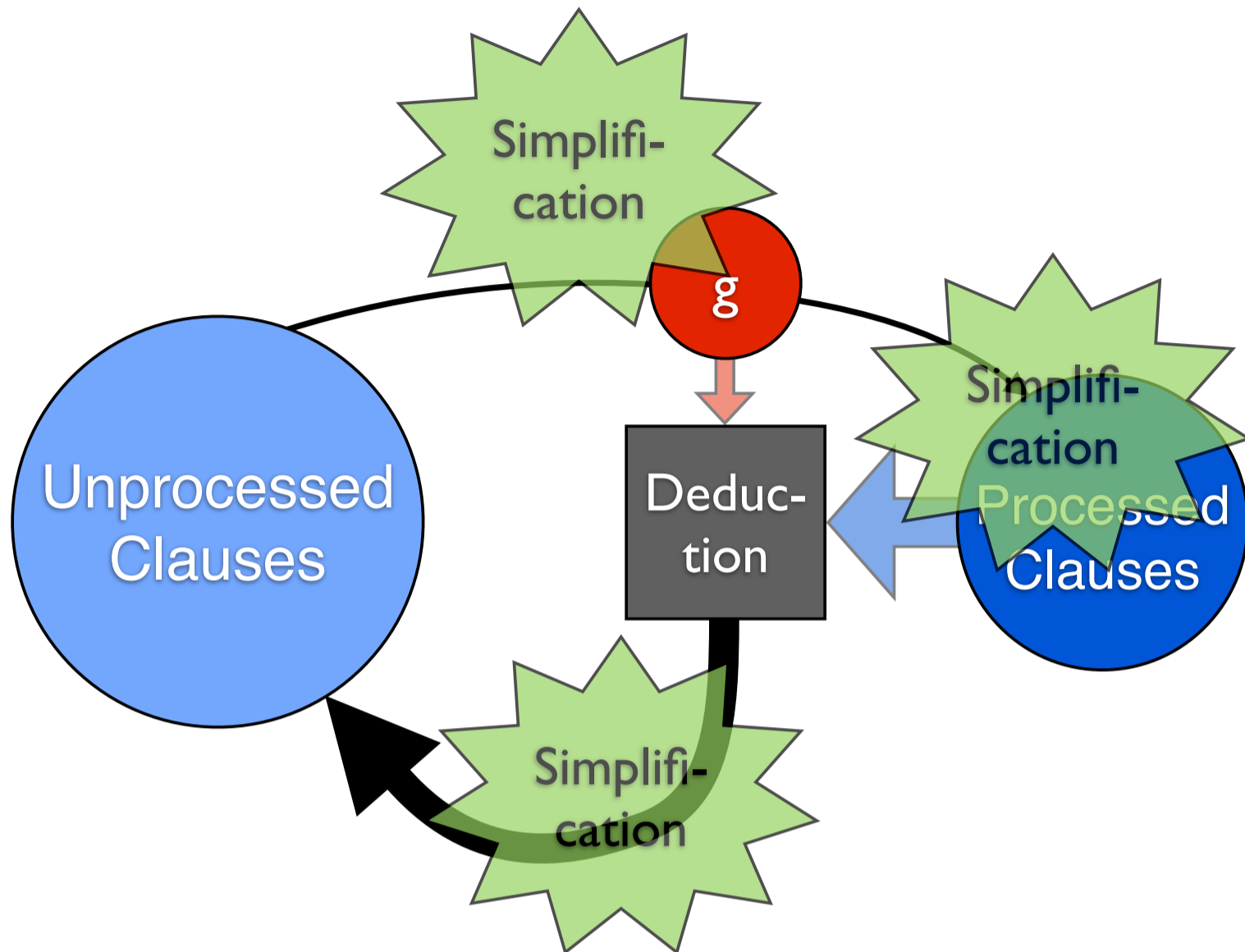


Fingerprint Indexing for Paramodulation

Stephan Schulz
schulz@eprover.org



Saturation with the Given-Clause Algorithm



Deduction

Compute all inferences where:

- ▶ **g** is at least one premise of the inference
- ▶ all other premises are processed clauses

»99%

paramodulation/resolution

Paramodulation

$$\frac{s \simeq t \vee S \quad u \simeq v \vee R}{\sigma(u[p \leftarrow t] \simeq v \vee S \vee R)}$$

- where $\sigma = mgu(s, u|_p)$
- possibly restricted by constraints (orderings, literal selection)



$$\frac{s \simeq t \vee S \quad u \simeq v \vee R}{\sigma(u[p \leftarrow t] \simeq v \vee S \vee R)}$$

$$\frac{s \simeq t \vee S \quad u \simeq v \vee R}{\sigma(u[p \leftarrow t] \simeq v \vee S \vee R)}$$

Paramodulate from g : Find suitable subterm $u|_p$ in processed clauses

Paramodulate into g : Find suitable LHS s in processed clauses

Reduction to Term Indexing

Given:

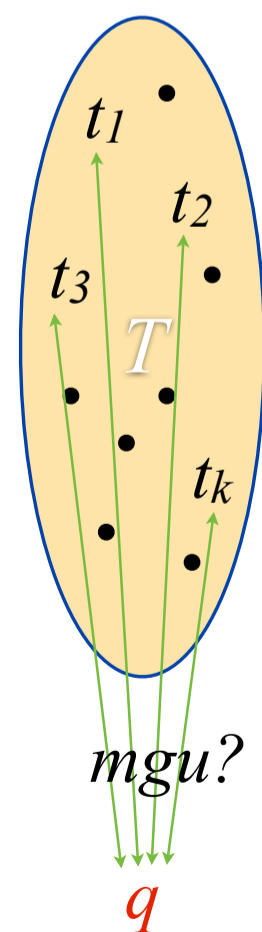
- ▶ A set of first-order terms T
- ▶ A query term q

Find (quickly):

- ▶ All $t \in T$ such that $mgu(q, t)$ exists

Side condition:

- ▶ Repeated queries with different q
- ▶ T is big ($\gg 10000$ terms)
- ▶ T evolves
 - ▶ Frequent additions
 - ▶ Infrequent deletions



Term Sampling

Consider t at (potential) position $p \in \mathbb{N}^*$

- ▶ \mathcal{N} : p does not exist in t or any instance
- ▶ \mathcal{B} : p is below a variable position in t
- ▶ \mathcal{A} : p denotes the occurrence of a variable in t
- ▶ f : t_p starts with function symbol f

Sampling function:

$$fps : Term(F, V) \times \mathbb{N}^* \rightarrow F \uplus \{\mathcal{A}, \mathcal{B}, \mathcal{N}\}$$

$$fps(t, p) = \begin{cases} \mathcal{A} & \text{if } p \in O(t), t|_p \in V \\ head(t|_p) & \text{if } p \in O(t), t|_p \notin V \\ \mathcal{B} & \text{if } p = q.r, q \in O(t) \text{ and } t|_q \in V \\ \mathcal{N} & \text{otherwise} \end{cases}$$

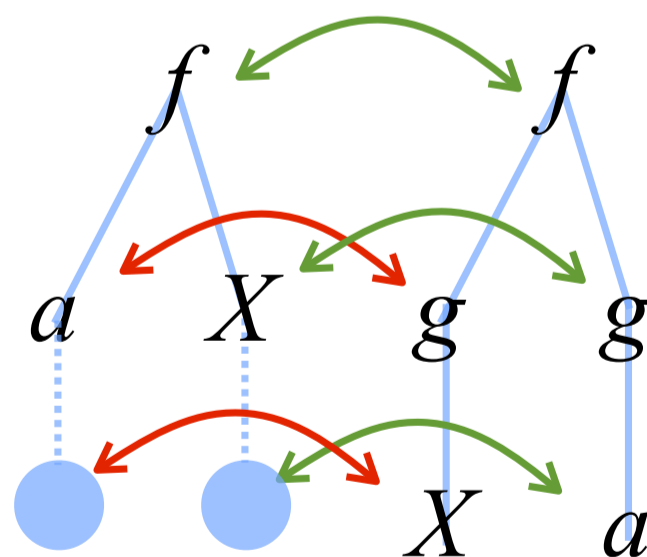
Sample Compatibility

Across: $fps(s, p)$

Down: $fps(t, p)$

Entry: Possibly unifiable?

| | f_1 | f_2 | \mathcal{A} | \mathcal{B} | \mathcal{N} |
|---------------|-------|-------|---------------|---------------|---------------|
| f_1 | Y | N | Y | Y | N |
| f_2 | N | Y | Y | Y | N |
| \mathcal{A} | Y | Y | Y | Y | N |
| \mathcal{B} | Y | Y | Y | Y | Y |
| \mathcal{N} | N | N | N | Y | Y |



Fingerprint Indexing

Fingerprint: Fixed length vector of term samples

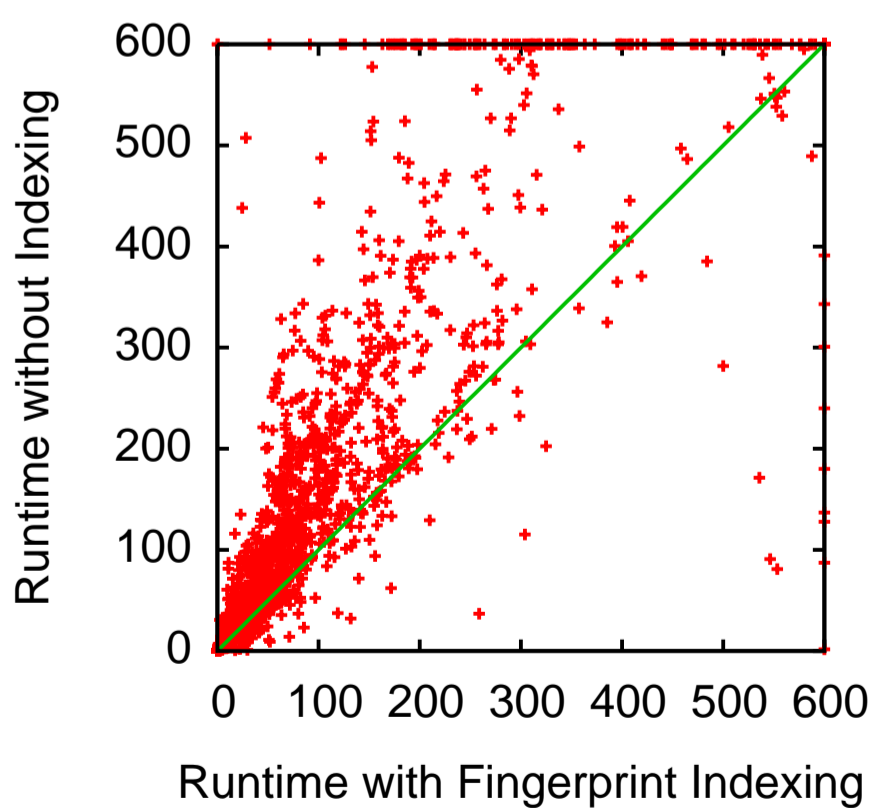
Fingerprints are organized in a **trie**

- ▶ Every term is represented only once in the index
- ▶ Index is compact (fingerprint length $\ll 10$)

Retrieval: Follow **all** compatible branches

- ▶ For identity only one choice
- ▶ For unification use preceding table
- ▶ For matching use alternative tables

Performance (very preliminary)

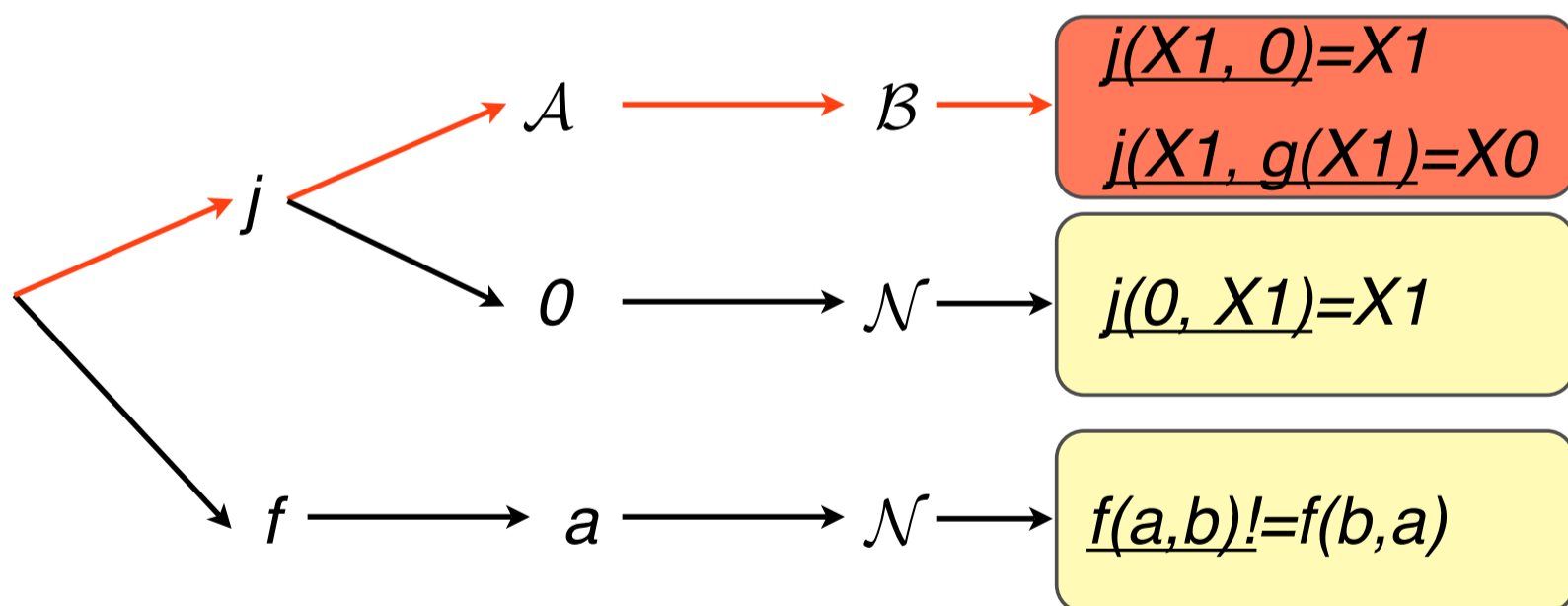


- ▶ E 1.2 (pre-release)
- ▶ TPTP 4.0.1
 - ≈ 14000 problems
 - ≈ 8500 solutions
- ▶ Abstract runtimes (normalized)
- ▶ 5-sample Fingerprint

Example

Fingerprint function: $fp_{3d}(t) = \langle fps(t, \epsilon), fps(t, 1), fps(t, 1.1) \rangle$

| Clause | Indexed term | Fingerprint |
|------------------------|----------------|---|
| $j(X1, 0) = X1$ | $j(X1, 0)$ | $\langle j, \mathcal{A}, \mathcal{B} \rangle$ |
| $j(0, X1) = X1$ | $j(0, X1)$ | $\langle j, 0, \mathcal{N} \rangle$ |
| $j(X1, g(X1)) = 0$ | $j(X1, g(X1))$ | $\langle j, \mathcal{A}, \mathcal{B} \rangle$ |
| $f(a, b) \neq f(b, a)$ | $f(a, b)$ | $\langle f, a, \mathcal{N} \rangle$ |



Query: $j(g(Y1), Y1)=0$

$fp_{3d}(j(g(Y1), Y1)) = \langle j, g, \mathcal{A} \rangle$