Elusive "Why" — An experiment in capturing proof intent

Leo Freitas (University of Newcastle)

AI4FM workshop, Edinburgh, 28/04/2011

Notions of "why" in theorem proving

- internal discussions on "how to say why"
  - Alan's proof scenarios + equational reasoning
  - Cliff's Why language / examples
  - Gudmund's lemma generation with Isa/Scheme
- attempts at different examples in Z and Event-B
  - Tokeneer and R^+-lemmas; Siemens and MIDAS
  - problem: large specs with "trivial" proofs
  - too much detail and tinkering is involved
- Leo's method for proof engineering

Problems with documenting intent

- path to failure can be tedious and verbose
- modelling motivations often do not take proof into account
  - "abstract" models serve specific agendas (customer agreements x design guides)
  - prover idioms x method/tool dependencies
- modelling for proving works both ways:
  - careless abstractions makes for unnecessarily harder proofs
  - concrete abstract modelling: easier proof x compromised clarity
- mechanisation idioms are often dissociated from problem
  - proof idioms can be problem independent
  - meaning they can be far from the problem too!
- needle-in-haystack search (*e.g.*, Mondex, Tokeeneer)
- proof decomposition/lemma suggestion are key; not straightforward
  - TIS Op: 112 vars; 45 p. goal; 60 $\exists$-witn on an "easy" proof

Approaches to documenting intent

1. active: Naur's diary log entries
   - captures the thinking process well
   - can be verbose and hard to mechanise
2. passive: user tracing + innovative HCI (?)
   - automatic, yet work-environment dependent
   - large amount of data: what is relevant?
3. mixed: descriptive tactical language (*e.g.*, Isar)
   - very clear, concise, and readiable
   - obfuscates paths / reasons to failure

Approaches to documenting intent

1. active: Naur's diary log entries
   - captures the thinking process well
   - can be verbose and hard to mechanise
2. passive: user tracing + innovative HCI (?)
   - automatic, yet work-environment dependent
   - large amount of data: what is relevant?
3. mixed: descriptive tactical language (*e.g.*, Isar)
   - very clear, concise, and readiable
   - obfuscates paths / reasons to failure

- **how could we measure progress?**

Peter Naur's reflection over Wirth's N-Queens problem

- Wirth's suggests: think about N-Queens problem before reading my paper
- Naur does just that by keeping a "diary"-like log of 38 entries about the problem
- curious style of presentation of ideas
  - bit verbose and conversational at times, if succinct.
  - nicely captures and conveys thinking processes.
- intent is pervasive with progress or leaps clearly documented
- **basis for our experiment on capturing "Why"**

Peter Naur's reflection over Wirth's N-Queens problem

- Wirth's suggests: think about N-Queens problem before reading my paper
- Naur does just that by keeping a "diary"-like log of 38 entries about the problem
- curious style of presentation of ideas
  - bit verbose and conversational at times, if succinct.
  - nicely captures and conveys thinking processes.
- intent is pervasive with progress or leaps clearly documented
- **basis for our experiment on capturing "Why"**

- **give us the dead ends you usually throw away!**

AI4FM experiment

- what model examples to use?
  - industry: *e.g.*, Tokeneer is large, mostly with simple proofs
  - academic: *e.g.*, could miss proof engineering issues
- mid-ground: proofs about transitive closure
  - used in a component for an embedded OS kernel model
  - part of a GC pilot project + few MSc projects (@ York)
- documented proof process similarly to Naur's logs (NCL TR)
- we wanted to instantiate Alan's proof scenarios
- and Cliff's models of why (earlier talks)

AI4FM experiment on R^+ - setup

AI4FM experiment on R^+ - setup

1. $(\{s\} \lhd R)^+ \stackrel{?}{=} \{s\} \lhd R^+$

AI4FM experiment on R^+ - setup

1. $(\{s\} \lhd R)^+ \stackrel{?}{=} \{s\} \lhd R^+$
   - hard witnesses with blunt proof attempt
   - counterexample when $s \in \mathrm{ran}\ R$ (e.g., flight patterns)

AI4FM experiment on R^+ - setup

1. $(\{s\} \lhd R)^+ \stackrel{?}{=} \{s\} \lhd R^+$
   - hard witnesses with blunt proof attempt
   - counterexample when $s \in \mathrm{ran}\ R$ (e.g., flight patterns)
2. $R^+ \mathrel{\widehat{=}} \bigcap \{Q \mid Q \fatsemi Q \subseteq Q \wedge R \subseteq Q\}$

AI4FM experiment on R^+ - setup

1. $(\{s\} \lhd R)^+ \stackrel{?}{=} \{s\} \lhd R^+$
   - hard witnesses with blunt proof attempt
   - counterexample when $s \in \mathrm{ran}\ R$ (e.g., flight patterns)
2. $R^+ \mathrel{\widehat{=}} \bigcap \{Q \mid Q \fatsemi Q \subseteq Q \wedge R \subseteq Q\}$
   - get the minimal transitively closed solution
   - counterexample shows alternative: iterative using $\bigcup$
   - good for inductive proofs + easier witnesses

AI4FM experiment on R^+ - setup

1. $(\{s\} \lhd R)^+ \stackrel{?}{=} \{s\} \lhd R^+$
   - hard witnesses with blunt proof attempt
   - counterexample when $s \in \mathrm{ran}\ R$ (e.g., flight patterns)
2. $R^+ \mathrel{\widehat{=}} \bigcap \{Q \mid Q \fatsemi Q \subseteq Q \wedge R \subseteq Q\}$
   - get the minimal transitively closed solution
   - counterexample shows alternative: iterative using $\bigcup$
   - good for inductive proofs + easier witnesses
3. propose lemma: $R^+ = \bigcup_{i \geq 1} R^i$, where $R^{i+1} = R \fatsemi R^i$; $i \geq 0$

AI4FM experiment on R^+ - setup

1. $(\{s\} \lhd R)^+ \stackrel{?}{=} \{s\} \lhd R^+$
   - hard witnesses with blunt proof attempt
   - counterexample when $s \in \mathrm{ran}\ R$ (e.g., flight patterns)
2. $R^+ \mathrel{\widehat{=}} \bigcap \{Q \mid Q \fatsemi Q \subseteq Q \wedge R \subseteq Q\}$
   - get the minimal transitively closed solution
   - counterexample shows alternative: iterative using $\bigcup$
   - good for inductive proofs + easier witnesses
3. propose lemma: $R^+ = \bigcup_{i \geq 1} R^i$, where $R^{i+1} = R \fatsemi R^i$; $i \geq 0$
   - proofs now are about finding right i^{th} iteration via induction
   - progress modulo type judgments over homogeneous relations $p \in R^+ \Rightarrow p \in T \times T$, $R^i \subseteq T \times T$, $\bigcup_{i \geq 1} R^i \subseteq T \times T$

AI4FM experiment on R^+ - induction

AI4FM experiment on R^+ - induction

4 base case: $(\{s\} \lhd R)^1 = \{s\} \lhd R^1$
induc. case: $(\{s\} \lhd R)^i = \{s\} \lhd R^i \Rightarrow (\{s\} \lhd R)^{i+1} = \{s\} \lhd R^{i+1}$

AI4FM experiment on R^+ - induction

4 base case: $(\{s\} \lhd R)^1 = \{s\} \lhd R^1$
induc. case: $(\{s\} \lhd R)^i = \{s\} \lhd R^i \Rightarrow (\{s\} \lhd R)^{i+1} = \{s\} \lhd R^{i+1}$

- extreme lemmas: $R^1 = R \quad R^0 = \mathrm{id}\ T$
- decomp. (right) lemma: $R^i = R \fatsemi R^{i-1}, i \geq 1$
- decomp. (left) lemma: $R^i = R^{i-1} \fatsemi R, i \geq 1$ FAILED!
- $\lhd$-$\fatsemi$-dist lemma:
$$\{s\} \lhd (R \fatsemi Q) = (\{s\} \lhd R) \fatsemi (\{s\} \lhd Q)$$

AI4FM experiment on R^+ - induction

4 base case: $(\{s\} \lhd R)^1 = \{s\} \lhd R^1$
induc. case: $(\{s\} \lhd R)^i = \{s\} \lhd R^i \Rightarrow (\{s\} \lhd R)^{i+1} = \{s\} \lhd R^{i+1}$

- extreme lemmas: $R^1 = R \quad R^0 = \mathrm{id}\ T$
- decomp. (right) lemma: $R^i = R \fatsemi R^{i-1}, i \geq 1$
- decomp. (left) lemma: $R^i = R^{i-1} \fatsemi R, i \geq 1$ FAILED!
- $\lhd$-$\fatsemi$-dist lemma:
$$\{s\} \lhd (R \fatsemi Q) = (\{s\} \lhd R) \fatsemi (\{s\} \lhd Q)$$

5 proof of $R^+ = \bigcup_{i \geq 1} R^i$ needs various lemmas

AI4FM experiment on R^+ - induction

4 base case: $(\{s\} \lhd R)^1 = \{s\} \lhd R^1$
induc. case: $(\{s\} \lhd R)^i = \{s\} \lhd R^i \Rightarrow (\{s\} \lhd R)^{i+1} = \{s\} \lhd R^{i+1}$

- extreme lemmas: $R^1 = R \quad R^0 = \mathrm{id}\ T$
- decomp. (right) lemma: $R^i = R \fatsemi R^{i-1}, i \geq 1$
- decomp. (left) lemma: $R^i = R^{i-1} \fatsemi R, i \geq 1$ FAILED!
- $\lhd$-$\fatsemi$-dist lemma:

$$\{s\} \lhd (R \fatsemi Q) = (\{s\} \lhd R) \fatsemi (\{s\} \lhd Q)$$

5 proof of $R^+ = \bigcup_{i \geq 1} R^i$ needs various lemmas

- iter-$\fatsemi$ splitting: $R^{i+j} = R^i \fatsemi R^j$ (induc)
- iter-closure on R^+: $R^i \subseteq R^+ \Rightarrow R^{i+1} \subseteq R^+, i \geq 1$ (induc)
- decomp. (left) lemma: $R^i = R^{i-1} \fatsemi R, i \geq 1$ SUCCEEDS!

AI4FM experiment on R^+ - summary

- goal decomposition summary
  - head-on collision with difficult $\exists$-witness
  - alternative inductive formulation with iteration
  - fiddling with prover's idioms on type judgements
  - needed lemmas for inductive goals
  - extra lemmas proved from the shape of goals
- "post-mortem" lemma (re-)organisation
  - meta-tagging of lemmas roles within proofs
  - determines what are type-judgements or rewrite rules
  - identify L-R order of equations and ability (on/off) lemmas
- repeated process for other operators $\trianglerighteq$, $\lhd$, $\rhd$, $\oplus$, *etc.*
- **Similar exercise done for Tokeeneer proofs**

AI4FM industrial experiment - Tokeneer

- instrumenting Eclipse for Naur's logs + PSP
- prototype on Cliff's models of "Why" + Alan's proof scenarios
- diff. analysis of proof scripts + model attempts + failures + ...
- HCI experiment setup for students over summer
- collect measurable data and define notions of "progress" (even if just experimentally - not stats. significant)

Conclusions

- key problems to tackle:
  - what is academically interesting from industry's proofs?
  - how to transfer academic practices to industry effectively?
  - what are the main barriers?
  - are the models of why a good impractical solution?
  - what can the lack of training tell about current practices?
- how to measure progress?
  - adapting CMU's PSP process for proof
  - Tokeneer experiments with undergraduates over summer
- current work
  - progress on Cliff's language to capturing proof intent
  - what should be the tagging principles / burden for the user?
  - integration with Isabelle/HOL; and/or other provers?
  - how to include machine learning in the process?
  - Isabelle/Eclipse + PSP + Models why prototype