Judging Granularity for Automated Mathematics Teaching[1]

Marvin Schiller, Christoph Benzmüller, Ann van de Veire

Saarland University

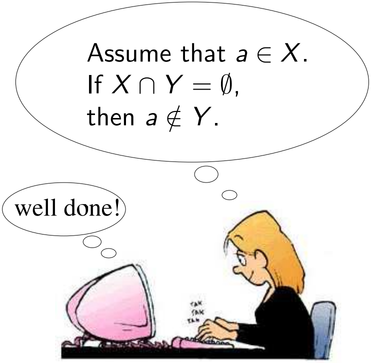
13th LPAR Conference, Phnom Penh November 12th - 17th, 2006

[1]This work has been funded by the DFG Collaborative Research Center on Resource-Adaptive Cognitive Processes, SFB 378.

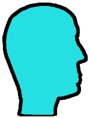
The SFB 378 DIALOG Project

Can we automate NL-based tutoring of mathematical proofs?

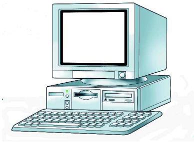
- NL analysis
- dialog management
- domain reasoning in mathematics (using the ΩMEGA environment)
- tutorial aspects
- NL generation



Example Dialog



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



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

Correct. Good start!

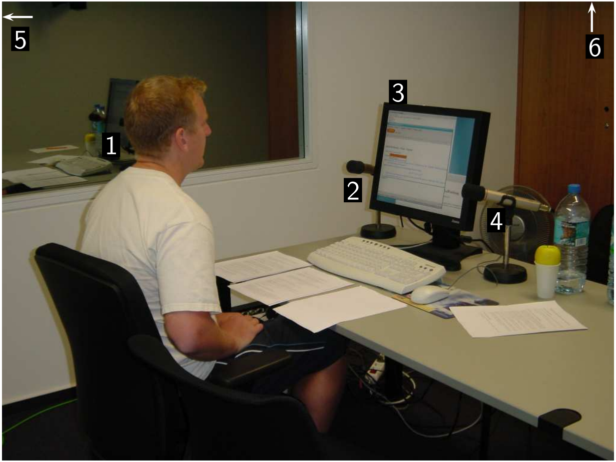
Then $\exists z$ such that (x, z) in $(R \cup S)$ and (z, y) in T

Correct!

Then . . .

Student Room

1 Subject
2 Audio Recording
3 Subject GUI
4 Audio Control
5 Dome Camera
6 Camera



Wizard Room

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

An Annotated Corpus of Tutorial Dialogs

Dialogs with human-simulated tutoring system [KI-06, LREC-06]

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.

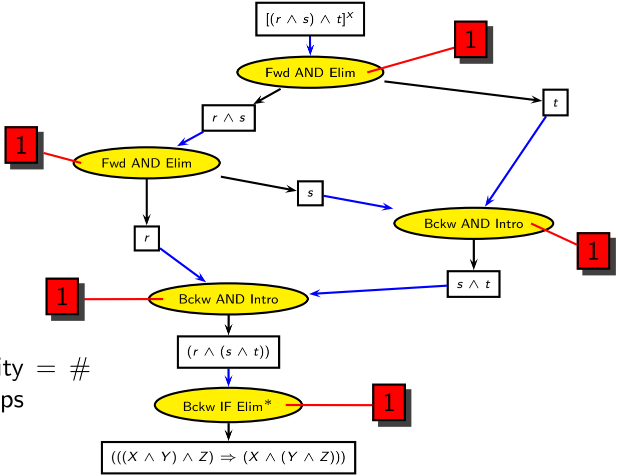
correct	appropriate	limited relevance

Goal: Automate correctness, granularity and relevance judgments.

Judging Granularity

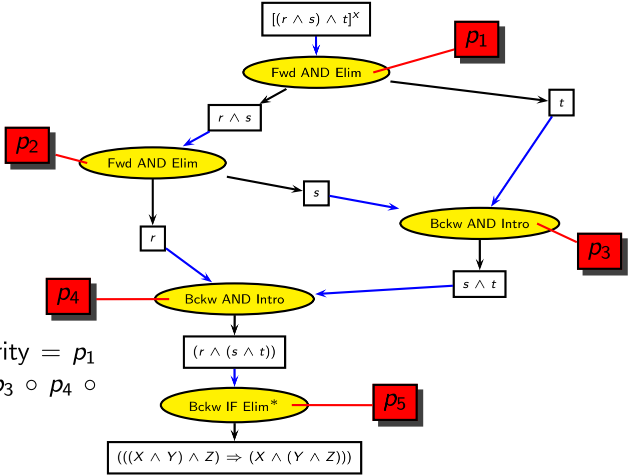
- Use ΩMEGA system
- Hypothesis: granularity of a proof step is sufficiently well related to proof size in a well chosen calculus
- Calculi studied first:
  Gentzen's ND [Gentzen-34] and "Psychology of Proof" [Rips-94]
- Implemented granularity analysis framework for inspecting proofs
- Implemented calculi in framework
- Evaluation: compare mechanical classification to expert's ratings

Granularity Evaluation Framework



Granularity = # proof steps

Granularity Evaluation Framework (ctd.)



Granularity $= p_1 \circ p_2 \circ p_3 \circ p_4 \circ p_5$

Relating Granularity to Calculus Level Steps

A: $(x, y) \in (S^{-1} \circ R^{-1}) \Leftrightarrow \exists z[(z, x) \in S \wedge (y, z) \in R]$

B: $\forall x \forall y[\exists z[(y, z) \in R \wedge (z, x) \in S] \rightarrow (y, x) \in (R \circ S)]$

C: therefore it follows: $(x, y) \in (S^{-1} \circ R^{-1}) \rightarrow (y, x) \in (R \circ S)$

	Statement A	Statement B	Statement C
Tutor	"too coarse-grained"	"appropriate"	"appropriate"
PSYCOP	5	2	10
[Gentzen34]	3	3	9

Number of justifying proof steps for PSYCOP and Gentzen's NK.

Evaluation Results (20 steps from the corpus)

Tutor's rating	Avg. proof step length at calculus level (with std. deviation)			
	PSYCOP calculus		Gentzen's ND calculus	
"too detailed"	1,00		0	
"appropriate"	5,27	(4,88)	5,00	(5,14)
"too coarse-grained"	11,67	(6,80)	10,33	(7,72)

Conclusion and Outlook

Conclusion

- Calculating proof sizes in neither ND nor PSYCOP seems sufficient
- 1 user's proof step $\approx$ 5.45 calculus level steps in ND
- ND and PSYCOP do not support rewriting or deep inference steps, however, the CORE calculus does.

Outlook

The study motivates to

- investigate other calculi: e.g., CORE calculus / deep inference for judging granularity
- incorporate a student and a teacher model