An Overview on the Projects LEO-II, DIALOG, and PLATO/OMEGA

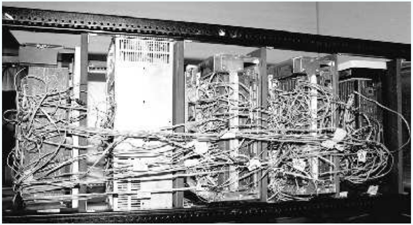
Christoph E. Benzmüller

Formal Mathematics Seminar
University of Bonn, June 13, 2008

LEO-II – A Cooperative Automatic Higher-Order Theorem Prover

Higher-Order Logic (HOL)

Some people say that HOL is like this:



I don't!

- Semantics (extensionality) **PhD-99,JSL-04**
- Proof theory **IJCAR-06**
- ATPs LEO and LEO-II **CADE-98,IJCAR-08**

Higher-Order Logic (HOL)
- on one slide -

Property	FOL	HOL	Example
Quantification over			
- individuals	✓	✓	$\forall x.\, P(F(x))$
- functions	–	✓	$\forall F.\, P(F(x))$
- predicates/sets/relations	–	✓	$\forall P.\, P(F(x))$
Unnamed			
- functions	–	✓	$(\lambda x.\, x)$
- predicates/sets/relations	–	✓	$(\lambda x.\, x \neq 2)$
Statements about			
- functions	–	✓	$\mathit{continuous}(\lambda x.\, x)$
- predicates/sets/relations	–	✓	$\mathit{reflexive}(=)$

Sets and Relations in HOL

$$A \cup B \quad := \qquad \{x \mid x \in A \vee x \in B\}$$

Sets and Relations in HOL

$$A \cup B \quad := \qquad (\lambda x . x \in A \vee x \in B)$$

Sets and Relations in HOL

$$\cup \quad := \quad \lambda A_{\blacksquare} \lambda B_{\blacksquare} (\lambda x_{\blacksquare} x \in A \vee x \in B)$$

Sets and Relations in HOL

$$\texttt{symmetric} \quad := \quad \lambda F_{\blacksquare}(\forall x, y_{\blacksquare} F(x,y) = F(y,x))$$

Sets and Relations in HOL

Theorem : $\quad \texttt{symmetric}(\cup)$

Sets and Relations in HOL

Sets and Relations in HOL

$$
\begin{array}{lll}
\in & := & \lambda x.\, \lambda A.\, A(x) \\
\emptyset & := & \lambda x.\, \bot \\
\cap & := & \lambda A.\, \lambda B.\, (\lambda x.\, x \in A \wedge x \in B) \\
\cup & := & \lambda A.\, \lambda B.\, (\lambda x.\, x \in A \vee x \in B) \\
\setminus & := & \lambda A.\, \lambda B.\, (\lambda x.\, x \in A \vee x \notin B) \\
\ldots & & \\
\subseteq & := & \lambda A.\, \lambda B.\, (\forall x.\, x \in A \Rightarrow x \in B) \\
\mathcal{P} & := & \lambda A.\, (\lambda B.\, B \subseteq A) \\
\ldots & & \\
\texttt{reflexive} & := & \lambda R.\, (\forall x.\, R(x, x)) \\
\texttt{transitive} & := & \lambda R.\, (\forall x, y, z.\, (R(x, y) \wedge R(y, z)) \Rightarrow R(x, z)) \\
\ldots & &
\end{array}
$$

Types are Needed

Typed Sets and Relations in HOL

$$
\begin{array}{lll}
\in & := & \lambda x_\alpha.\lambda A_{\alpha\to o}.A(x) \\
\emptyset & := & \lambda x_\alpha.\bot \\
\cap & := & \lambda A_{\alpha\to o}.\lambda B_{\alpha\to o}.(\lambda x_\alpha.x \in A \wedge x \in B) \\
\cup & := & \lambda A_{\alpha\to o}.\lambda B_{\alpha\to o}.(\lambda x_\alpha.x \in A \vee x \in B) \\
\setminus & := & \lambda A_{\alpha\to o}.\lambda B_{\alpha\to o}.(\lambda x_\alpha.x \in A \vee x \notin B) \\
\ldots & &
\end{array}
$$

Polymorphism is a Challenge for Automation

- Another source of indeterminism / blind guessing

TPHOLs-WP-07

Automation of HOL: A Nightmare?

Undecidable and Infinitary Unification

$\exists F_{\iota\to\iota}\text{.}\, F(g(x)) = g(F(x))$

(1) $F \leftarrow \lambda y_i\text{.}\, y$
(2) $F \leftarrow \lambda y_i\text{.}\, g(y)$
(3) $F \leftarrow \lambda y_i\text{.}\, g(g(y))$
(4) $\ldots$

Automation of HOL: A Nightmare?

Primitive Substitution

Example Theorem: $\exists S_{\bullet}\, reflexive(S)$

Negation and Expansion of Definitions:

$$\neg\exists S_{\bullet}\,(\forall x_{\iota\bullet}\, S(x,x))$$

Clause Normalisation ($a(S)$ Skolem term):

$$\neg S(a(S), a(S))$$

Guess some suitable instances for S

$S \leftarrow \lambda y_{\bullet}\, \lambda z_{\bullet}\, \top$

$\rightsquigarrow \neg\top$

$S \leftarrow \lambda y_{\bullet}\, \lambda z_{\bullet}\, V(y,z) = W(y,z)$

$\rightsquigarrow V(a(S), a(S)) \neq W(a(S), a(S))$

$S \leftarrow \ldots$

Automation of HOL: A Nightmare?

Cut rule

$$\frac{A \Rightarrow B}{A \Rightarrow C \quad C \Rightarrow B}$$

considered as bad in ATP

Automation of HOL: A Nightmare?

Cut rule

$$\frac{A \Rightarrow B}{A \Rightarrow C \quad C \Rightarrow B}$$

considered as bad in ATP

IJCAR-06: Axioms that imply Cut

- Axiom of excluded middle
- Comprehension axioms
- Functional and Boolean extensionality
- Leibniz and other definitions of equality
- Axiom of induction
- Axiom of choice
- Axiom of description

Automation of HOL: A Nightmare?

Cut rule

$$\frac{A \Rightarrow B}{A \Rightarrow C \quad C \Rightarrow B}$$

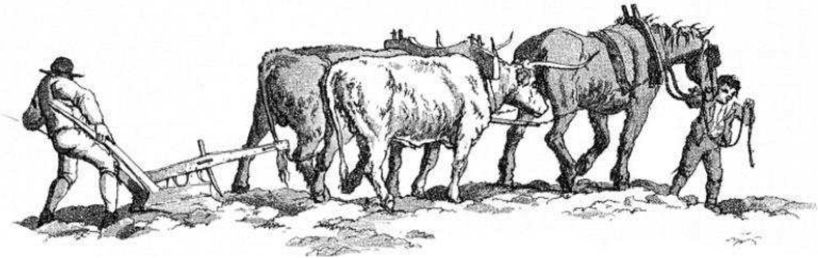
considered as bad in ATP

Calculi that avoid axioms

- Axiom of excluded middle ✓
- Comprehension axioms ✓
- Functional and Boolean extensionality ✓[CADE-98,PhD-99]
- Leibniz and other definitions of equality ✓[CADE-99,PhD-99]
- Axiom of induction ?
- Axiom of choice –
- Axiom of description –

LEO-II

An Effective Higher-Order Theorem Prover



LEO-II employs FO-ATPs: E, Spass, Vampire

Motivation for LEO-II

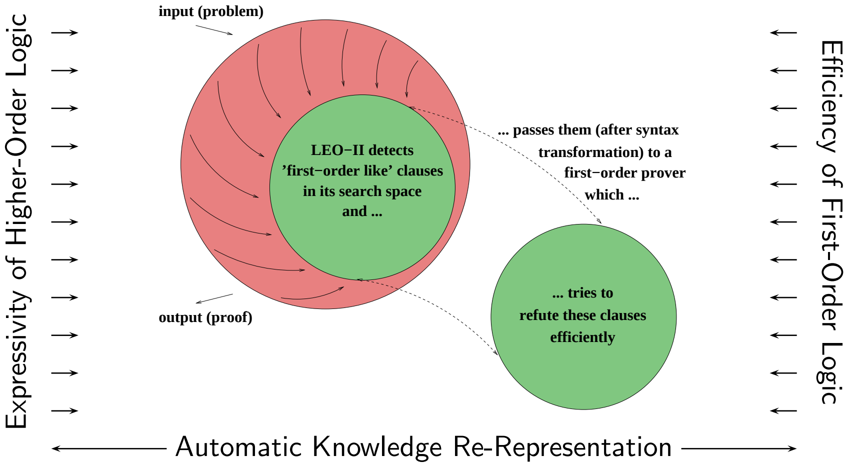
- TPS system of Peter Andrews et al.

- LEO hardwired to ΩMEGA(predecessor of LEO-II)
- Agent-based architecture Ω-ANTS
  (with V. Sorge) **AIMSA-98,EPIA-99,Calculemus-00**

- Collaboration of LEO with FO-ATP via Ω-ANTS
  (with V. Sorge) **KI-01,LPAR-05,JAL-07**

- Progress in Higher-Order Termindexing
  (with F. Theiss and A. Fietzke) **IWIL-06**

Architecture of LEO-II



Solving Lightweight Problems

Example: TPTP Problem SET171+3

Axiomatization in FO Set Theory

Assumptions:

$\forall B, C, x. (x \in (B \cup C) \Leftrightarrow x \in B \vee x \in C)$

$\forall B, C, x. (x \in (B \cap C) \Leftrightarrow x \in B \wedge x \in C)$

$\forall B, C. (B \subseteq C \Leftrightarrow \forall x. x \in B \Rightarrow x \in C)$

$\forall B, C. (B = C \Leftrightarrow B \subseteq C \wedge C \subseteq B)$

Proof Goal:

$$\forall B, C, D.\quad B \cup (C \cap D) = (B \cup C) \cap (B \cup D)$$

Example: TPTP Problem SET171+3

Axiomatization in FO Set Theory

Assumptions:

$\forall B, C, x. (x \in (B \cup C) \Leftrightarrow x \in B \vee x \in C)$

$\forall B, C, x. (x \in (B \cap C) \Leftrightarrow x \in B \wedge x \in C)$

$\forall B, C. (B \subseteq C \Leftrightarrow \forall x. x \in B \Rightarrow x \in C)$

$\forall B, C. (B \cup C = C \cup B)$

$\forall B, C. (B \cap C = C \cap B)$

$\forall B, C. (B = C \Leftrightarrow B \subseteq C \wedge C \subseteq B)$

$\forall B, C. (B = C \Leftrightarrow \forall x. x \in B \Leftrightarrow x \in C)$

Proof Goal:

$$\forall B, C, D. \quad B \cup (C \cap D) = (B \cup C) \cap (B \cup D)$$

Example: TPTP Problem SET171+3

Axiomatization in FO Set Theory

Assumptions:

$\forall B, C, x_{\blacksquare}(x \in (B \cup C) \Leftrightarrow x \in B \vee x \in C)$

$\forall B, C, x_{\blacksquare}(x \in (B \cap C) \Leftrightarrow x \in B \wedge x \in C)$

$\forall B, C_{\blacksquare}(B \subseteq C \Leftrightarrow \forall x_{\blacksquare} x \in B \Rightarrow x \in C)$

$\forall B, C_{\blacksquare}(B \cup C = C \cup B)$

$\forall B, C_{\blacksquare}(B \cap C = C \cap B)$

$\forall B, C_{\blacksquare}(B = C \Leftrightarrow B \subseteq C \wedge C \subseteq B)$

$\forall B, C_{\blacksquare}(B = C \Leftrightarrow \forall x_{\blacksquare} x \in B \Leftrightarrow x \in C)$

Proof Goal:

$\forall B, C, D_{\blacksquare}$

$$B \cup (C \cap D) = (B \cup C) \cap (B \cup D)$$

Performance: FO-ATPs

```
% SPASS---3.0
% Problem : SET171+3
% SPASS beiseite: Ran out of time.

% E---0.999
% Problem : SET171+3
% Failure: Resource limit exceeded
(time)

% Vampire---9.0
% Problem : SET171+3
% Result : Theorem 68.6s
```

Example: TPTP Problem SET171+3

Axiomatization in FO Set Theory

Assumptions:

$\forall B, C, x. (x \in (B \cup C) \Leftrightarrow x \in B \vee x \in C)$

$\forall B, C, x. (x \in (B \cap C) \Leftrightarrow x \in B \wedge x \in C)$

$\forall B, C. (B \subseteq C \Leftrightarrow \forall x. x \in B \Rightarrow x \in C)$

$\forall B, C. (B \cup C = C \cup B)$

$\forall B, C. (B \cap C = C \cap B)$

$\forall B, C. (B = C \Leftrightarrow B \subseteq C \wedge C \subseteq B)$

$\forall B, C. (B = C \Leftrightarrow \forall x. x \in B \Leftrightarrow x \in C)$

Proof Goal:

$\forall B, C, D.$

$$B \cup (C \cap D) = (B \cup C) \cap (B \cup D)$$

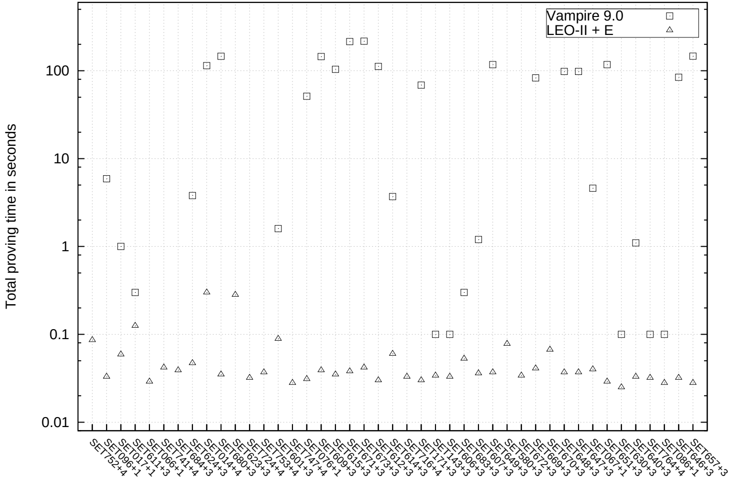
Performance: FO-ATPs

```
% SPASS---3.0
% Problem : SET171+3
% SPASS beiseite: Ran out of time.

% E---0.999
% Problem : SET171+3
% Failure: Resource limit exceeded
(time)

% Vampire---9.0
% Problem : SET171+3
% Result : Theorem 68.6s
```

Performance: LEO-II + E

```
Eureka --- Thanks to Corina!
Total Reasoning Time: 0.03s
LEO-II (Proof Found!)
```

Results



Vampire 9.0
LEO-II + E
Total proving time in seconds
100
10
1
0.1
0.01
SET752+4
SET096+1
SET017+1
SET611+3
SET066+1
SET741+4
SET684+3
SET624+3
SET014+4
SET680+3
SET623+3
SET724+4
SET753+4
SET601+3
SET747+4
SET076+1
SET609+3
SET615+3
SET671+3
SET673+3
SET612+3
SET614+3
SET716+4
SET171+3
SET143+3
SET606+3
SET683+3
SET607+3
SET649+3
SET580+3
SET672+3
SET669+3
SET670+3
SET648+3
SET647+3
SET067+1
SET651+3
SET630+3
SET640+3
SET764+4
SET086+1
SET646+3
SET657+3

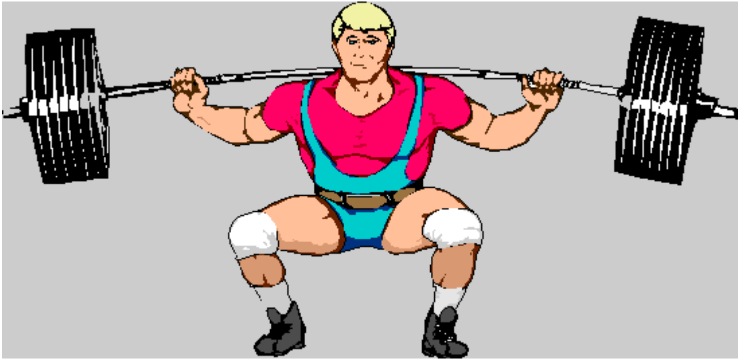
Results

Problem	Vamp. 9.0	LEO+Vamp.	LEO-II+E
014+4	114.5	2.60	0.300
017+1	1.0	5.05	0.059
066+1	–	3.73	0.029
067+1	4.6	0.10	0.040
076+1	51.3	0.97	0.031
086+1	0.1	0.01	0.028
096+1	5.9	7.29	0.033
143+3	0.1	0.31	0.034
171+3	68.6	0.38	0.030
580+3	0.0	0.23	0.078
601+3	1.6	1.18	0.089
606+3	0.1	0.27	0.033
607+3	1.2	0.26	0.036
609+3	145.2	0.49	0.039
611+3	0.3	4.00	0.125
612+3	111.9	0.46	0.030
614+3	3.7	0.41	0.060
615+3	103.9	0.47	0.035
623+3	–	2.27	0.282
624+3	3.8	3.29	0.047
630+3	0.1	0.05	0.025
640+3	1.1	0.01	0.033
646+3	84.4	0.01	0.032
647+3	98.2	0.12	0.037

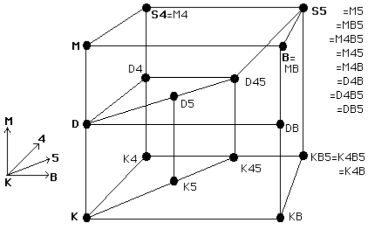
Problem	Vamp. 9.0	LEO+Vamp.	LEO-II+E
648+3	98.2	0.12	0.037
649+3	117.5	0.25	0.037
651+3	117.5	0.09	0.029
657+3	146.6	0.01	0.028
669+3	83.1	0.20	0.041
670+3	–	0.14	0.067
671+3	214.9	0.47	0.038
672+3	–	0.23	0.034
673+3	217.1	0.47	0.042
680+3	146.3	2.38	0.035
683+3	0.3	0.27	0.053
684+3	–	3.39	0.039
716+4	–	0.40	0.033
724+4	–	1.91	0.032
741+4	–	3.70	0.042
747+4	–	1.18	0.028
752+4	–	516.00	0.086
753+4	–	1.64	0.037
764+4	0.1	0.01	0.032

Vamp. 9.0: 2.80GHz, 1GB memory, 600s time limit
LEO+Vamp.: 2.40GHz, 4GB memory, 120s time limit
LEO-II+E: 1.60GHz, 1GB memory, 60s time limit

Solving Less Lightweight Problems



Modal Logics Challenge

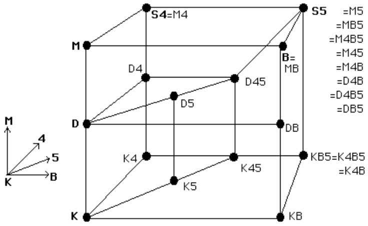


John Halleck (U Utah):
`http://www.cc.utah.edu/~nahaj/`

$100 Modal Logic Challenge:
`www.tptp.org`

Logic Systems Interrelationships

Modal Logics Challenge



John Halleck (U Utah):
`http://www.cc.utah.edu/~nahaj/`

$100 Modal Logic Challenge:
`www.tptp.org`

Example

$$
\begin{aligned}
S4 &= K \\
&+ M(T): \quad \Box_R A \Rightarrow A \\
&+ 4: \quad \Box_R A \Rightarrow \Box_R \Box_R A
\end{aligned}
$$

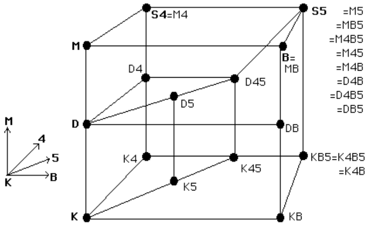
Theorems:

$$S4 \not\subseteq K \tag{1}$$

$$(M \wedge 4) \Leftrightarrow (\mathit{refl.}(R) \wedge \mathit{trans.}(R)) \tag{2}$$

Logic Systems Interrelationships

Modal Logics Challenge



John Halleck (U Utah):
`http://www.cc.utah.edu/~nahaj/`

$100 Modal Logic Challenge:
`www.tptp.org`

Example

$$
\begin{array}{rcll}
S4 & = & K & \\
 & + & M(T): & \Box_R A \Rightarrow A \\
 & + & 4: & \Box_R A \Rightarrow \Box_R \Box_R A
\end{array}
$$

Theorems:

$$
\begin{array}{rcll}
S4 & \not\subseteq & K & (1) \\
(M \wedge 4) & \Leftrightarrow & (refl.(R) \wedge trans.(R)) & (2)
\end{array}
$$

Experiments

	FO-ATPs **[SutcliffeEtal-07]**	LEO-II + E **[BePa-08]**
(1)	16min + 2710s	17.3s
(2)	???	2.4s

Even simpler: Reasoning within Multimodal Logics

Problem	LEO-II + E
valid($\Box_r\ \top$)	0.025s
valid($\Box_r\ a \Rightarrow \Box_r\ a$)	0.026s
valid($\Box_r\ a \Rightarrow \Box_s\ a$)	–
valid($\Box_s\ (\Box_r\ a \Rightarrow \Box_r\ a)$)	0.026s
valid($\Box_r\ (a \wedge b) \Leftrightarrow (\Box_r\ a \wedge \Box_r\ b)$)	0.044s
valid($\Diamond_r\ (a \Rightarrow b) \Rightarrow \Box_r\ a \Rightarrow \Diamond_r\ b$)	0.030s
valid($\neg \Diamond_r\ a \Rightarrow \Box_r\ (a \Rightarrow b)$)	0.029s
valid($\Box_r\ b \Rightarrow \Box_r\ (a \Rightarrow b)$)	0.026s
valid($(\Diamond_r\ a \Rightarrow \Box_r\ b) \Rightarrow \Box_r\ (a \Rightarrow b)$)	0.027s
valid($(\Diamond_r\ a \Rightarrow \Box_r\ b) \Rightarrow (\Box_r\ a \Rightarrow \Box_r\ b)$)	0.029s
valid($(\Diamond_r\ a \Rightarrow \Box_r\ b) \Rightarrow (\Diamond_r\ a \Rightarrow \Diamond_r\ b)$)	0.030s

(Normal) Multimodal Logic in HOL

Simple, Straightforward Encoding of Multimodal Logic

- base type ι: set of possible worlds
  certain terms of type $\iota \rightarrow o$: multimodal logic formulas
- multimodal logic operators:

$$
\begin{aligned}
\neg_{(\iota\rightarrow o)\rightarrow(\iota\rightarrow o)} &= \lambda A_{\iota\rightarrow o} \centerdot (\lambda x_\iota \centerdot \neg A(x)) \\
\vee_{(\iota\rightarrow o)\rightarrow(\iota\rightarrow o)\rightarrow(\iota\rightarrow o)} &= \lambda A_{\iota\rightarrow o}, B_{\iota\rightarrow o} \centerdot (\lambda x_\iota \centerdot A(x) \vee B(x)) \\
\Box_{R\,(\iota\rightarrow\iota\rightarrow o)\rightarrow(\iota\rightarrow o)\rightarrow(\iota\rightarrow o)} &= \lambda R_{\iota\rightarrow\iota\rightarrow o}, A_{\iota\rightarrow o} \centerdot \\
&\qquad (\lambda x_\iota \centerdot \forall y_\iota \centerdot R(x,y) \Rightarrow A(y))
\end{aligned}
$$

Related Work

[Gallin-73], [Carpenter-98], [Merz-99],
[Brown-05], [Hardt&Smolka-07], [Kaminski&Smolka-07]

Encoding of Validity

$$\texttt{valid} \quad := \quad \lambda A_{\iota\rightarrow o} \centerdot (\forall w_{\iota} \centerdot A(w))$$

Example Proof: $\mathbf{valid}(\Box_s (\Box_r a \Rightarrow \Box_r a))$

Initialisation of problem

$$\neg\mathbf{valid}(\Box_s (\Box_r a \Rightarrow \Box_r a))$$

Example Proof: $\textbf{valid}(\Box_s (\Box_r a \Rightarrow \Box_r a))$

Initialisation of problem

$$\neg\textbf{valid}(\Box_s (\Box_r a \Rightarrow \Box_r a))$$

Definition expansion

$$\neg(\forall x_\iota \centerdot \forall y_\iota \centerdot \neg s(x,y) \vee ((\neg(\forall u_\iota \centerdot \neg r(y,u) \vee a(u))) \vee (\forall v_\iota \centerdot \neg r(y,v) \vee a(v)))$$

Example Proof: $\textbf{valid}(\Box_s (\Box_r a \Rightarrow \Box_r a))$

Initialisation of problem

$$\neg\textbf{valid}(\Box_s (\Box_r a \Rightarrow \Box_r a))$$

Definition expansion

$$\neg(\forall x_\iota\text{.}\forall y_\iota\text{.}\neg s(x, y) \vee ((\neg(\forall u_\iota\text{.}\neg r(y, u) \vee a(u))) \vee (\forall v_\iota\text{.}\neg r(y, v) \vee a(v)))$$

Normalisation (x, y, u are now Skolem constants, V is a variable)

$$\begin{array}{ll} s(x, y) & \neg a(u) \\ r(y, u) & a(V) \vee \neg r(y, V) \end{array}$$

Example Proof: valid($\Box_s (\Box_r a \Rightarrow \Box_r a)$)

Initialisation of problem

$$\neg\mathbf{valid}(\Box_s (\Box_r a \Rightarrow \Box_r a))$$

Definition expansion

$$\neg(\forall x_\iota \blacksquare \forall y_\iota \blacksquare \neg s(x, y) \vee ((\neg(\forall u_\iota \blacksquare \neg r(y, u) \vee a(u))) \vee (\forall v_\iota \blacksquare \neg r(y, v) \vee a(v)))$$

Normalisation (x, y, u **are now Skolem constants,** V **is a variable)**

$$\begin{array}{ll} s(x, y) & \neg a(u) \\ r(y, u) & a(V) \vee \neg r(y, V) \end{array}$$

Translation to first-order logic

$$\begin{array}{ll} @_{(io)\_i}(@_{(i(io))\_i}(s, x), y) & \neg @_{(\iota o)\_\iota}(a, u) \\ @_{(io)\_i}(@_{(i(io))\_i}(r, y), u) & @_{(\iota o)\_\iota}(a, V) \vee \neg @_{(io)\_i}(@_{(i(io))\_i}(r, y), V) \end{array}$$

More Examples ...

A simple equation between modal logic formulas

$$\forall R_\blacksquare \forall A_\blacksquare \forall B_\blacksquare (\Box_R (A \vee B)) = (\Box_R (B \vee A))$$

- initialisation, definition expansion and normalisation:

$$(\lambda X_\iota . \forall Y_{\iota\blacksquare} \neg((r\, X)\, Y) \vee (a\, Y) \vee (b\, Y))$$
$$\neq$$
$$(\lambda X_\iota . \forall Y_{\iota\blacksquare} \neg((r\, X)\, Y) \vee (b\, Y) \vee (a\, Y))$$

More Examples . . .

A simple equation between modal logic formulas

$$\forall R_\bullet \forall A_\bullet \forall B_\bullet (\Box_R (A \vee B)) = (\Box_R (B \vee A))$$

- functional and Boolean extensionality:

$$\neg((\forall Y_{\iota\bullet} \neg((r\,w)\,Y) \vee (a\,Y) \vee (b\,Y))$$
$$\Leftrightarrow$$
$$(\forall Y_{\iota\bullet} \neg((r\,w)\,Y) \vee (b\,Y) \vee (a\,Y)))$$

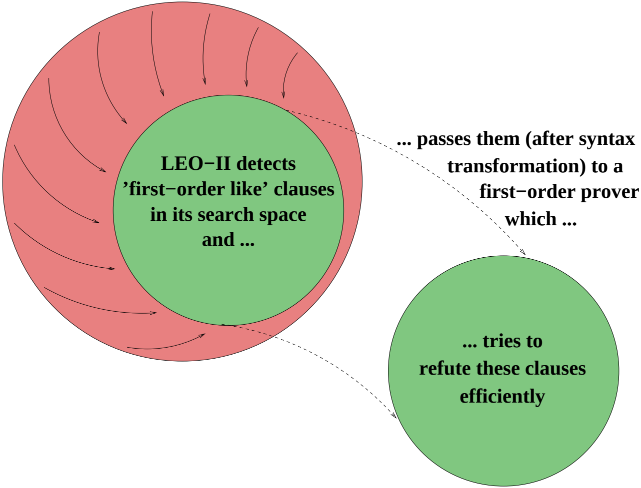
More Examples ...

A simple equation between modal logic formulas

$$\forall R. \forall A. \forall B. (\Box_R (A \vee B)) = (\Box_R (B \vee A))$$

- normalisation:
  $40 : (b\,V) \vee (a\,V) \vee \neg((r\,w)\,V) \vee \neg((r\,w)\,W) \vee (b\,W) \vee (a\,W)$
  $41 : ((r\,w)\,z) \vee ((r\,w)\,v)$
  $42 : \neg(a\,z) \vee ((r\,w)\,v)$
  $43 : \neg(b\,z) \vee ((r\,w)\,v)$
  $44 : ((r\,w)\,z) \vee \neg(a\,v)$
  $45 : \neg(a\,z) \vee \neg(a\,v)$
  $46 : \neg(b\,z) \vee \neg(a\,v)$
  $47 : ((r\,w)\,z) \vee \neg(b\,v)$
  $48 : \neg(a\,z) \vee \neg(b\,v)$
  $49 : \neg(b\,z) \vee \neg(b\,v)$
- total proving time (notebook with 1.60GHz, 1GB): 0.071s

Architecture of LEO-II



More Examples ...

A simple equation between modal logic formulas

$$\forall R_\blacksquare \forall A_\blacksquare \forall B_\blacksquare (\Box_R (A \vee B)) \doteq (\Box_R (B \vee A))$$

where $\doteq$ is defined as $\lambda X, Y.\forall P_\blacksquare (P\, X) \Rightarrow (P\, Y)$

- initialisation, definition expansion and normalisation:

$$(p\,(\lambda X_\iota.\forall Y_{\iota}{}_\blacksquare \neg((r\, X)\, Y) \vee (a\, Y) \vee (b\, Y)))$$
$$\neg(p\,(\lambda X_\iota.\forall Y_{\iota}{}_\blacksquare \neg((r\, X)\, Y) \vee (b\, Y) \vee (a\, Y)))$$

More Examples . . .

A simple equation between modal logic formulas

$$\forall R_{\blacksquare} \forall A_{\blacksquare} \forall B_{\blacksquare} (\Box_R (A \vee B)) \doteq (\Box_R (B \vee A))$$

where $\doteq$ is defined as $\lambda X, Y. \forall P_{\blacksquare} (P\, X) \Rightarrow (P\, Y)$

- resolution:

$$(p\, (\lambda X_\iota . \forall Y_{\iota \blacksquare} \neg ((r\, X)\, Y) \vee (a\, Y) \vee (b\, Y)))$$
$$\neq$$
$$(p\, (\lambda X_\iota . \forall Y_{\iota \blacksquare} \neg ((r\, X)\, Y) \vee (b\, Y) \vee (a\, Y)))$$

More Examples . . .

A simple equation between modal logic formulas

$$\forall R_\blacksquare \forall A_\blacksquare \forall B_\blacksquare (\Box_R (A \lor B)) \doteq (\Box_R (B \lor A))$$

where $\doteq$ is defined as $\lambda X, Y. \forall P_\blacksquare (P\, X) \Rightarrow (P\, Y)$

- decomposition:

$$(\lambda X_\iota . \forall Y_{\iota\blacksquare} \neg((r\, X)\, Y) \lor (a\, Y) \lor (b\, Y))$$
$$\neq$$
$$(\lambda X_\iota . \forall Y_{\iota\blacksquare} \neg((r\, X)\, Y) \lor (b\, Y) \lor (a\, Y))$$

More Examples . . .

A simple equation between modal logic formulas

$$\forall R_\blacksquare \forall A_\blacksquare \forall B_\blacksquare (\Box_R (A \vee B)) \doteq (\Box_R (B \vee A))$$

where $\doteq$ is defined as $\lambda X, Y. \forall P_\blacksquare (P\,X) \Rightarrow (P\,Y)$

- functional and Boolean extensionality:

$$\neg((\forall Y_{\iota\blacksquare} \neg((r\,w)\,Y) \vee (a\,Y) \vee (b\,Y))$$
$$\Leftrightarrow$$
$$(\forall Y_{\iota\blacksquare} \neg((r\,w)\,Y) \vee (b\,Y) \vee (a\,Y)))$$

More Examples ...

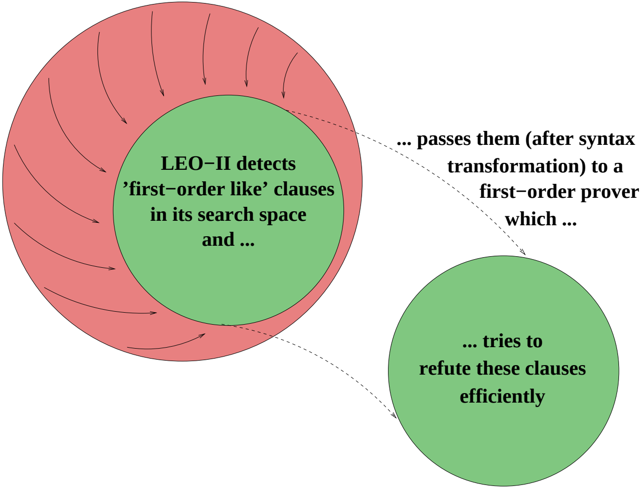
A simple equation between modal logic formulas

$$\forall R_{\blacksquare}\forall A_{\blacksquare}\forall B_{\blacksquare}(\Box_R\,(A \vee B)) \doteq (\Box_R\,(B \vee A))$$

where $\doteq$ is defined as $\lambda X, Y.\forall P_{\blacksquare}(P\,X) \Rightarrow (P\,Y)$

- normalisation: ...see previous example ...
- total proving time is $0.166s$

Architecture of LEO-II



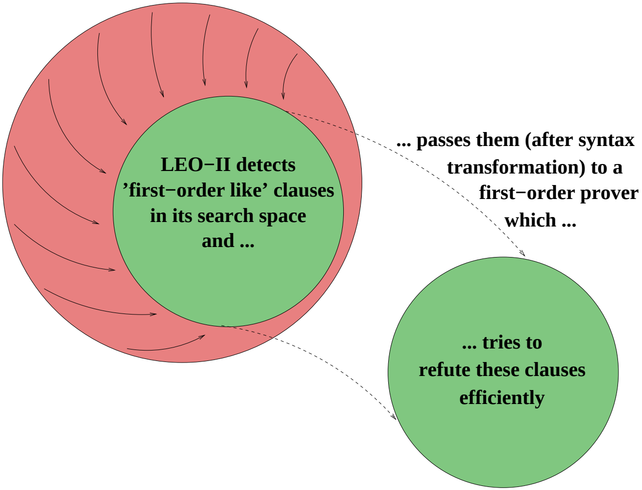
More Examples ...

In modal logic **K**, the axioms T and 4 are equivalent to reflexivity and transitivity of the accessibility relation R

$$\forall R.(\forall A.\mathtt{valid}(\Box_R A \Rightarrow A) \wedge \mathtt{valid}(\Box_R A \Rightarrow \Box_R \Box_R A)) \\ \Leftrightarrow (\mathtt{reflexive}(R) \wedge \mathtt{transitive}(R))$$

- processing in LEO-II analogous to previous example
- now 70 clauses are passed to E
- E generates 21769 clauses before finding the empty clause
- total proving time 2.4s
- this proof cannot be found in LEO-II alone

Architecture of LEO-II



More Examples . . .

$S4 \not\subseteq K$: Axioms T and 4 are not valid in modal logic $\mathbf{K}$

$$\neg\forall R. \forall A. \forall B. (\mathrm{valid}(\Box_R A \Rightarrow A)) \wedge (\mathrm{valid}(\Box_R B \Rightarrow \Box_R \Box_R B))$$

- LEO-II shows that axiom T is not valid
- R is instantiated with $\neq$ via primitive substitution
- total proving time 17.3s

More Examples . . .

$\mathtt{T} \not\subseteq \mathtt{K}$: Axiom T is not valid in modal logic K

$$\neg \forall R. \forall A. (\mathrm{valid}(\Box_R A \Rightarrow A))$$

- initialisation, definition expansion and normalization generates:

$$((R\,W)\,s^{A,W,R}) \vee (A\,W)$$
$$\neg(A\,s^{A,W,R}) \vee (A\,W)$$

where $s^{A,W,R} = (((s\,A)\,W)\,R)$ is a new Skolem term

More Examples . . .

$\mathtt{T} \not\subseteq \mathtt{K}$: Axiom T is not valid in modal logic **K**

$$\neg\forall R.\forall A.(\mathrm{valid}(\Box_R A \Rightarrow A))$$

- the refutation employs only the former clause

$$((R\,W)\,s^{A,W,R}) \vee (A\,W)$$

More Examples ...

$\texttt{T} \not\subseteq \texttt{K}$: Axiom T is not valid in modal logic **K**

$$\neg\forall R. \forall A. (\mathrm{valid}(\Box_R\, A \Rightarrow A))$$

- $((R\, W)\, s^{A,W,R}) \vee (A\, W)$
- LEO-II 'guesses' the instantiations

$$R \leftarrow \lambda X, Y.((M\, X)\, Y) \neq ((N\, X)\, Y)$$
$$A \leftarrow \lambda X.(O\, X) \neq (P\, X)$$

with primitive substitution rule (M, N, O, P are new free variables) ...

More Examples . . .

$\mathtt{T} \not\subseteq \mathtt{K}$: Axiom T is not valid in modal logic K

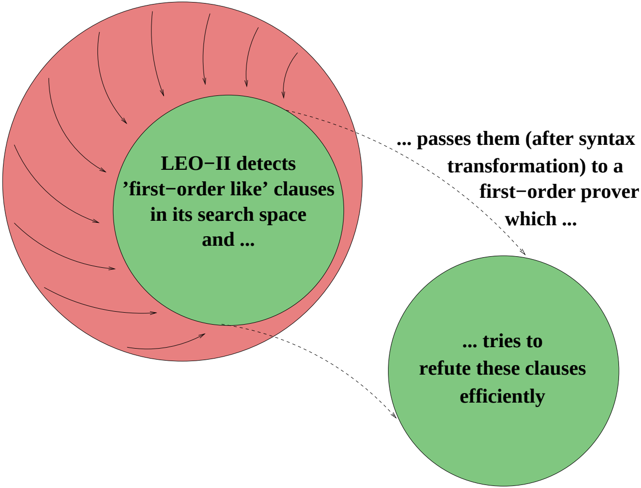
$$\neg\forall R.\forall A.(\text{valid}(\Box_R\, A \Rightarrow A))$$

- . . . and applies them

$$((M\,(R\,W))\,s^{A,W,R}) \neq ((N\,(R\,W))\,s^{A,W,R})$$
$$\vee$$
$$(O\,W) \neq (P\,W)$$

- such flex-flex unification constraints are always solvable!
- total proving time 9.0s

Architecture of LEO-II



Limitations

LEO-II cannot prove the following example:

Modal logic $K4$ (which adds only axiom 4 to K) is not entailed in K:

$$\neg\forall R_\blacksquare \forall B_\blacksquare (\text{valid}(\Box_R B \Rightarrow \Box_R \Box_R B))$$

Limitations

LEO-II also cannot prove this related example:

$$\neg \forall R. \mathrm{trans}(R)$$

Limitations

LEO-II also cannot prove this related example:

$$\neg\forall R_{\bullet}\mathrm{trans}(R)$$

- reason: not a theorem; domain of possible worlds may well just consist of a single world w.

Limitations

LEO-II also cannot prove this related example:

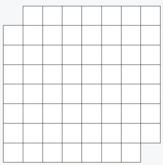
$$\neg\forall R.\mathrm{trans}(R)$$

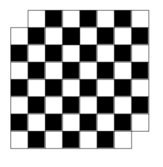
- reason: not a theorem; domain of possible worlds may well just consist of a single world w.
- LEO-II can in fact prove the latter example under the additional assumption

$$\neg\forall X.\forall Y.X = Y$$

Representation (and the right System Architecture) Matters!

A general lesson in AI . . .





. . . and a specific lesson here

FOL	HOL
+	+
FO-ATP	LEO-II + FO-ATP

... there is much left to be done!

LEO-II

- Equational Reasoning
- Termindexing
- Handling of Definitions

... there is much left to be done!

LEO-II

- Equational Reasoning
- Termindexing
- Handling of Definitions

Cooperat. with Specialist Reasoners

- Monadic Second-Order Logic, Prop. Logic, Arithmetic, . . .
- Logic Translations
- Feedback for LEO-II
- Proof Transf./Verification
- Agent-based Architecture

… there is much left to be done!

LEO-II

- Equational Reasoning
- Termindexing
- Handling of Definitions

Cooperat. with Specialist Reasoners

- Monadic Second-Order Logic, Prop. Logic, Arithmetic, . . .
- Logic Translations
- Feedback for LEO-II
- Proof Transf./Verification
- Agent-based Architecture

Integration into Proof Assistants

- Relevance of Axioms
- Proof Transf./Verification

... there is much left to be done!

LEO-II

- Equational Reasoning
- Termindexing
- Handling of Definitions

Cooperat. with Specialist Reasoners

- Monadic Second-Order Logic, Prop. Logic, Arithmetic, ...
- Logic Translations
- Feedback for LEO-II
- Proof Transf./Verification
- Agent-based Architecture

Integration into Proof Assistants

- Relevance of Axioms
- Proof Transf./Verification

International Infrastructure

- TPTP Language(s) for HOL
- Repository of Proof Problems
- HOL Prover Contest

... there is much left to be done!

LEO-II

- Equational Reasoning
- Termindexing
- Handling of Definitions

Cooperat. with Specialist Reasoners

- Monadic Second-Order Logic, Prop. Logic, Arithmetic, ...
- Logic Translations
- Feedback for LEO-II
- Proof Transf./Verification
- Agent-based Architecture

Integration into Proof Assistants

- Relevance of Axioms
- Proof Transf./Verification

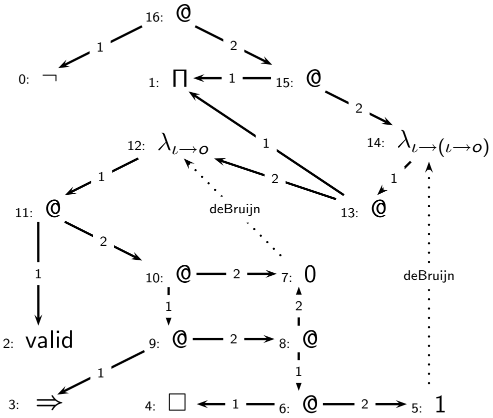
International Infrastructure

- TPTP Language(s) for HOL
- Repository of Proof Problems
- HOL Prover Contest

Applications

Logic System Interrelationships, Ontology Reasoning (SUMO, CYC), Formal Methods, CL, ...

Term Graph for: $\neg\forall R.\forall A.(\mathbf{valid}(\Box_R A \Rightarrow A))$



Term graph videos: `http://www.ags.uni-sb.de/~chris/leo/art`

DIALOG - Tutorial Natural Language Dialog on Proofs

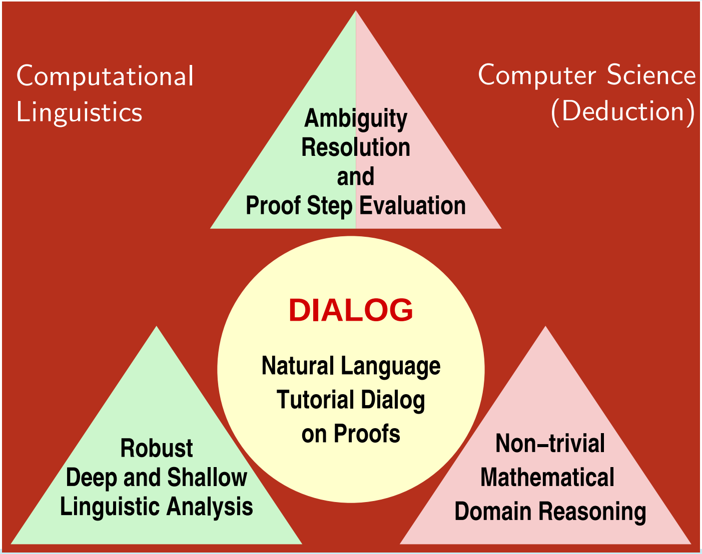
The DIALOG Project

Tutorial NL Dialog for Mathematical Proofs.

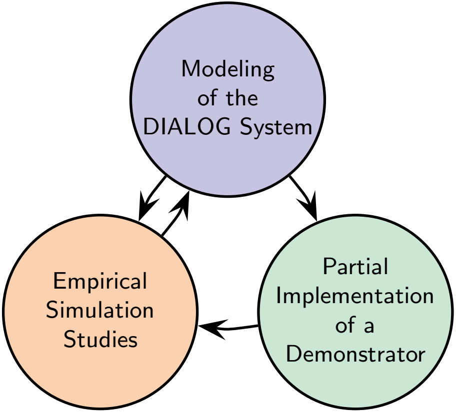
- Natural language analysis
- Mathematical domain reasoning (using proof assistant OMEGA)
- Dialog management
- Output generation and verbalization

Assume that $a \in X$.
If $X \cap Y = \emptyset$,
then $a \notin Y$.

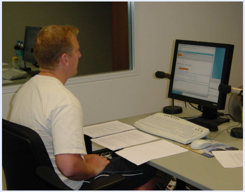
The DIALOG Project in the SFB 378



Progressive Refinement



Empirical Investigations



Let R, S and T be relations in an arbitrary set M. It holds: $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$. Do the proof interactively with the system!

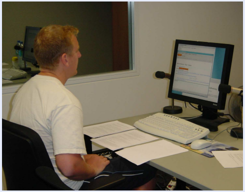
Let $(x, y) \in (R \cup S) \circ T$

Correct! Good start!

Then $\exists z$ such that (x, z) in $(R \cup S)$ and (z, y) in T

Correct!

Empirical Investigations (Wizard-of-Oz)



Let R, S and T be relations in an arbitrary set M. It holds: $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$. Do the proof interactively with the system!

Let $(x, y) \in (R \cup S) \circ T$

Correct! Good start!

Then $\exists z$ such that (x, z) in $(R \cup S)$ and (z, y) in T

Correct!

Wizard Room

1 Audio Recording
2 Video Recording
3 Experimenter
4 Overall Control
5 Wizard
6 Wizard GUI

Dialog Excerpt I (VM group)

S33: Nach Aufgabe W ist $(S \circ (S{\cup}R)^{-1})^{-1} = ((S{\cup}R)^{-1})^{-1} \circ S^{-1}$
T34: Das ist richtig!
S34: Dies ist nach Theorem 1 gleich $(S{\cup}R) \circ S^{-1}$
T35: Das ist auch richtig!
S35: Ein Element (a,b) ist genau dann in dieser Menge, wenn es ein $z{\in}M$ gibt mit $(a,z){\in}S{\cup}R$ und $(z,b){\in}S^{-1}$
T36: Das ist korrekt!
S36: Also $(b,z){\in}S$.
T37: Auch das ist korrekt! **S37:** DAmit [sic] kann aber nicht die geforderte Gleichheit bewiesen werden. **T38:** Das haben Sie gut erkannt! Was fehlt Ihnen dazu?

Dialog Excerpt I (FM group)

S25: $(R \cup S) \circ S = \{(x,y) \mid \exists\, z(z \in M \wedge ((x,z) \in R \vee (x,z) \in S) \wedge (z,y) \in S\}$
T28: Diese Beziehung hat Relevanz fuer die Aufgabe. Sie muessen sich also etwas dabei gedacht haben, als Sie diese vermuteten. Koennen Sie begruenden, wie sie mit der Aufgabe zusammenhaengt?
T29: Ihre Formel ist richtig.
T30: Koennen Sie die Behauptung unter zusaetzlichen Voraussetzungen beweisen?
S26: $\{(x,y) \mid \exists\, z(z \in M \wedge ((x,z) \in R \vee (x,z) \in S) \wedge (z,y) \in S\} = \{(x,y) \mid \exists\, z(z \in M \wedge (y,z) \in S \wedge ((x,z) \in R \vee (x,z) \in S))\} \Leftrightarrow ((y,z) \in S \wedge (z,y) \in S)\}$
T31: Auf der rechten Seite ist z nicht spezifiziert. Meinten Sie vielleicht $\Leftrightarrow$ oder etwas Aehnliches statt $\wedge$?

Corpora and more

http://www.ags.uni-sb.de/~dialog

Correctness, Granularity and Relevance

student: $(x, y) \in (R \circ S)^{-1}$
tutor: Now try to draw inferences from that!
student: $(x, y) \in S^{-1} \circ R^{-1}$
tutor: One cannot directly deduce that.

You need some intermediate steps!

correct	too coarse-grained	relevant

student: $(x, y) \in (R \circ S)^{-1}$ if according to the inverse relation it holds that $(y, x) \in (R \circ S)$
tutor: That is correct, but try to use $(x, y) \in (R \circ S)^{-1}$ as a precondition.

correct	appropriate	limited relevance

Research Challenges

Perspective of Mathematical Domain Reasoning (MDR):

- Support for resolution of Ambiguities and Underspecification
- Proof Step Evaluation
  - Soundness: proof step verifiable by formal system?
  - Granularity: size/argumentative complexity of proof step?
  - Relevance: proof step needed/useful in achieving the goal?

Perspective of NL Analysis: [... not in this talk ...]

Perspective of Dialog Management: [... not in this talk ...]

Perspective of Tutoring Proofs: [... not in this talk ...]

Research Challenges

Perspective of Mathematical Domain R… ning (MDR):

- Support for resolution of Am… s and Underspecification
- Proof Step Evaluation
  - Soundness: pr… verifiable by formal system?
  - Granulari… argumentative complexity of proof step?
  - Relevance… proof step needed/useful in achieving the goal?

Logical vs Tutorial Dimension

Perspective of NL Analysis: [. . . not in this talk . . .]

Perspective of Dialog Management: [. . . not in this talk . . .]

Perspective of Tutoring Proofs: [. . . not in this talk . . .]

Proof Step Evaluation

(DM-1) ...
(DM-2) ...

?

(G) ...

Given: (DM-1) $\overline{X \cup Y} = \overline{X} \cap \overline{Y}$
(DM-2) $\overline{X \cap Y} = \overline{X} \cup \overline{Y}$

Task: Please show $\overline{(A \cup B) \cap (C \cup D)} = (\overline{A} \cap \overline{B}) \cup (\overline{C} \cap \overline{D})$

New: By deMorgan $\overline{(A \cup B) \cap (C \cup D)} = \overline{(A \cup B)} \cup \overline{(C \cup D)}$.

Proof Step Evaluation

(DM-1) ...
(DM-2) ...

?

(G) ...

Given: (DM-1) $\overline{X \cup Y} = \overline{X} \cap \overline{Y}$
(DM-2) $\overline{X \cap Y} = \overline{X} \cup \overline{Y}$

Task: Please show $\overline{(A \cup B) \cap (C \cup D)} = (\overline{A} \cap \overline{B}) \cup (\overline{C} \cap \overline{D})$

New: By deMorgan $\overline{(A \cup B) \cap (C \cup D)} = \overline{(A \cup B)} \cup \overline{(C \cup D)}$.

(DM-1) ...
(DM-2) ...
(New) ... $\downarrow$

?

(G) ...

Soundness: yes
Granularity: 1x(DM-2)
Relevance: yes

(DM-1) ...
(DM-2) ...

?

(New) ... $\uparrow$
(G) ...

Soundness: yes
Granularity: 2x(DM-1)
Relevance: yes

Proof Step Evaluation: How?

New:

We show E.

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$

?

(G) $D \vee E$

(1) . . .
(2) . . .
(3) . . .
(4) . . .

?

(G') E
(G) . . .

PSE:

Soundness

Granularity

Relevance

Proof Step Evaluation: How?

New:

We show E.

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$

?

(G) $D \vee E$

(1) ...
(2) ...
(3) ...
(4) ...

?

(G') E
(G) ...

PSE:

Soundness

- $(G') \vdash^{?} (G)$
- any proof

Granularity

Relevance

Proof Step Evaluation: How?

New:

We show E.

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$

?

(G) $D \vee E$

(1) ...
(2) ...
(3) ...
(4) ...

?

(G') E
(G) ...

PSE:

Soundness

- $(G') \vdash^{?} (G)$
- any proof

Granularity

- size-of($(G') \vdash^{?} (G)$)
- cognitively adequate proofs

Relevance

Proof Step Evaluation: How?

New:

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$

?

(G) $D \vee E$

We show E.

(1) ...
(2) ...
(3) ...
(4) ...

?

(G') E
(G) ...

PSE:

Soundness

- $(G') \vdash^? (G)$
- any proof

Granularity

- size-of($(G') \vdash^? (G)$)
- cognitively adequate proofs

Relevance

- $(1), (2), (3), (4) \vdash^? (G')$
- detours?, shorter proofs?

Example: Assertion Level Proof Reconstruction

1. Tutor: Show $(r \circ s)^{-1} = s^{-1} \circ r^{-1}$!

Exercise: $\vdash (r \circ s)^{-1} = s^{-1} \circ r^{-1}$

Example: Assertion Level Proof Reconstruction

1. Tutor: Show $(r \circ s)^{-1} = s^{-1} \circ r^{-1}$!
2. Student: Let $(x, y) \in (r \circ s)^{-1}$.

$$\Delta_1$$

$$\dfrac{\dfrac{\dfrac{}{\texttt{s1:}\quad (x,y) \in (r \circ s)^{-1} \vdash (x,y) \in s^{-1} \circ r^{-1}}}{\vdash (r \circ s)^{-1} \subseteq s^{-1} \circ r^{-1}}\,\text{Def.}\subseteq \qquad \vdash s^{-1} \circ r^{-1} \subseteq (r \circ s)^{-1}}{\texttt{Exercise:}\quad \vdash (r \circ s)^{-1} = s^{-1} \circ r^{-1}}\,\text{Def.} =$$

Example: Assertion Level Proof Reconstruction

1. Tutor: Show $(r \circ s)^{-1} = s^{-1} \circ r^{-1}$!
2. Student: Let $(x, y) \in (r \circ s)^{-1}$.
3. Student: Hence $(y, x) \in (r \circ s)$.

Δ_2

s2: $(y, x) \in (r \circ s) \vdash$ "—"

Def. $^{-1}$

Δ_1

s1: $(x, y) \in (r \circ s)^{-1} \vdash (x, y) \in s^{-1} \circ r^{-1}$

Def. $\subseteq$

$\vdash (r \circ s)^{-1} \subseteq s^{-1} \circ r^{-1}$ $\quad$ $\vdash s^{-1} \circ r^{-1} \subseteq (r \circ s)^{-1}$

Def. $=$

Exercise: $\vdash (r \circ s)^{-1} = s^{-1} \circ r^{-1}$

Example: Assertion Level Proof Reconstruction

1. Tutor: Show $(r \circ s)^{-1} = s^{-1} \circ r^{-1}$!
2. Student: Let $(x, y) \in (r \circ s)^{-1}$.
3. Student: Hence $(y, x) \in (r \circ s)$.
4. Student: Hence $(y, z) \in r \wedge (z, x) \in s$.

Δ_3 s3: $(y, z) \in r \wedge (z, x) \in s \vdash$ "—"

Def.$\circ$

Δ_2 s2: $(y, x) \in (r \circ s) \vdash$ "—"

Def.$^{-1}$

Δ_1

s1: $(x, y) \in (r \circ s)^{-1} \vdash (x, y) \in s^{-1} \circ r^{-1}$

Def. $\subseteq$

$\vdash (r \circ s)^{-1} \subseteq s^{-1} \circ r^{-1}$ $\qquad$ $\vdash s^{-1} \circ r^{-1} \subseteq (r \circ s)^{-1}$

Def. $=$

Exercise: $\vdash (r \circ s)^{-1} = s^{-1} \circ r^{-1}$

Example: Assertion Level Proof Reconstruction

1. Tutor: Show $(r \circ s)^{-1} = s^{-1} \circ r^{-1}$!
2. Student: Let $(x, y) \in (r \circ s)^{-1}$.
3. Student: Hence $(y, x) \in (r \circ s)$.
4. Student: Hence $(y, z) \in r \land (z, x) \in s$.
5. Student: Hence $(z, y) \in r^{-1} \land (x, z) \in s^{-1}$.

$$\vdots$$

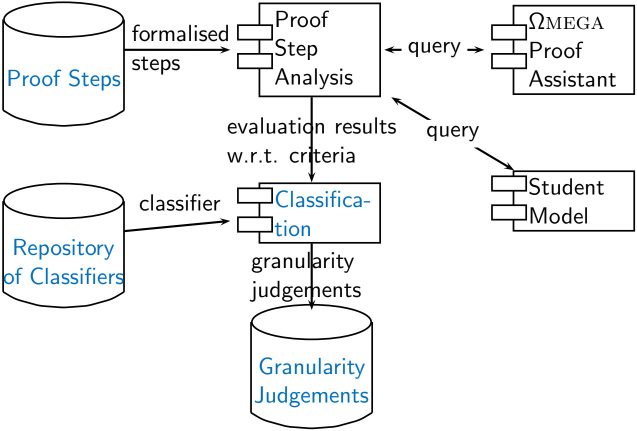
$$\Delta_4 \qquad \dfrac{\dfrac{\vdots}{\texttt{s4:}\quad (z,y) \in r^{-1} \land (x,z) \in s^{-1} \vdash \text{"—"}}}{(y,z) \in r \land (x,z) \in s^{-1} \vdash \text{"—"}}\ \text{Def}^{-1}$$

$$\Delta_3 \qquad \dfrac{(y,z) \in r \land (x,z) \in s^{-1} \vdash \text{"—"}}{\texttt{s3:}\quad (y,z) \in r \land (z,x) \in s \vdash \text{"—"}}\ \text{Def.}^{-1}$$

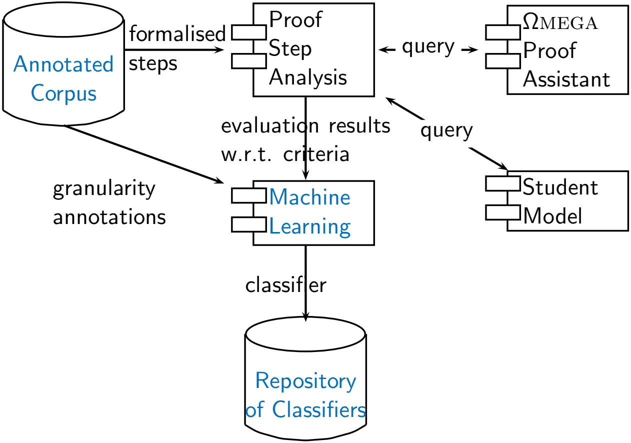
$$\Delta_2 \qquad \dfrac{\texttt{s3:}\quad (y,z) \in r \land (z,x) \in s \vdash \text{"—"}}{\texttt{s2:}\quad (y,x) \in (r \circ s) \vdash \text{"—"}}\ \text{Def.}\circ$$

$$\Delta_1 \qquad \dfrac{\dfrac{\dfrac{\texttt{s2:}\quad (y,x) \in (r \circ s) \vdash \text{"—"}}{\texttt{s1:}\quad (x,y) \in (r \circ s)^{-1} \vdash (x,y) \in s^{-1} \circ r^{-1}}\ \text{Def.}^{-1}}{\vdash (r \circ s)^{-1} \subseteq s^{-1} \circ r^{-1}}\ \text{Def.} \subseteq \qquad \vdash s^{-1} \circ r^{-1} \subseteq (r \circ s)^{-1}}{\texttt{Exercise:}\quad \vdash (r \circ s)^{-1} = s^{-1} \circ r^{-1}}\ \text{Def.} =$$

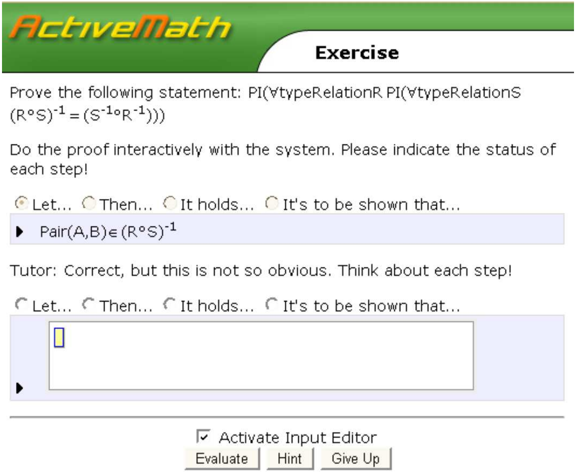
Judgment Module



Trainer Module



Integration with ActiveMath



Conclusion

Finding 'a proof' automatically		Reasoning about human-constructed proof(step)s
Essential criterion: soundness	vs	Novel criteria: granularity & relevance
Traditional TP challenges: completeness, efficiency		Novel TP challenges: qualitative aspects of proofs, enumeration of proofs

PLATO – Mediator between Texteditors and Proof Assistants