First-Order Term-Indexing

Alexander Maringele

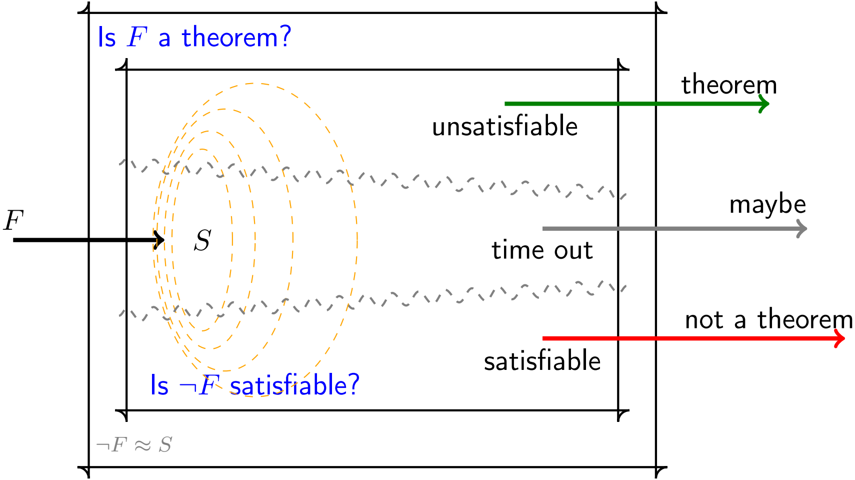
January 27th, 2016

References

R. Sekar, I. V. Ramakrishnan, and Andrei Voronkov, *Term indexing*, Handbook of Automated Reasoning (Alan Robinson and Andrei Voronkov, eds.), Elsevier Science Publishers B. V., Amsterdam, The Netherlands, 2001, pp. 1853–1964.

Outline

1. Motivation
2. Term Structure
3. Path Indexing
4. Discrimination Trees
5. Substitution Trees
6. Search times


Is F a theorem?
theorem
unsatisfiable
F
S
maybe
time out
not a theorem
satisfiable
Is $\neg F$ satisfiable?
$\neg F \approx S$

Clausal form

$$\{\ \mathsf{P}(\mathsf{f}(x)) \lor \mathsf{f}(x) \not\approx \mathsf{a},\ \ \mathsf{g}(x,y) \approx \mathsf{a} \lor \neg\mathsf{Q}(x,y),\ \ \mathcal{C}_3\ \}$$

$$\equiv$$

$$\forall x\,(\mathsf{P}(\mathsf{f}(x)) \lor \mathsf{f}(x) \not\approx \mathsf{a})$$

$$\land$$

$$\forall xy\,(\mathsf{g}(x,y) \approx \mathsf{a} \lor \neg\mathsf{Q}(x,y))$$

$$\land$$

$$\forall\,\mathcal{V}\mathsf{ar}(\mathcal{C}_3)\,(\mathcal{C}_3)$$

Goal

A sound and refutation complete calculus.

Resolution (without equality)

Resolve and factor all clauses and literals in an unsatisfiable set

$$\frac{A \vee \mathcal{C} \quad \neg B \vee \mathcal{D}}{(\mathcal{C} \vee \mathcal{D})\sigma}\ (\sigma)\ \text{resolution} \qquad \frac{A \vee B \vee \mathcal{C}}{(A \vee \mathcal{C})\sigma}\ (\sigma)\ \text{factoring}$$

$$\sigma = \mathrm{mgu}(A, B)$$

and the empty clause will be derived eventually.

Observation

Usually the set grows too fast to obtain a result.

Goal

A sound, refutation complete, and *effective* calculus.

1. *Reduce* search space
   - Ordered Resolution, Strategies, ...
   - ...with selection functions for clauses and literals
2. *Reduce* redundancy
   - e.g. discard clauses that are subsumed by other clauses
   - ...depending on the calculus

Example (forward subsumption)

$$S = \{{}^{1:}\mathsf{P}(x,y), {}^{2:}\neg\mathsf{P}(\mathsf{a},z)\} \cup \{{}^{3:}\mathsf{P}(\mathsf{a},z')\}$$ t_1 subsumes t_3

$$\frac{\mathsf{P}(x,y) \quad \neg\mathsf{P}(\mathsf{a},z)}{\square}\ \{x \mapsto \mathsf{a}, y \mapsto z\}$$ Resolution

$$S\bot = \{\mathsf{P}(\bot,\bot), \neg\mathsf{P}(\mathsf{a},\bot), \mathsf{P}(\mathsf{a},\bot)\}$$ InstGen / SMT

Goal

A sound, refutation complete, and effective calculus.

3. Quickly find
   - *variants* — variant removal
   - *instances* — backward subsumption
   - *generalizations* — forward subsumption
   - *unifiable terms* — resolution, demodulation

   of a query term in a given set of terms.

Observation

Deduction rate drops quickly with sequential search.

Term Indexing

Data structures and algorithms for fast retrieval of matching terms.

Positions of a term

$$\mathcal{P}\mathsf{os}(t) = \begin{cases} \{\epsilon\} & \text{if } t = x \in \mathcal{V} \\ \{\epsilon\} \cup \{ip \mid 1 \leq i \leq n \wedge p \in \mathcal{P}\mathsf{os}(t_i)\} & \text{if } t = f(t_1, \ldots, t_n) \end{cases}$$

Traversals of $\mathsf{h}(\mathsf{f}(\mathsf{a}, y))$

$$\begin{array}{ccc} & \langle \epsilon, \mathsf{h} \rangle & \\ & | & \\ & \langle 1, \mathsf{f} \rangle & \\ / & & \backslash \\ \langle 11, \mathsf{a} \rangle & & \langle 12, \mathsf{y} \rangle \end{array}$$

$$\mathcal{P}\mathsf{os}(\mathsf{h}(\mathsf{f}(\mathsf{a}, \mathsf{y}))) = \{\epsilon, 1, 11, 12\}$$

$$\mathsf{h}(\mathsf{f}(\mathsf{a}, \mathsf{y}))|_{12} = y \qquad \langle 12, y \rangle$$

$\langle \epsilon, \mathsf{h} \rangle \langle 1, \mathsf{f} \rangle \langle 12, y \rangle$ root to leaf y

$\langle \epsilon, \mathsf{h} \rangle \langle 1, \mathsf{f} \rangle \langle 11, \mathsf{a} \rangle \langle 12, y \rangle$ pre-order

Variables

Different terms generate the same position strings when

- variable names are ignored $\mathsf{f}(z, y), \mathsf{f}(y, x), \mathsf{f}(x, x) \Rightarrow \mathsf{f}(*, *)$
- or normalized $\mathsf{f}(z, y), \mathsf{f}(z, x) \Rightarrow \mathsf{f}(*_1, *_2)$
  $\mathsf{f}(x, x) \Rightarrow \mathsf{f}(*_1, *_1)$

In the second case only variants of terms generate the same strings.

Notation

We abbreviate

- path strings $\langle \epsilon, \mathsf{h}\rangle\langle 1, \mathsf{f}\rangle\langle 12, *\rangle$ $\mathsf{h}.1.\mathsf{f}.2.*$
- and traversal strings $\langle \epsilon, \mathsf{h}\rangle\langle 1, \mathsf{f}\rangle\langle 11, *\rangle\langle 12, *\rangle$ $\mathsf{h}.\mathsf{f}.\mathsf{a}.*$
  when traversal order and arity of symbols are fixed.

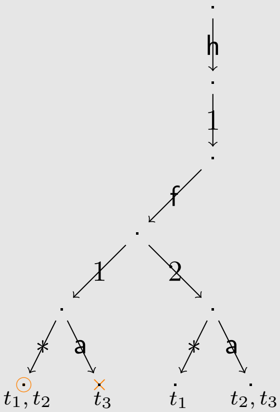
Build

$$^{t_1:}\mathsf{h}(\mathsf{f}(x,y)),\ ^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{a})),\ ^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{a},\mathsf{a}))$$

$$t_1 \Rightarrow \{\mathsf{h}.1.\mathsf{f}.1.*, \mathsf{h}.1.\mathsf{f}.2.*\}$$

$$t_2 \Rightarrow \{\mathsf{h}.1.\mathsf{f}.1.*, \mathsf{h}.1.\mathsf{f}.2.\mathsf{a}\}$$

$$t_3 \Rightarrow \{\mathsf{h}.1.\mathsf{f}.1.\mathsf{a}, \mathsf{h}.1.\mathsf{f}.2\mathsf{a}\}$$



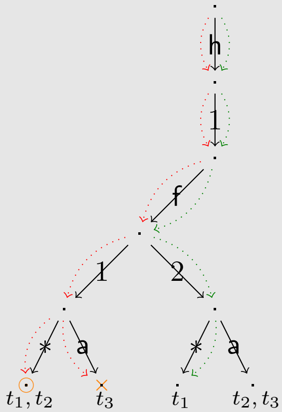
Retrieve

$^{t_1:}\mathsf{h}(\mathsf{f}(x,y)),\ ^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{a})),\ ^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{a},\mathsf{a}))$

$$\mathsf{h}(\mathsf{f}(x,\mathsf{b}))) \Rightarrow \{\mathsf{h.f.}*, \mathsf{h.f.b}\}$$

$$u : \mathsf{h}(\mathsf{f}(x',\mathsf{b})) \mapsto \{t_1, t_2, t_3\} \cap \{t_1, t_3\}$$
$$i : \mathsf{h}(\mathsf{f}(x',\mathsf{b})) \mapsto \{t_1, t_2, t_3\} \cap \{\}$$
$$g : \mathsf{h}(\mathsf{f}(x',\mathsf{b})) \mapsto \{t_1, t_2\} \cap \{t_1\}$$
$$v : \mathsf{h}(\mathsf{f}(x',\mathsf{b})) \mapsto \{t_1, t_2\} \cap \{\}$$

$$v : \mathsf{h}(\mathsf{f}(x',x')) \mapsto \{t_1, t_2\} \cap \{t_1\}$$



Unit Superposition Inference Rules

$$\frac{s \approx t \quad L[s']}{(L[t]) \cdot \sigma} \text{ unit paramodulation}$$

where $\sigma = \mathrm{mgu}(s, s')$, $s' \notin \mathcal{V}$, $t\sigma \not\succeq s\sigma$

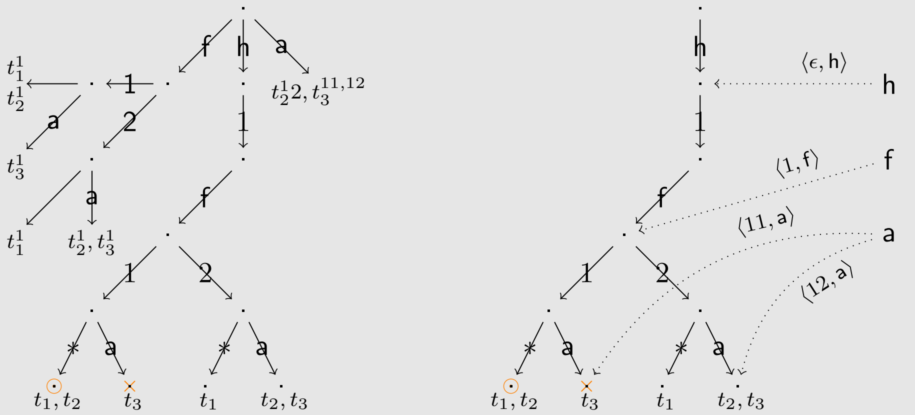
$$\frac{s \approx t \quad u[s'] \not\approx v}{(u[t] \not\approx v) \cdot \sigma} \text{ unit superposition } \frac{s \approx t \quad u[s'] \approx v}{(u[t] \approx v) \cdot \sigma}$$

where $\sigma = \mathrm{mgu}(s, s')$, $s' \notin \mathcal{V}$, $t\sigma \not\succeq s\sigma$, $v\sigma \not\succeq u[s']\sigma$

$$\frac{s \not\approx t}{\square} \text{ unit equality resolution} \qquad \frac{A \quad \neg B}{\square} \text{ unit resolution}$$

where s and t (A and B respectively) are unifiable

$^{t_4:}\mathsf{f}(x, \mathsf{a}) \approx x$



$^{t_1:}\mathsf{h}(\mathsf{f}(x, y)), {}^{t_2:}\mathsf{h}(\mathsf{f}(x, \mathsf{a})), {}^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{a}, \mathsf{a}))$

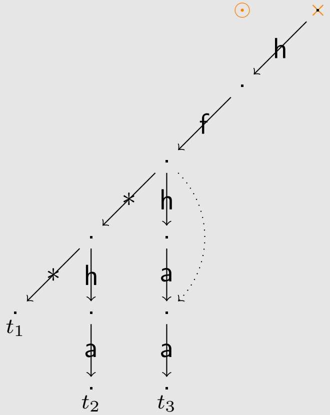
Insert

$^{t_1:}\mathsf{h}(\mathsf{f}(x,y)), {}^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{h}(\mathsf{a}))), {}^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{h}(\mathsf{a}),\mathsf{a}))$

$$t_1 \Rightarrow \mathsf{h.f.}{*}.{*}$$

$$t_2 \Rightarrow \mathsf{h.f.}{*}.\mathsf{h.a}$$

$$t_3 \Rightarrow \mathsf{h.f.h.a.a}$$

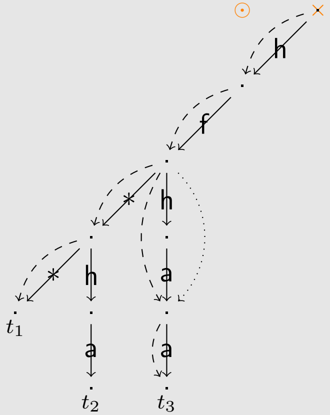


Retrieve

$^{t_1:}\mathsf{h}(\mathsf{f}(x,y)), {}^{t_2:}\mathsf{h}(\mathsf{f}(x,\mathsf{h}(\mathsf{a}))), {}^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{h}(\mathsf{a}),\mathsf{a}))$

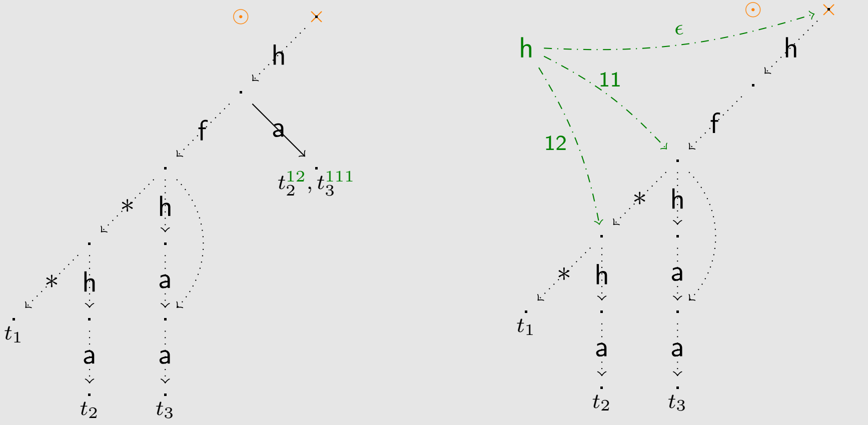
$$\mathsf{h}(\mathsf{f}(x',\mathsf{a})) \Rightarrow \mathsf{h.f.}{*}\mathsf{.a}$$

$$u : \mathsf{h}(\mathsf{f}(x',\mathsf{a})) \mapsto \{t_1, t_3\}$$
$$i : \mathsf{h}(\mathsf{f}(x',\mathsf{a})) \mapsto \{t_3\}$$
$$g : \mathsf{h}(\mathsf{f}(x',\mathsf{a})) \mapsto \{t_1\}$$
$$v : \mathsf{h}(\mathsf{f}(x',\mathsf{a})) \mapsto \{\ \}$$



Subterms

$^{t_1:}\mathsf{h}(\mathsf{f}(x, y)), ^{t_2:}\mathsf{h}(\mathsf{f}(x, \mathsf{h}(\mathsf{a}))), ^{t_3:}\mathsf{h}(\mathsf{f}(\mathsf{h}(\mathsf{a}), \mathsf{a}))$



`TPTP/Problems/HWV/HWV134-1.p`
`2 332 428 formulae, 6 570 884 literals`

literals new	total	(ℓ_1, ℓ_2)	$A, \neg B$	sequential search	index search	speed up
1 000	1 000	500 000	761	726ms	70ms	10
1 000	2 000	1 500 000	812	2s	69ms	29
1 000	4 000	3 500 000	723	4s	75ms	53
1 000	8 000	7 500 000	433	9s	125ms	72
1 000	16 000	15 500 000	742	21s	221ms	95
1 000	32 000	31 500 000	592	40s	489ms	82
1 000	64 000	63 500 000	1167	*80s*	697ms	115
1 000	128 000	127 500 000	1479	*160s*	13s	12
1 000	256 000	255 500 000	1097	*320s*	440s	1
1 000	512 000	511 500 000	1440	*640s*	348s	2
1 000	1 024 000	1023 500 000	1534	*1280s*	348s	4