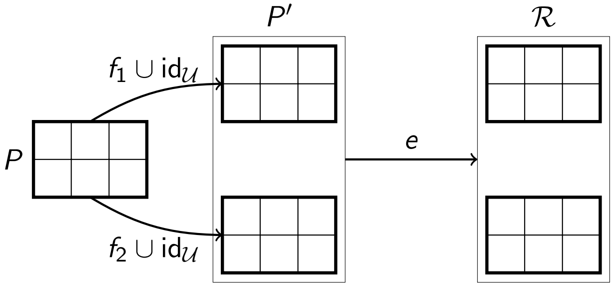
Idea

If $c_1(P)$ and $c_2(P)$ only agree on L,
Then express 'agree on R' by equalities between $c_1(P), c_2(P)$

Functional constraints and equality-generating constraints

Let $f_1, f_2 : \mathcal{V}_P \to \mathcal{V}$ be injections with:

$$f_1 =_L f_2 \text{ and } \mathrm{range}(f_1|_{\mathcal{V}_P \setminus L}) \cap \mathrm{range}(f_2|_{\mathcal{V}_P \setminus L}) = \emptyset$$



Proposition

$(P, L \to R)$ is equivalent to

$$((f_1 \cup id_{\mathcal{U}})(P) \cup (f_2 \cup id_{\mathcal{U}})(P), \{f_1(\$r) = f_2(\$r) \mid \$r \in R\})$$

Functional Dependencies

Proposition (reflexivity)

Let P be a pattern. If $R \subseteq L \subseteq \mathcal{V}_P$, then $(P, L \to R)$.

Proposition (augmentation)

If $(P, L \to R)$ and $V \subseteq \mathcal{V}_P$, then $(P, L \cup V \to R \cup V)$.

Proposition (transitivity)

If $(P, V_1 \to V_2)$ and $(P, V_2 \to V_3)$, then $(P, V_1 \to V_3)$.

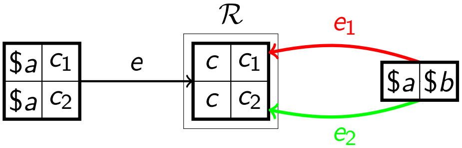
Lemma

$(P, L \to R)$ is equivalent to $\{(P, L \to \$r) \mid \$r \in R\}$

Incomplete axiomatization: Inconsistencies

Example

Does $(\{(\$a, \$b)\}, \$a \rightarrow \$b)$ imply $(\{(\$a, c_1), (\$a, c_2)\}, \emptyset \rightarrow \$a)$?



if embedding e exists, then $c_1 = e_1(\$b) \neq e_2(\$b) = c_2$

Overview

Chase for equality-generating constraints

Chase($\mathcal{C}$, (P, E)): answers $\mathcal{C} \models (P, E)$?

$\mathfrak{T} \leftarrow P$

while $\exists\ (P', E') \in \mathcal{C}$, embedding e of P' into $\mathfrak{T}$ with
  $\exists\ \{t_1 = t_2\} \in E'$ and $e(t_1) \neq e(t_2)$ **do**

  if $e(t_1), e(t_2)$ are both constants **then**

    return true

  else

    apply $(P', \{t_1 = t_2\})$ on $\mathfrak{T}$ by replacing $e(t_1)$ by $e(t_2)$
    (taking constants into account)

  end if

end while

return $\mathfrak{T} \mathrel{|}\!\equiv (P, E)$

Chase for functional constraints

Chase-FC$(\mathcal{C}_{\mathrm{FC}}, (P, L \to \$r))$: answers $\mathcal{C}_{\mathrm{FC}} \models (P, L \to \$r)$?

let $f_1, f_2 : \mathcal{V}_P \to \mathcal{V}$ be injections with $f_1 =_L f_2$ and
$\mathrm{range}(f_1|_{\mathcal{V}_P \setminus L}) \cap \mathrm{range}(f_2|_{\mathcal{V}_P \setminus L}) = \emptyset$

$\mathfrak{T} \leftarrow (f_1 \cup \mathrm{id}_{\mathcal{U}})(P) \cup (f_2 \cup \mathrm{id}_{\mathcal{U}})(P)$

while $\exists\ (P', L' \to \$r') \in \mathcal{C}_{\mathrm{FC}}$, embeddings e_1, e_2 of P' into $\mathfrak{T}$ with $e_1 =_L e_2$ and $e_1(\$r') \neq e_2(\$r')$ **do**

  if $e(\$r'), e(\$r')$ are both constants **then**

    return true

  else

    apply $(P', L' \to \$r')$ on $\mathfrak{T}$ by replacing $e_1(\$r_1)$ by $e_2(\$r_2)$ (taking constants into account)

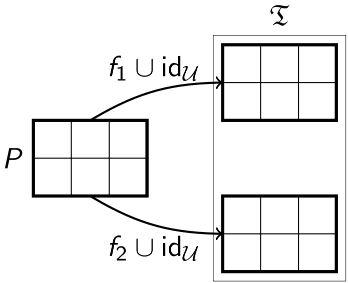
  end if

end while

return $\mathfrak{T} \models (P, L \to \$r)$

Initialization of **Chase-FC**

We chase for $(P, L \to \$r)$:

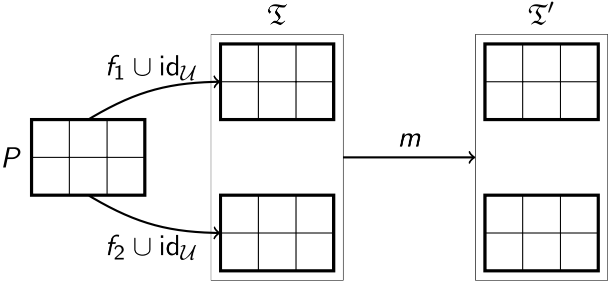


Definition (state of tableau)

$$\mathbf{S}_{\mathfrak{T}}(P, L, f_1, f_2)$$

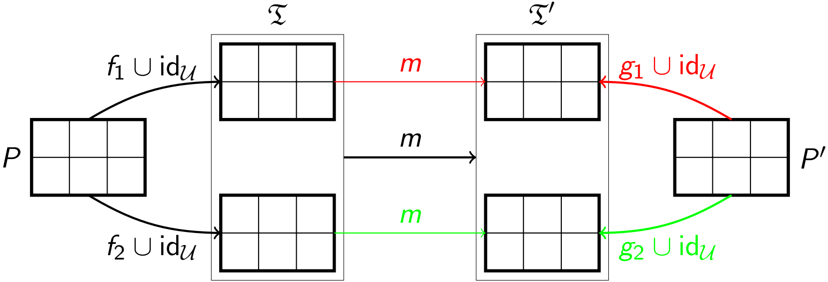
Symmetry preservation

Sequence of chase steps on $\mathfrak{T}$ resulting in $\mathfrak{T}' = m(\mathfrak{T})$.



Symmetry preservation

Sequence of chase steps on $\mathfrak{T}$ resulting in $\mathfrak{T}' = m(\mathfrak{T})$.



Definition (symmetry-preserving)

Mapping m is *symmetry-preserving* if there exists a tableau state $\mathbf{S}_{\mathfrak{T}'}(P', L', g_1, g_2)$ such that (for $i \in \{1, 2\}$):

$$m((f_i \cup \mathrm{id}_{\mathcal{U}})(P)) = (g_i \cup \mathrm{id}_{\mathcal{U}})(P').$$

Chasing can be symmetry-preserving

Theorem

Consider the chase for $\mathcal{C}_{\mathrm{FC}} \models C$ with $C = (P, L \rightarrow \$r)$. If initial a chase step is possible with $C_i = (P', L' \rightarrow \$r')$, then a symmetry preserving sequence of at most two chase steps with C_i is possible.

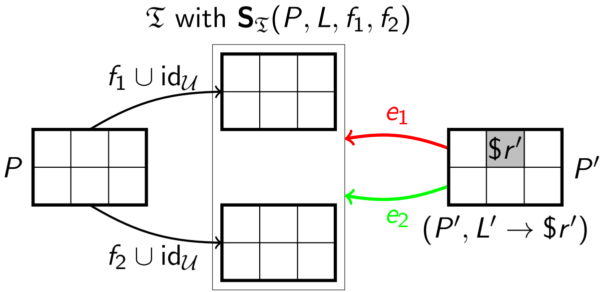
Step to axiomatization

Initially we have $\mathbf{S}_{\mathfrak{T}}(P, L, f_1, f_2)$, after chase step(s) with C_i we have $\mathbf{S}_{m(\mathfrak{T})}(P', L', f_1', f_2')$. We shall prove that there exists a functional constraint $C' = (P', L' \rightarrow \$r')$ with:

1. $\{C_i, C'\} \models C$
2. $C \models C'$

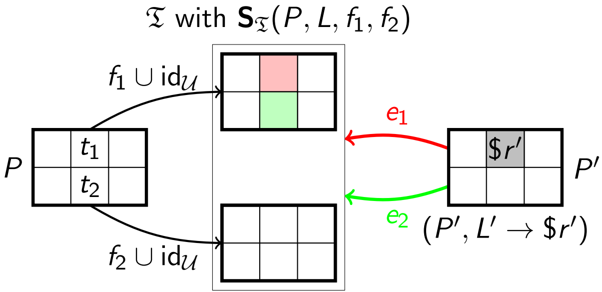
Initial steps of the chase algorithm

- $\mathcal{C}_{\text{FC}} \models (P, L \to \$r)$?
- First chase step with $(P', L' \to \$r')$ and embeddings e_1, e_2
- Initially we have $\mathbf{S}_{\mathfrak{T}}(P, L, f_1, f_2)$



Perform case distinction on $e_1(\$r')$ and $e_2(\$r')$.

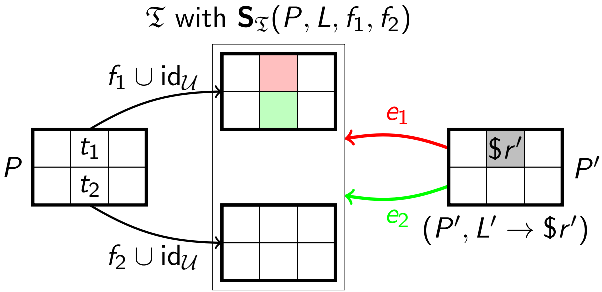
Initial steps of the chase algorithm: case 1a



Case $f_1(t_1) = e_1(\$r')$, $f_1(t_2) = e_2(\$r')$, $t_1 \in L \cup \mathcal{U}$, $t_2 \in L$:

- Left-hand sides coincide: symmetry preserved

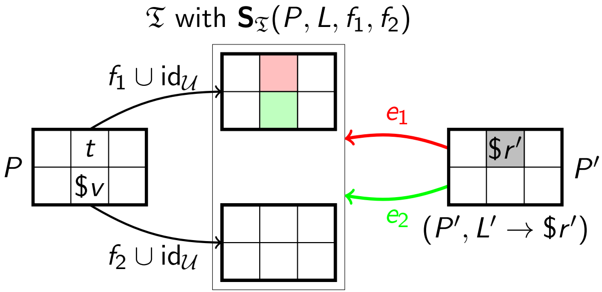
Initial steps of the chase algorithm: case 1b



Case $f_1(t_1) = e_1(\$r')$, $f_2(t_2) = e_2(\$r')$, $t_1 \in \mathcal{V} \cup \mathcal{U}$, $t_2 \in \mathcal{V} \setminus L$:

- $\varepsilon_i = \Phi_{f_2 \hookleftarrow f_1} \circ e_i, i \in \{1, 2\}$ embed P' into $\mathfrak{T}$
- After chase step: $\varepsilon_1 =_{L'} \varepsilon_2$ and $\varepsilon_1(\$r') \neq \varepsilon_2(\$r')$
- Changes in $(f_1 \cup \mathrm{id}_{\mathcal{U}})(P)$ mimicked by $\varepsilon_1, \varepsilon_2$ in $(f_2 \cup \mathrm{id}_{\mathcal{U}})(P)$

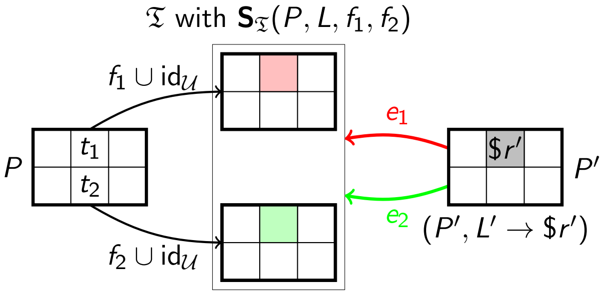
Initial steps of the chase algorithm: case 1a, 1b



Afterwards

$\mathbf{S}_{\mathfrak{T}}\left(\phi_{t \hookleftarrow \$v}(P), L \setminus \{\$v\}, f_1|_{\mathcal{V}_P \setminus \{\$v\}}, f_2|_{\mathcal{V}_P \setminus \{\$v\}}\right)$

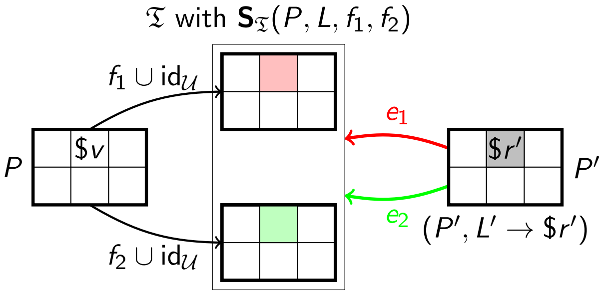
Initial steps of the chase algorithm: case 2



Case $f_1(t_1) = e_1(\$r')$, $f_1(t_1) = e_2(\$r')$:

- We must have: $t_1 \in \mathcal{V}_P$ and $t_1 \notin L$
- Change: f_1 and f_2 agree on t_1
- t_1 added to L

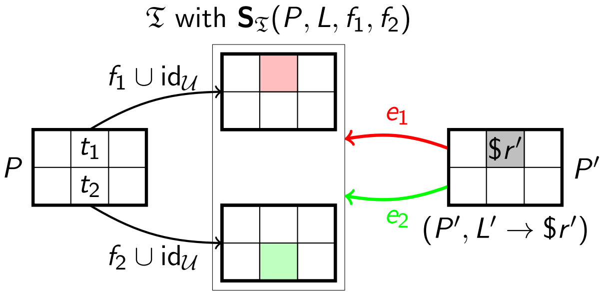
Initial steps of the chase algorithm: case 2



Afterwards

$\mathbf{S}_{\mathfrak{T}}\left(P, L \cup \{\$v\}, \phi_{f_1(\$v) \hookleftarrow f_2(\$v)} \circ f_1, \phi_{f_1(\$v) \hookleftarrow f_2(\$v)} \circ f_2\right)$

Initial steps of the chase algorithm: case 3



Case $f_1(t_1) = e_1(\$r')$, $f_2(t_2) = e_2(\$r')$:

- $\varepsilon_i = \Phi_{f_1 \hookleftarrow f_2} \circ e_i, i \in \{1, 2\}$ embed P' into $\mathfrak{T}$
- We have $(f_1 \cup \mathrm{id}_{\mathcal{U}})(t_1) = \varepsilon_1(\$r')$ and $(f_1 \cup \mathrm{id}_{\mathcal{U}})(t_2) = \varepsilon_2(\$r')$
- Reduces to cases 1a and 1b

Symmetry-preserving chase for functional constraints

S-Chase-FC($\mathcal{C}_{\mathrm{FC}}$, $(P, L \to \$r)$): answers $\mathcal{C}_{\mathrm{FC}} \models (P, L \to \$r)$?

let $f_1, f_2 : \mathcal{V}_P \to \mathfrak{V}$ be injections with $f_1 =_L f_2$ and
$\mathrm{range}(f_1|_{\mathcal{V}_P \setminus L}) \cap \mathrm{range}(f_2|_{\mathcal{V}_P \setminus L}) = \emptyset$
$\mathfrak{T} \leftarrow (f_1 \cup \mathrm{id}_{\mathcal{U}})(P) \cup (f_2 \cup \mathrm{id}_{\mathcal{U}})(P)$
while $\exists$ $(P', L' \to \$r') \in \mathcal{C}_{\mathrm{FC}}$, embeddings e_1, e_2 of P' into
$\mathfrak{T}$ with $e_1 =_L e_2$, $e_1(\$r') \neq e_2(\$r')$, $e_2(\$r') \notin \mathcal{U}$ **do**
    perform the corresponding symmetry-preserving step
end while
if *inconsistency termination* **then**
    return TRUE
else
    return $\mathfrak{T} \models (P, L \to \$r)$
end if

Overview

Axiomatization for symmetry-preserving steps

Complete Axiomatization

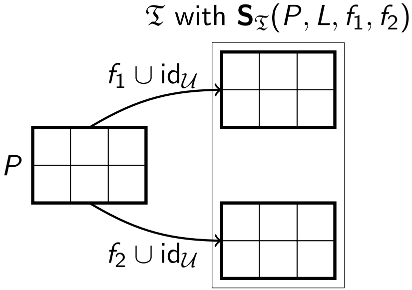
If $\mathcal{C}_{\mathrm{FC}} \models (P, L \to \$r)$ then $\mathcal{C}_{\mathrm{FC}} \vdash (P, L \to \$r)$

Proof strategy

Simulate each possible initial step of symmetry-preserving chases

- Direct termination via $\mathfrak{T} \models\!\!\!\equiv (P, L \to \$r)$
- Direct termination via inconsistency
- A symmetry-preserving step is performed

Direct termination via $\mathfrak{T} \models (P, L \to \$r)$

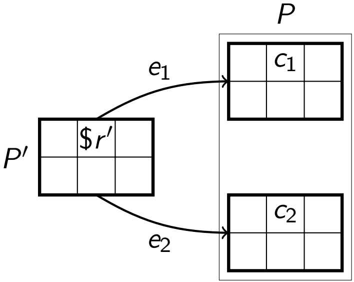


No changes made to $\mathfrak{T}$, hence $\$r \in L$, hence reflexivity

Direct termination via inconsistency

Proposition (Inconsistency)

If $(P', L' \to \$r')$, *if there exist two embeddings of* P' *into a pattern* P *which agree on* $L' \in \mathcal{V}_{P'}$ *and map* $\$r'$ *to different constants of* $\mathcal{U}$, *and if* $\$r \in \mathcal{V}_P$, *then* $(P, L \to \$r)$.



A symmetry-preserving step is performed

Theorem

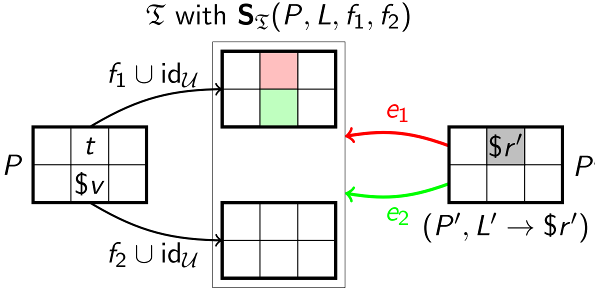
Consider the chase for $\mathcal{C}_{\mathrm{FC}} \models C$ with $C = (P, L \to \$r)$. If initial a chase step is possible with $C_i = (P', L' \to \$r')$, then a symmetry preserving sequence of at most two chase steps with C_i is possible.

Induction on symmetry-preserving steps

Initially we have $\mathbf{S}_{\mathfrak{I}}(P, L, f_1, f_2)$, after chase step(s) with C_i we have $\mathbf{S}_{m(\mathfrak{I})}(P', L', f_1', f_2')$. We shall prove that there exists a functional constraint $C' = (P', L' \to \$r')$ with:

1. $\{C_i, C'\} \vdash C$ (which implies $\{C_i, C'\} \models C$)
2. $C \models C'$ (which implies $\mathcal{C}_{\mathrm{FC}} \models C'$)

A symmetry-preserving step is performed: case 1a, 1b



Afterwards

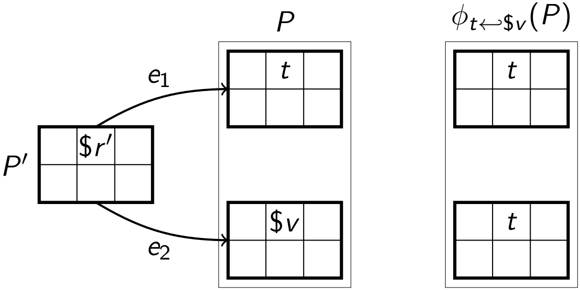
$\mathbf{S}_{\mathfrak{T}}\left(\phi_{t \hookleftarrow \$v}(P), L \setminus \{\$v\}, f_1|_{\mathcal{V}_P \setminus \{\$v\}}, f_2|_{\mathcal{V}_P \setminus \{\$v\}}\right)$

$$C' = (\phi_{t_1 \hookleftarrow \$v_2}(P), \phi_{t_1 \hookleftarrow \$v_2}(L) \cap \mathcal{V}_P \to \{\phi_{t_1 \hookleftarrow \$v_2}(\$r)\} \cap \mathcal{V}_P)$$

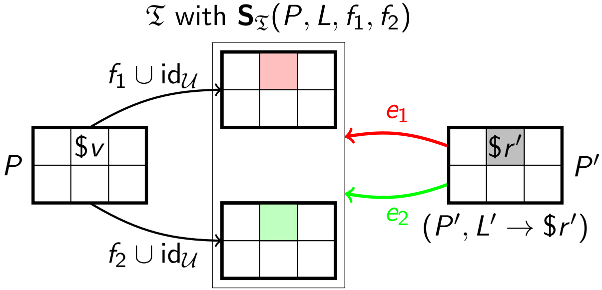
A symmetry-preserving step is performed: case 1a, 1b

Proposition (Pattern-modification)

If $(\phi_{t\hookleftarrow\$v}(P), \phi_{t\hookleftarrow\$v}(L) \cap \mathcal{V}_P \rightarrow \{\phi_{t\hookleftarrow\$v}(\$r)\} \cap \mathcal{V}_P)$, $(P', L' \rightarrow \$r')$, *and if there exists two embeddings of* P' *into* P *which agree on* L' *and map* $\$r'$ *to* t *and* $\$v$, *then* $(P, L \rightarrow \$r)$.



A symmetry-preserving step is performed: case 2



Afterwards

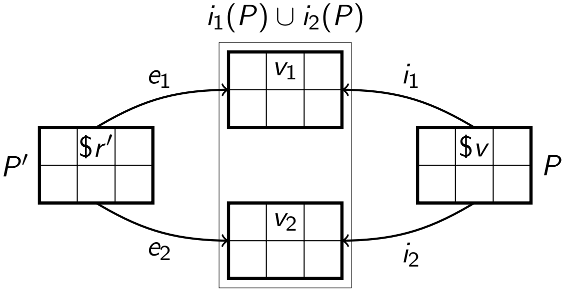
$\mathbf{S}_{\mathfrak{T}}\left(P, L \cup \{\$v\}, \phi_{f_1(\$v)\leftrightarrow f_2(\$v)} \circ f_1, \phi_{f_1(\$v)\leftrightarrow f_2(\$v)} \circ f_2\right)$

$$C' = (P, L \cup \{\$v\} \to \$r)$$

A symmetry-preserving step is performed: case 2

Proposition (Left-modification)

Let $i_1, i_2 : \mathcal{V}_P \to \mathcal{V}$ *be injective functions with* $i_1(\$v) \neq i_2(\$v)$*,* $i_1 =_L i_2$*, and* $range(i_1|_{\mathcal{V}_P \setminus L}) \cap range(i_2|_{\mathcal{V}_P \setminus L}) = \emptyset$*. If* $(P', L' \to \$r')$*,* $(P, L \cup \{\$v\} \to \$r)$*, and if there exist two embeddings from* P' *into* $(i_1 \cup id_{\mathcal{U}})(P) \cup (i_2 \cup id_{\mathcal{U}})(P)$ *which agree on* L' *and map* $\$r'$ *to* $i_1(\$v)$ *and* $i_2(\$v)$*, then* $(P, L \to \$r)$*.*



Overview

Results

- Sound and complete axiomatization for functional constraints
- Proof based on restructuring chase procedure
- Provides an answer to Constraints in RDF (Akhtar et al.)

Future work

- Complexity of implication for functional constraints
- Extending functional constraints with constructs of the form

$$(P, \$v = c), \$v \in \mathcal{V}_P, c \in \mathcal{U}$$

(used in conditional functional dependencies (Fan et al.))