

Nancy, France, April 14th 2005



Can a Higher-Order and a First-Order Theorem Prover Cooperate?

Christoph Benzmüller (Saarland U, Germany)

joint work with:

Volker Sorge (U Birmingham, UK)

Manfred Kerber (U Birmingham, UK)

Mateja Jamnik (U Cambridge, UK)

Overview: Issues of this Talk



- Computer-supported Mathematics
- Automation of Mathematical Reasoning
- Automation of Higher-Order Theorem Proving (HOTP)
- Architectures supporting System Integrations
- Problem Libraries such as TPPTP

Computer Math: Representation Matters

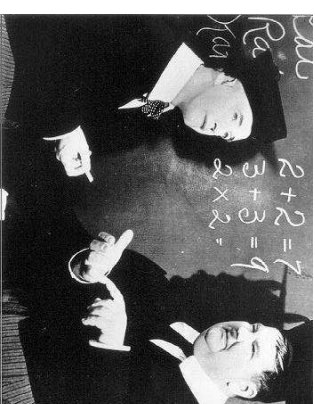


- Computer-supported Mathematics / Mathematics Assistance Systems
 - full automatization not realistic and only partly desirable
 - support for collaboration mathematician and computer is needed
 - interaction should be based on expressive languages
 - fact: maths in practice uses higher-order constructs
 - fact also: prominent proof assistants already support higher-order logic

- Example:

	textbooks	higher-order logic	first-order logic
$\mathcal{P}(A)$	$\{x x \subseteq A\}$ $\mathcal{P}(\emptyset)$ is finite	$\lambda x.x \subseteq A$ finite($\mathcal{P}(\emptyset)$)	$x \in \mathcal{P}(A) \Leftrightarrow x \subseteq A$... less nice ...
$\text{Im}(F, A)$	$\{y \exists x.x \in A \wedge y = F(x)\}$	$\lambda y.\exists x.x \in A \wedge y = F(x)$	see TPTP (terrible)

- Idea: Start with higher-order representations in a mathematics assistance system and **combine** higher-order and first-order (and propositional) reasoning (supported by transformational mappings)



- Test Problems:
 - 45 theorems on sets, relations, and functions
 - taken from the TPTP domain "SET"
 - also used in paper on Saturate system [GanzingerStuber-IJCAR-04]
 - we added some problems that cannot be solved by any FOTP
- Conciseness of Higher-Order Representations:
 - 45 problem formulations (required definitions + theorems) fit on 1,5 page
 - not possible in first-order without λ -abstraction

- Examples of Basic Definitions on Sets and Relations

$- \in -$	$::= \lambda x, A. [Ax]$
\emptyset	$::= [\lambda x. \perp]$
$- \cap -$	$::= \lambda A, B. [\lambda x. x \in A \wedge x \in B]$
$- \cup -$	$::= \lambda A, B. [\lambda x. x \in A \vee x \in B]$
$- \setminus -$	$::= \lambda A, B. [\lambda x. x \in A \vee x \notin B]$
Meets($-, -$)	$::= \lambda A, B. [\exists x. x \in A \wedge x \in B]$
Pair($-, -$)	$::= \lambda x, y. [\lambda u, v. u = x \wedge v = y]$
$- \times -$	$::= \lambda A, B. [\lambda u, v. u \in A \wedge v \in B]$
Subrel($-, -$)	$::= \lambda R, Q. [\forall x, y. Rxy \Rightarrow Qxy]$
IsRelOn($-, -, -$)	$::= \lambda R, A, B. [\forall x, y. Rxy \Rightarrow x \in A \wedge y \in B]$
RestrictRDom($-, -$)	$::= \lambda R, A, B. [\lambda x, y. x \in A \wedge Rxy]$

Display in UI as
 $A \times B$
 $=$
 $\{(u, v) \mid u \in A \wedge v \in B\}$

...

- Examples of the Test Problems

SET171 + 3	.67	$\forall X, Y, Z. X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$
SET611 + 3	.44	$\forall X, Y. (X \cap Y = \emptyset) \Leftrightarrow (X \setminus Y = X)$
SET624 + 3	.67	$\forall X, Y, Z. \text{Meets}(X, Y \cap Z) \Leftrightarrow \text{Meets}(X, Y) \vee \text{Meets}(X, Z)$
SET646 + 3	.56	$\forall x, y. \text{Subrel}(\text{Pair}(x, y), (\lambda u. \text{T}) \times (\lambda v. \text{T}))$
SET670 + 3	1.0	$\forall Z, R, X, Y. \text{IsRelOn}(R, X, Y) \Rightarrow \text{IsRelOn}(\text{RestrictRDom}(R, Z), Z, Y)$

...

- Observation:
 - complete encodings of set theory in higher-order (comprehension via λ -abstraction, Boolean and functional extensionality, ...)
- vs.
- incomplete and sometimes artificially tailored (useful lemmata) problem formulations in TPPT

- Example: TPPTP171+3

Assumptions:	$\forall B, C, x.x \in (B \cup C) \Leftrightarrow (x \in B \vee x \in C)$	(1)
	$\forall B, C, x.x \in (B \cap C) \Leftrightarrow (x \in B \wedge x \in C)$	(2)
	$\forall B, C.B = C \Leftrightarrow (B \subseteq C \wedge C \subseteq B)$	(3)
	$\forall B, C.B \cup C = C \cup B$	(4) derivable from 1,3,6
	$\forall B, C.B \cap C = C \cap B$	(5) derivable from 2,3,6
	$\forall B, C.B \subseteq C \Leftrightarrow (\forall x.x \in B \Rightarrow x \in C)$	(6)
	$\forall B, C.B = C \Leftrightarrow (\forall x.x \in B \Leftrightarrow x \in C)$	(7) derivable from 3,6
Proof Goal:	$\forall X, Y, Z.X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$	(8)

- Hence: Our Comparison is **Unfair**
→ **our higher-order problem formulations are more general and non-tailored**

HOTP may outperform FOTP



- Observation not new:
 - TPS [see papers on TPS]
 - LEO [CADE-98, Benz Müller-PhD]
 - OMEGA-OANTS [KI-01]
 - others ...
- New:
 - Combination of HOTP and FOTP may even perform better
- Approach:
 - Make use of complementary strengths of both worlds
 - Our HOTP of choice: LEO (extensional higher-order resolution)
 - Our FOTP of choice: Bliksem [Nivelle-99]
 - Our integration means of choice: Ω ANTS [AIMSA-98, Sorge-PhD]

SET171 + 3: A Motivating Example



Problem:

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

$$[\forall B, C, D. C \cup (B \cap D) = (C \cup B) \cap (C \cup D)]^F$$

↓ def.-expansion, cnf
↓ B, C, D Skolem const.

$$[(\lambda x. Bx \vee (Cx \wedge Dx)) = (\lambda x. (Bx \wedge Cx) \vee (Cx \wedge Dx))]^F$$

↓ unification constraint

$$[(\lambda x. Bx \vee (Cx \wedge Dx)) =? (\lambda x. (Bx \wedge Cx) \vee (Cx \wedge Dx))]^F$$

↓ f-extensionality
↓ x new Skolem constant

$$[(Bx \vee (Cx \wedge Dx)) =? ((Bx \wedge Cx) \vee (Cx \wedge Dx))]^F$$

↓ B-extensionality

$$[(Bx \vee (Cx \wedge Dx)) \Leftrightarrow ((Bx \wedge Cx) \vee (Cx \wedge Dx))]^F$$

↓ cnf, factor., subsumption

$$[Bx]^F$$

$$[Bx]^T \vee [Cx]^T$$

$$[Bx]^T \vee [Dx]^T$$

$$[Cx]^F \vee [Dx]^F$$

Propositional Problem!!!

within LEO or within FOTP? ↓ propositional reasoning

□

SET624+3: Direct Mapping into FO



Problem:

$$\forall X, Y, Z. \text{Meets}(X, Y \cap Z) \Leftrightarrow \text{Meets}(X, Y) \vee \text{Meets}(X, Z)$$

$$[\forall X, Y, Z. \text{Meets}(X, Y \cap Z) \Leftrightarrow \text{Meets}(X, Y) \vee \text{Meets}(X, Z)]^F$$

↓ def.-expansion

$$[\exists x. (Bx \wedge (Cx \vee Dx)) \Leftrightarrow ((\exists x. Bx \wedge Cx) \vee (\exists x. Bx \wedge Dx))]^F$$

↓ cnf

26 FO-like clauses

within LEO?
within FOTF?

SET646+3: No Proof Search



Problem:

$$\forall x_\alpha, y_\beta. \text{Subrel}(\text{Pair}(x, y), (\lambda u_\alpha. T) \times (\lambda v_\beta. T))$$

$$[\forall x, y. \text{Subrel}(\text{Pair}(x, y), (\lambda u. T) \times (\lambda v. T))]^F$$

↓ def.-expansion

$$[\forall x, y, u, v. (u = x \wedge v = y) \Rightarrow ((\neg \perp) \wedge (\neg \perp))]^F$$

↓ cnf

$$\dots$$
$$[\perp]^T \vee [\perp]^T = \square$$

...

SET611 + 3: Repeated Extensionality



Problem:

$$\forall A, B. (A \cap B = \emptyset) \Leftrightarrow (A \setminus B = A)$$

$$[\forall A, B. (A \cap B = \emptyset) \Leftrightarrow (A \setminus B = A)]^F$$

↓ def.-expansion

$$[\forall A, B. (\lambda x. (Ax \wedge Bx)) = (\lambda x. \perp) \Leftrightarrow (\lambda x. (Ax \wedge \neg Bx)) = (\lambda x. Ax)]^F$$

↓ cnf, A, B Skolem

$$\begin{aligned} (1) & [(\lambda x. (Ax \wedge Bx)) = (\lambda x. \perp)]^T \vee [(\lambda x. (Ax \wedge \neg Bx)) = (\lambda x. Ax)]^T \\ (2) & [(\lambda x. (Ax \wedge Bx)) = ? (\lambda x. \perp)] \vee [(\lambda x. (Ax \wedge \neg Bx)) = ? (\lambda x. Ax)] \\ & [(\lambda x. (Ax \wedge Bx)) = (\lambda x. \perp)]^F \vee [(\lambda x. (Ax \wedge \neg Bx)) = (\lambda x. Ax)]^F \end{aligned}$$

↓ several rounds
↓ of B&f-ext.
↓ and cnf

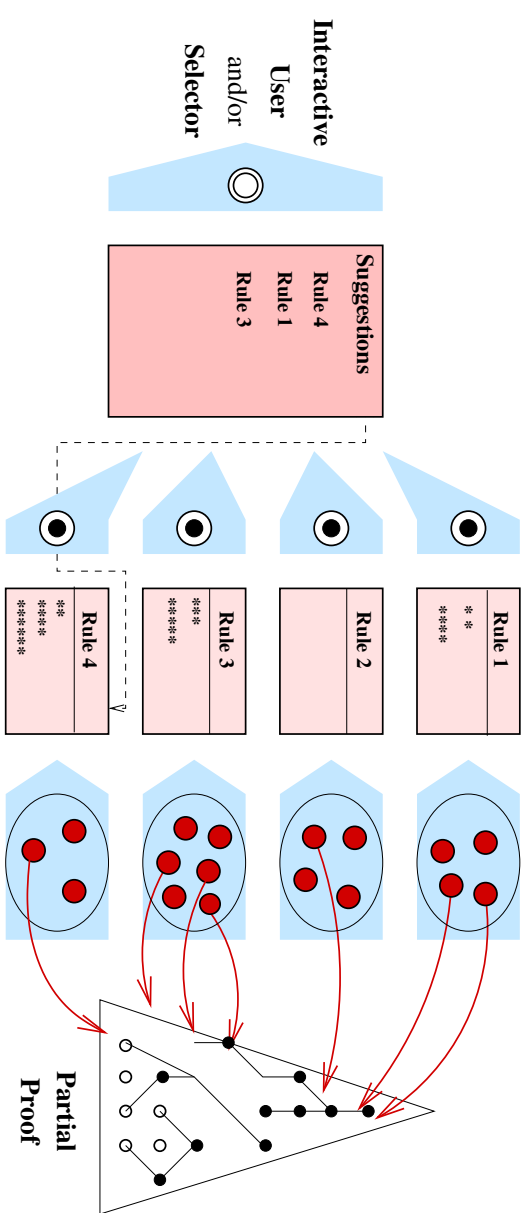
inconsistent set of FO-like clauses

within LEO?
within FOTP?

HOTP-FOTP: Modeling in Ω ANTS



- Ω ANTS:
 - distributed suggestion mechanism for interactive theorem proving
 - backboard architecture, supports redefinition of agents at run-time
 - automation of proof search possible [Calulemus-00]



HOTP-FOTP: Modeling in Ω ANTS



OLD Solution

- $\frac{\text{HO-goal}}{\text{HO-goal}}$ LEO (LEO-params)
- $\frac{\text{Conjunction-of-FO-clauses}}{\text{HO-goal}}$ LEO-with-partial-result (LEO-params)
- $\frac{\text{FO-goal}}{\text{FO-goal}}$ FOTP (FOTP-params)

NEW Solution

- $\frac{\text{HO-goal}}{\text{HO-goal}}$ LEO (LEO-params)
- $\frac{\text{HO-goal}}{\text{HO-goal}}$ LEO+FOTP (LEO-partial-proof, FO-clauses, FO-proof, LEO-params)

Experiments: Results (I)



SET	Rat.	Vampire 7	Strat.	LEO Cl.	Time	Cl.	Time	LEO + FOTP	FOcl	FOTm	GenCl
014+4	.67	.01	ST	41	.16	34	6.76	19	.01		7
017+1	.56	.03	EXT	3906	57.52	25	8.54	16	.01		74
066+1	1.00	--	--	--	--	26	6.80	20	10		56
067+1	.56	.04	ST	6	.02	13	.32	16	.01		12
076+1	.67	.00	--	--	--	10	.47	18	.01		35
086+1	.22	.04	ST	4	.01	4	.01	N/A	N/A		N/A
096+1	.56	.03	--	--	--	27	7.99	14	.01		25
143+1	.67	68.71	EIR	37	.38	33	7.93	18	.01		19
171+3	.67	108.31	EIR	36	.56	25	4.75	19	.01		20
580+3	.44	14.71	EIR	25	.19	6	2.73	8	.01		13
601+3	.22	168.40	EIR	145	2.20	55	4.96	8	.01		13
606+3	.78	62.02	EIR	21	.33	17	10.8	15	.01		5
607+3	.67	65.57	EIR	22	.31	17	7.79	15	.01		6
609+3	.89	161.78	EIR	37	.60	26	6.50	19	10		17
611+3	.44	60.20	EIR	996	12.69	72	32.14	38	.01		101
612+3	.89	113.33	EIR	41	.54	18	3.95	6	.01		7
614+3	.67	157.88	EIR	38	.46	19	4.34	16	.01		17
615+3	.67	109.01	EIR	38	.57	17	3.59	6	.01		9
623+3	1.00	--	EXT	43	8.84	23	9.54	10	.01		14
624+3	.67	.04	ST	4942	34.71	54	9.61	46	.01		212
630+3	.44	60.39	EIR	11	.07	6	.08	8	10		4

Experiments: Results (II)



SET	Rat.	Vampire 7	Strat.	LEO Cl.	Time Cl.	Time	LEO + FOTP	FOcl	FOTm	GenCl
640+3	.22	70.41	EIR	2	2	.01	N/A	N/A	N/A	N/A
646+3	.56	59.63	EIR	2	2	.01	N/A	N/A	N/A	N/A
647+3	.56	64.21	EIR	26	13	.30	13	.01	.01	15
648+3	.56	64.22	EIR	26	14	.30	13	.01	.01	16
649+3	.33	63.77	EIR	45	29	5.49	12	.01	.01	16
651+3	.44	63.88	EIR	20	11	.16	10	10	10	11
657+3	.22	1.44	EIR	2	2	.01	N/A	N/A	N/A	N/A
669+3	.56	.34	EI	35	35	.23	N/A	N/A	N/A	N/A
670+3	1.00	--	EXT	15	17	.36	16	.01	.01	6
671+3	.78	218.02	EIR	78	7	2.71	10	.01	.01	14
672+3	1.00	--	EXT	27	30	.70	21	.01	.01	14
673+3	.78	47.86	EIR	78	14	5.66	14	.01	.01	16
680+3	.33	.07	ST	185	29	4.61	18	.01	.01	24
683+3	.22	.06	ST	46	35	8.90	18	10	10	24
684+3	.78	.33	ST	275	46	5.95	26	.01	.01	47
686+3	.56	.11	ST	274	46	5.37	26	.01	.01	46
716+4	.89	--	ST	39	18	3.81	18	.01	.01	118
724+4	.89	--	EXT	154	18	7.21	15	10	10	23
741+4	1.00	--	--	--	--	--	--	--	--	--
747+4	.89	--	ST	34	25	1.11	18	10	10	10
752+4	.89	--	--	--	50	6.60	48	.01	.01	4363
753+4	.89	--	--	--	15	3.07	12	10	10	19
764+4	.56	.02	EI	9	8	.04	N/A	N/A	N/A	N/A
770+4	.89	--	--	--	--	--	--	--	--	--

Soundness

- LEO's calculus is sound
- Bliksem's calculus is sound
- Crucial part:
 - transformation from FO-like clause in LEO to real FO clauses in Bliksem must preserve satisfiability
 - we use TRAMPs [Meier00] injective mapping

$$P(f(a)) \dashrightarrow \mathcal{Q}_{\text{pred}}^1(P, \mathcal{Q}_{\text{fun}}^1(f, a))$$

Completeness

- LEO's calculus is Henkin complete (the implementation of LEO is not though)
- Completeness of the cooperative approach relies on the completeness of LEO

Generation of proof-objects

- How can we obtain a common proof object?
 - solved since Tuesday (LPAR “programming session” with Volker)

Leibniz equality (and other definitions of equality)

- Leibniz equality: $=$ can be defined as $\lambda x. \lambda y. \forall P. P(x) \Rightarrow P(y)$
- Example: $a = b \Rightarrow f(a) = f(b)$

Primitive equality	Leibniz equality
$[a = b]^T$ $[f(a) = f(b)]^F$	$[P(a)]^F \vee [P(b)]^T$ $[Q(f(a))]^T$ $[Q(f(b))]^F$
refutable in LEO and Bliksem	refutable only in LEO $P \leftarrow \lambda x. Q(f(x))$

Related Work



- Denzinger/Fuchs [IJCAI-99]:

TECHS system

- only cooperation of first-order systems

Andreas Meier
[CADE-00]

Joe Hurd
[CADE-02]

Jia Meng, Larry Paulson
[IJCAR-04]

TRAMP, generic
interface between
OMEGA and FOTPs

generic interface
between
HOL and FOTPs

interface
between
Isabelle and Vampire

- no calls to FOTP from within automated HO proof search

- Computer-supported Mathematics
 - representation does matter
- Automation of Mathematical Reasoning
 - higher-order may outperform first-order in certain domains
- Automation of Higher-Order Theorem Proving (HOTP)
 - cooperation with a first-order theorem proving (FO_{TP}) is beneficial
- Architectures supporting System Integrations
 - agent-based reasoning with OANTS
- Problem Libraries such as TP_{TP}
 - should support alternative (e.g. higher-order) problem representations

And Finally ...



I can fully recommend TEX_{MACS} as scientific editor

