Natural language based Interaction with a Mathematics Assistance System

Christoph Benzmüller

Joint work with: SFB378 DIALOG Project

Special thanks to: Marvin Schiller, Marc Buckley, Magda Wolska

DIALOG Project

Computer Science & Comput. Ling.
Saarland University
Saarbrücken, Germany
`http://www.ags.uni-sb.de/~chris/dialog/`

Ringvorlesung, 8 February 2006, Saarbrücken, Germany

UNIVERSITÄT DES SAARLANDES

Motivation and Goal of the Project

DIALOG Project

- Support for e-learning in mathematics:

  (interactive) tutoring of mathematical proof

  - in addition to traditional classroom courses
  - as part of an e-learning course

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

well done!

UNIVERSITÄT DES SAARLANDES

Why Focus on “Mathural”?

- “Mathural”: natural language (NL) and mathematical formulas

Why Focus on “Mathural”?

- “Mathural”: natural language (NL) and mathematical formulas
- [Moore 93]: use of NL supports active learning

Why Focus on “Mathural”?

- “Mathural”: natural language (NL) and mathematical formulas
- [Moore 93]: use of NL supports active learning
- Understanding “mathural” is an interesting challenge for CL

Why Focus on “Mathural”?

DIALOG Project

- “Mathural”: natural language (NL) and mathematical formulas
- [Moore 93]: use of NL supports active learning
- Understanding “mathural” is an interesting challenge for CL
- Progress in project $\longrightarrow$ probably we can one day automatically proof check mathematical publications

Theorem: $\sqrt{2}$ is irrational.
Proof: (by contradiction)
Assume $\sqrt{2}$ is rational, that is, there exist natural numbers m, n with no common divisor such that $\sqrt{2} = m/n$. Then $n\sqrt{2} = m$, and thus $2n^2 = m^2$. Hence m^2 is even and, since odd numbers square to odds, m is even; say $m = 2k$. Then $2n^2 = (2k)^2 = 4k^2$, that is, $n^2 = 2k^2$. Thus, n^2 is even too, and so is n. That means that both n and m are even, contradicting the fact that they do not have a common divisor.

q.e.d.

A Demo ...

DIALOG Project

. . . of the DIALOG proof tutor

WOZ-Experiment $\longrightarrow$ Own Corpus

DIALOG Project

Subject (Student) Room:

Wizard Room:

WOZ-Experiment → Own Corpus

Wizard (Tutor) Room:

Subject Room:

Wizard of Oz Experiment

- research experiment in which subjects interact with a computer system that subjects believe to be autonomous, but which is actually being operated or partially operated by an unseen human being

Wizard of Oz Experiment

- research experiment in which subjects interact with a computer system that subjects believe to be autonomous, but which is actually being operated or partially operated by an unseen human being
- missing system functionality that the wizard provides may be implemented in later versions of the system, but its precise details are generally considered irrelevant to the study

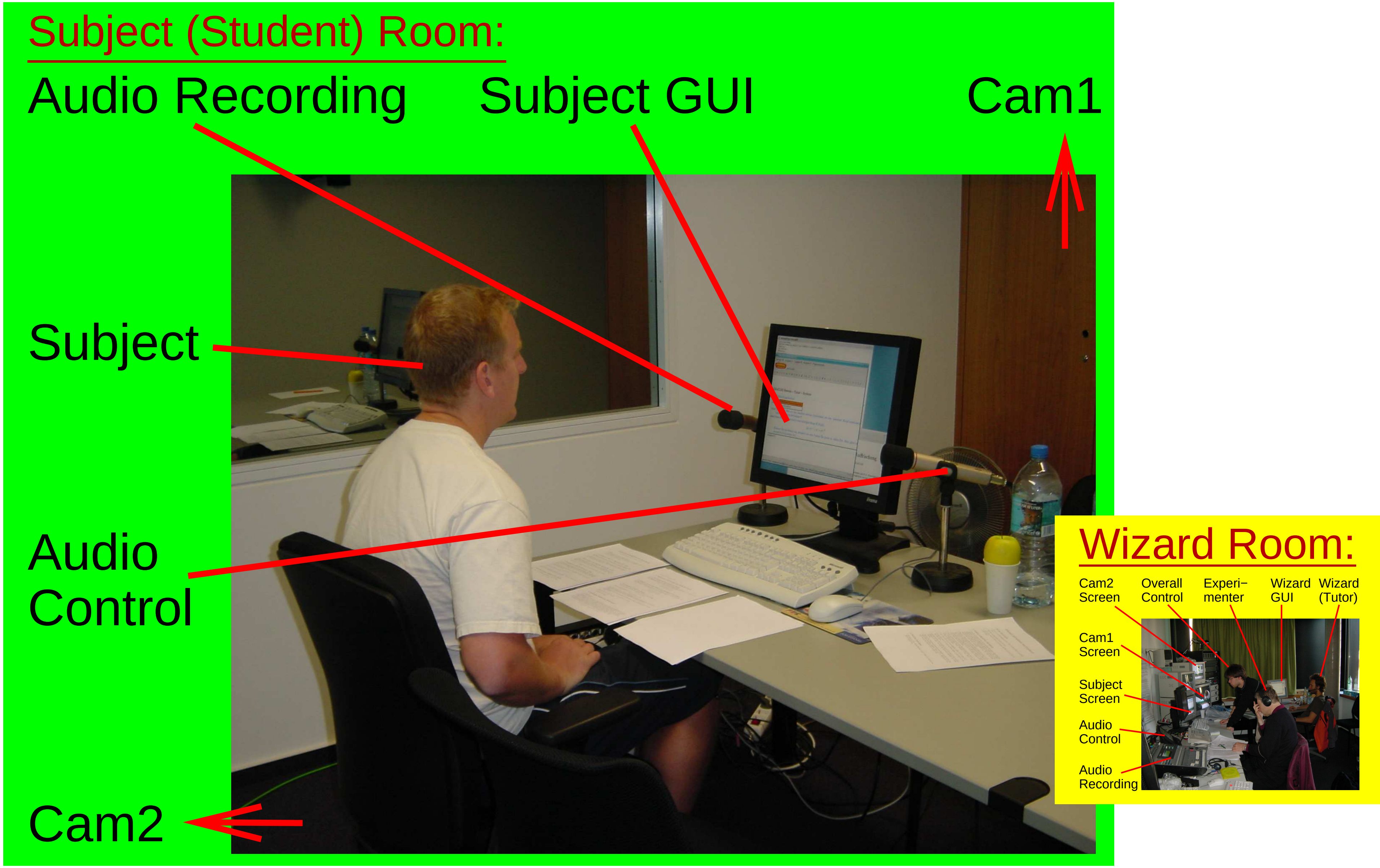
Wizard of Oz Experiment

DIALOG Project

- research experiment in which subjects interact with a computer system that subjects believe to be autonomous, but which is actually being operated or partially operated by an unseen human being
- missing system functionality that the wizard provides may be implemented in later versions of the system, but its precise details are generally considered irrelevant to the study

- the name of the experiment comes from The Wonderful Wizard of Oz story, in which an ordinary man hides behind a curtain and pretends to be a powerful wizard

WOZ-Experiment → Own Corpus

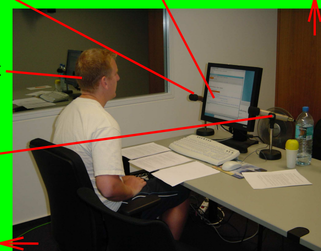


WOZ-Experiment $\longrightarrow$ Own Corpus

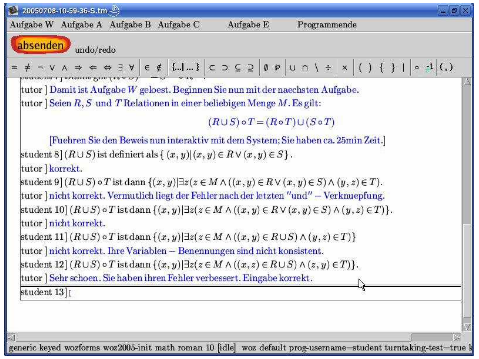
DIALOG Project

Wizard (Tutor) Room:

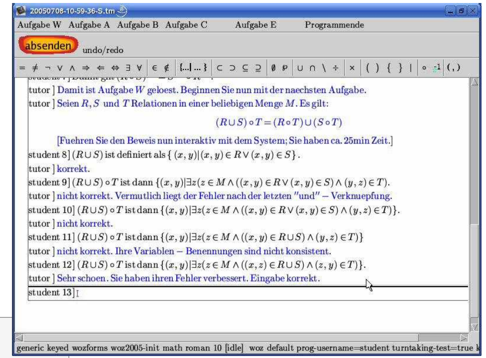
Subject Room:

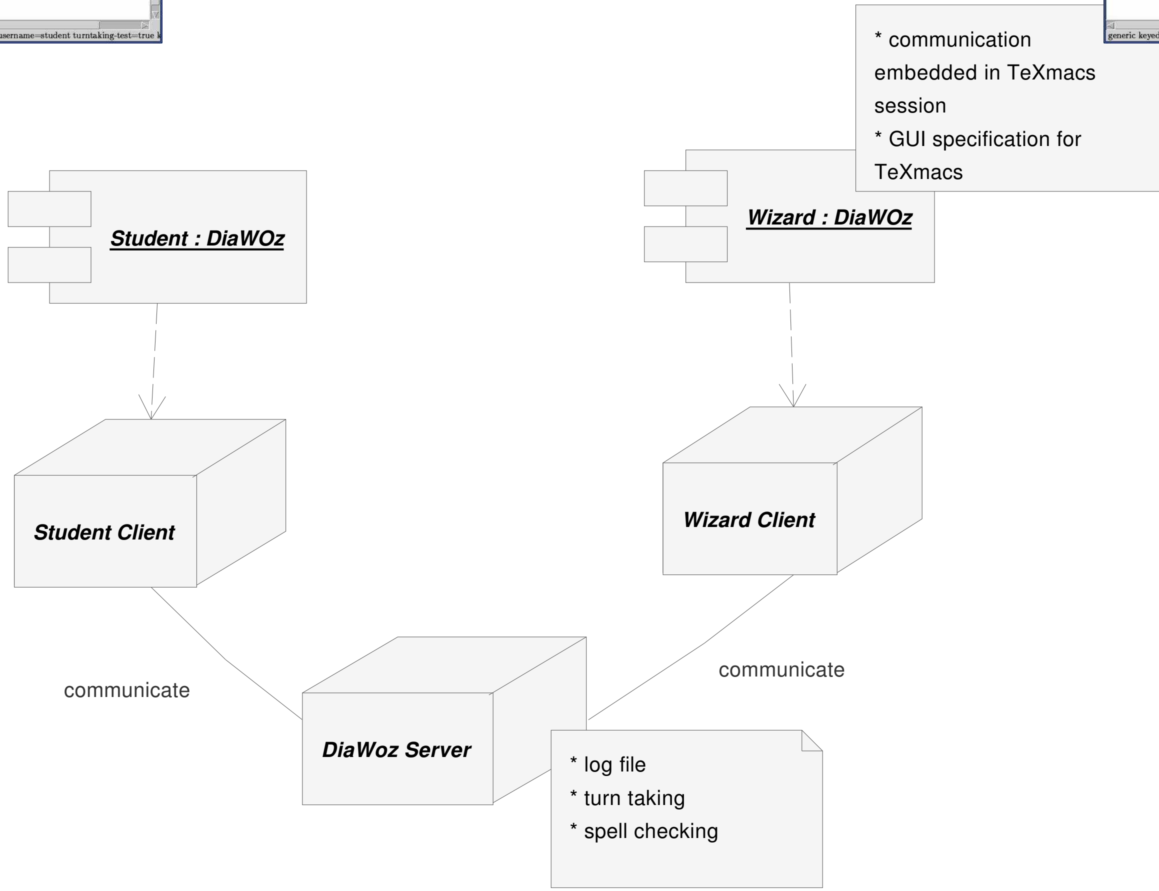


Architecture of DIAWOZ-II



More than a configurable chat tool for maths . . .





Outline of Remaining Lecture

DIALOG Project

- The DIALOG project in the SFB 378

Outline of Remaining Lecture

- The DIALOG project in the SFB 378
- Corpora we obtained from experiments

Outline of Remaining Lecture

- The DIALOG project in the SFB 378
- Corpora we obtained from experiments
- The components of the DIALOG system

Outline of Remaining Lecture

- The DIALOG project in the SFB 378
- Corpora we obtained from experiments
- The components of the DIALOG system
- A research challenge of the DIALOG project:

Outline of Remaining Lecture

DIALOG Project

- The DIALOG project in the SFB 378
- Corpora we obtained from experiments
- The components of the DIALOG system
- A research challenge of the DIALOG project:
  - proof step evaluation and granularity

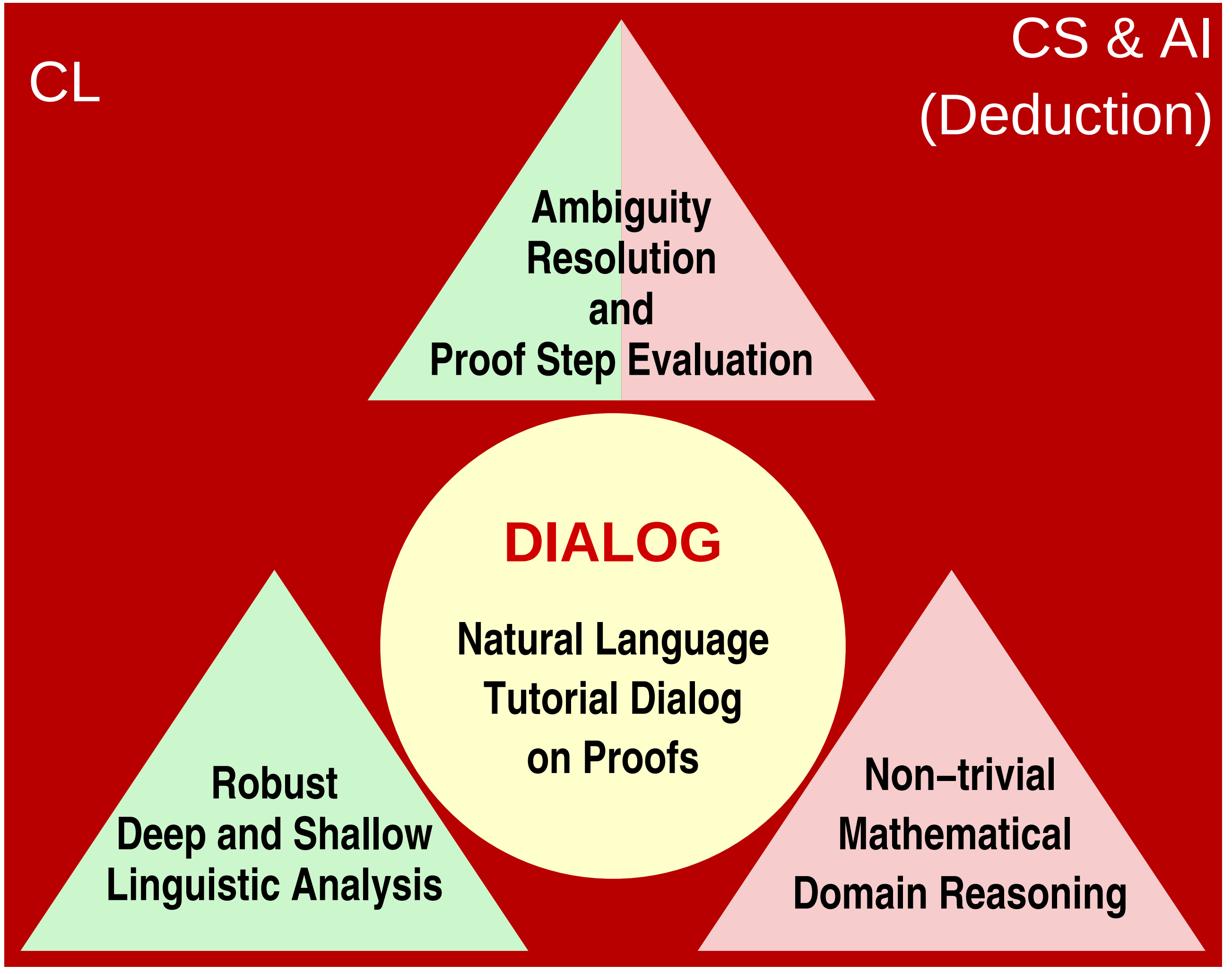
Outline of Remaining Lecture

- The DIALOG project in the SFB 378
- Corpora we obtained from experiments
- The components of the DIALOG system
- A research challenge of the DIALOG project:
  - proof step evaluation and granularity
- An agent-based dialog manager

Outline of Remaining Lecture

- The DIALOG project in the SFB 378
- Corpora we obtained from experiments
- The components of the DIALOG system
- A research challenge of the DIALOG project:
  - proof step evaluation and granularity
- An agent-based dialog manager
- Conclusion

The DIALOG Project in the SFB 378



The project also strives Psychology and Philosophy

The DIALOG Team

DIALOG Project

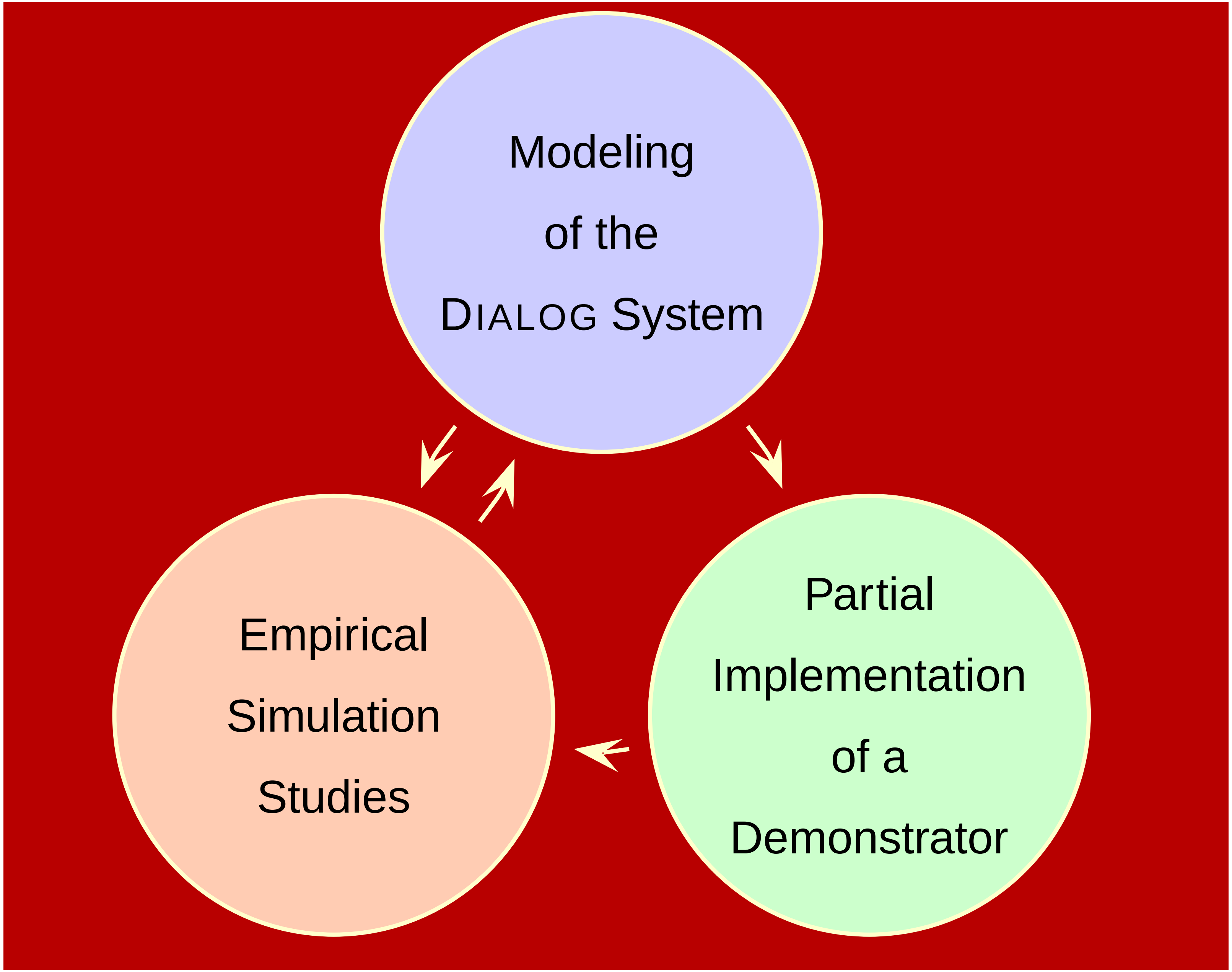
- Computational Linguistics
  Dr. Ivana Kruijff-Korbayova
  Prof. Dr. Manfred Pinkal
  Magdalena Wolska

The DIALOG Team

DIALOG Project

- Computational Linguistics
  Dr. Ivana Kruijff-Korbayova
  Prof. Dr. Manfred Pinkal
  Magdalena Wolska
- Computer Science
  HD Dr. Christoph Benzmüller
  Dr. Chad Brown
  Mark Buckley
  Dr. Armin Fiedler
  PD Dr. Helmut Horacek
  Marvin Schiller
  Henri Lesourd Prof. Dr. Jörg Siekmann
  Dimitra Tsovaltzis

Method: Progressive Refinement



Different Corpora

Experiment 1 (2003):

- sets and simple operations on sets

Experiment 2 (2005):

- relations and simple operations on relations

Different Corpora

Experiment 1 (2003):

- sets and simple operations on sets
- 1 maths teacher (wizard)

Experiment 2 (2005):

- relations and simple operations on relations
- 4 maths teachers (wizards)

Different Corpora

Experiment 1 (2003):

- sets and simple operations on sets
- 1 maths teacher (wizard)
- 24 students / two hours for each student

Experiment 2 (2005):

- relations and simple operations on relations
- 4 maths teachers (wizards)
- 37 students / two hours for each student

Different Corpora

Experiment 1 (2003):

- sets and simple operations on sets
- 1 maths teacher (wizard)
- 24 students / two hours for each student
- 66 dialogs / av. 12 turns / 1115 sentences

Experiment 2 (2005):

- relations and simple operations on relations
- 4 maths teachers (wizards)
- 37 students / two hours for each student
- 902 student turns

Different Corpora

DIALOG Project

Experiment 1 (2003):

- sets and simple operations on sets
- 1 maths teacher (wizard)
- 24 students / two hours for each student
- 66 dialogs / av. 12 turns / 1115 sentences

Experiment 2 (2005):

- relations and simple operations on relations
- 4 maths teachers (wizards)
- 37 students / two hours for each student
- 902 student turns
- some information on the students: different study subjects, 27 had maths as “specialized course” in their A-levels (Abitur)

Different Corpora

Corpora available at:

`www.ags.uni-sb.de/~chris/dialog/`

`www.ags.uni-sb.de/~schiller/dialog2005study.html`

Corpus Example: 1. Experiment (2003)

T1: Bitte zeigen Sie: $K((A \cup B) \cap (C \cup D)) = (K(A) \cap K(B)) \cup (K(C) \cap K(D))$!
[Please show: $K((A \cup B) \cap (C \cup D)) = (K(A) \cap K(B)) \cup (K(C) \cap K(D))$!]

S1: nach deMorgan-Regel-2 ist $K((A \cup B) \cap (C \cup D)) = (K(A \cup B) \cup K(C \cup D))$.
[by deMorgan-Rule-2 $K((A \cup B) \cap (C \cup D)) = (K(A \cup B) \cup K(C \cup D))$ holds.]

T2: Das ist richtig!
[This is correct!]

S2: $K(A \cup B)$ ist laut deMorgan-1 $K(A) \cap K(B)$
[$K(A \cup B)$ is $K(A) \cap K(B)$ according to deMorgan-1]

T3: Das stimmt auch.
[That is also right.]

S3: und $K(C \cup D)$ ist ebenfalls laut deMorgan-1 $K(C) \cap K(D)$
[and $K(C \cup D)$ is also $K(C) \cap K(D)$ according to deMorgan-1]

...

Corpus Example: 2. Experiment (2005)

S0: was ist $\circ$

T0: Das Relationenprodukt, auch Komposition von Relationen genannt. Bitte schauen Sie sich die Definition unter Abschnitt 4 an. `(k.A.; k.A.; k.A.)`

S1: $(R\circ S)^{-1} = \{(z,x) | \exists y((x,y) \in R \wedge (y,z) \in S\}$

T1: Das ist korrekt. `(korrekt; angemessen; relevant)`

S2: $R^{-1} = \{(x,y) | (y,x) \in R\}$

T2: Ebenfalls korrekt. `(korrekt; angemessen; relevant)`

S3: Also ist $S^{-1} \circ R^{-1} = \{(v,x) | v \in S^{-1} \wedge x \in R^{-1}\}$

T3: Nein. Auch die inversen Relationen, S^{-1} und R^{-1}, sind binaere Relationen! `(inkorrekt; angemessen; relevant)`

S4: Also ist $S^{-1} \circ R^{-1} = \{(v,x) | \exists z((v,z) \in S^{-1} \wedge (z,x) \in R^{-1}\}$

T4: ...

Different Corpora

Observation in 2005 corpus:

- Two dialog fragments for exercise W: $(\mathsf{R} \circ \mathsf{S})^{-1} = \mathsf{S}^{-1} \circ \mathsf{R}^{-1}$

Different Corpora

DIALOG Project

Observation in 2005 corpus:

- Two dialog fragments for exercise W: $(\mathsf{R} \circ \mathsf{S})^{-1} = \mathsf{S}^{-1} \circ \mathsf{R}^{-1}$
- Student A

student]	One needs to show equality between two sets
tutor]	That's right! How do you proceed?
student]	I use the extensionality principle
tutor]	That's correct!

Different Corpora

Observation in 2005 corpus:

- Two dialog fragments for exercise W: $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$
- Student A

student]	One needs to show equality between two sets
tutor]	That's right! How do you proceed?
student]	I use the extensionality principle
tutor]	That's correct!

- Student B

student]	$(R \circ S)^{-1} = \{(x, y) \vert (y, x) \in (R \circ S)\}$
tutor]	correct
student]	$(R \circ S)^{-1} = \{(x, y) \vert (y, x) \in \{(x, y) \vert \exists z (z \in M \wedge (x, z) \in R \wedge (z, y) \in S\}\}$
tutor]	okay, but can be done simpler.

Different Corpora

Observation in 2005 corpus:

- Two dialog fragments for exercise W: $(\mathsf{R} \circ \mathsf{S})^{-1} = \mathsf{S}^{-1} \circ \mathsf{R}^{-1}$
- Student A

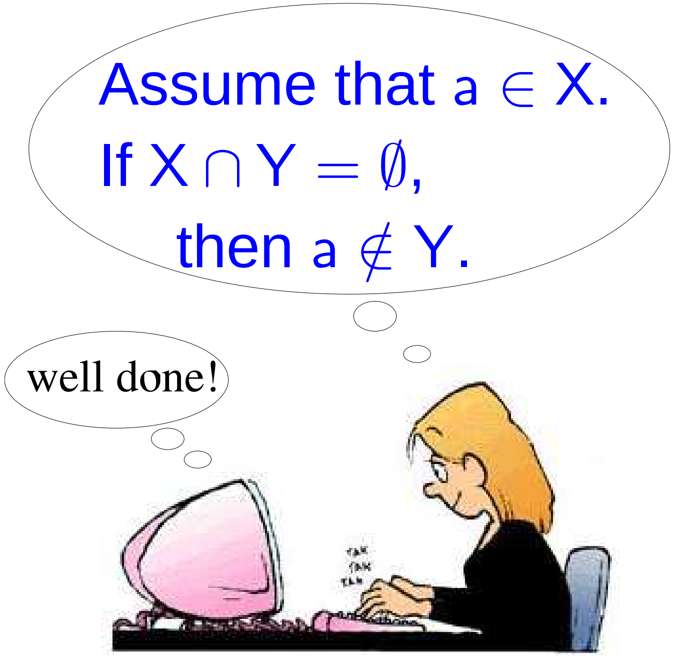
student]	One needs to show equality between two sets
tutor]	That's right! How do you proceed?
student]	I use the extensionality principle
tutor]	That's correct!

- Student B

student]	$(\mathsf{R} \circ \mathsf{S})^{-1} = \{(x, y) \mid (y, x) \in (\mathsf{R} \circ \mathsf{S})\}$
tutor]	correct
student]	$(\mathsf{R} \circ \mathsf{S})^{-1} = \{(x, y) \mid (y, x) \in \{(x, y) \mid \exists z (z \in \mathsf{M} \wedge (x, z) \in \mathsf{R} \wedge (z, y) \in \mathsf{S}\}\}$
tutor]	okay, but can be done simpler.

Research Goal

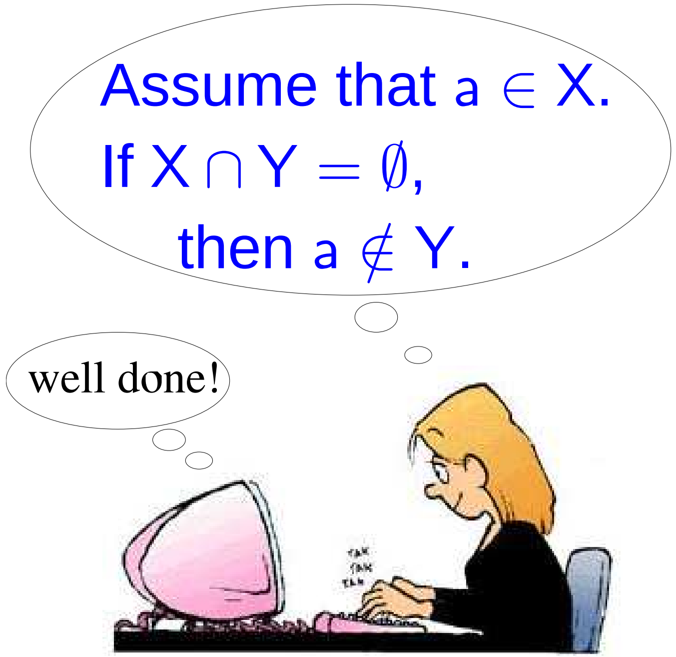
- Can we mechanize and automate the maths teachers tasks?



Research Goal

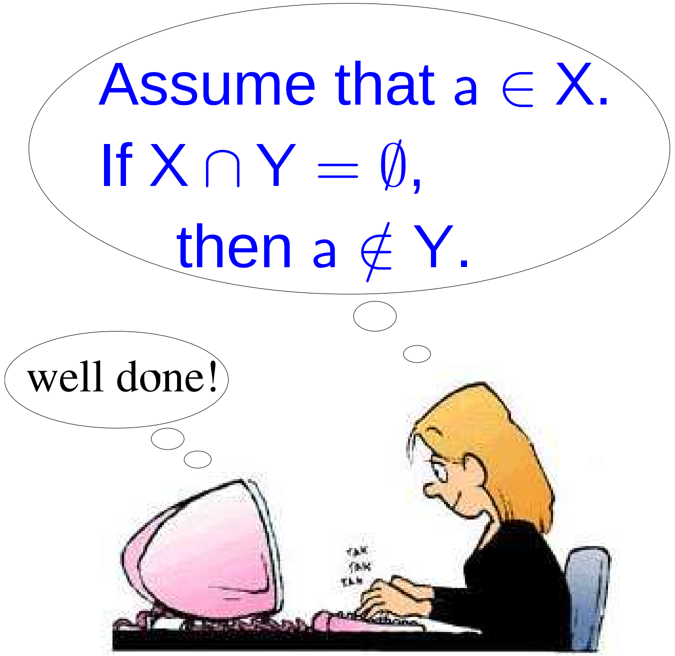
DIALOG Project

- Can we mechanize and automate the maths teachers tasks?
- What are the main challenges?

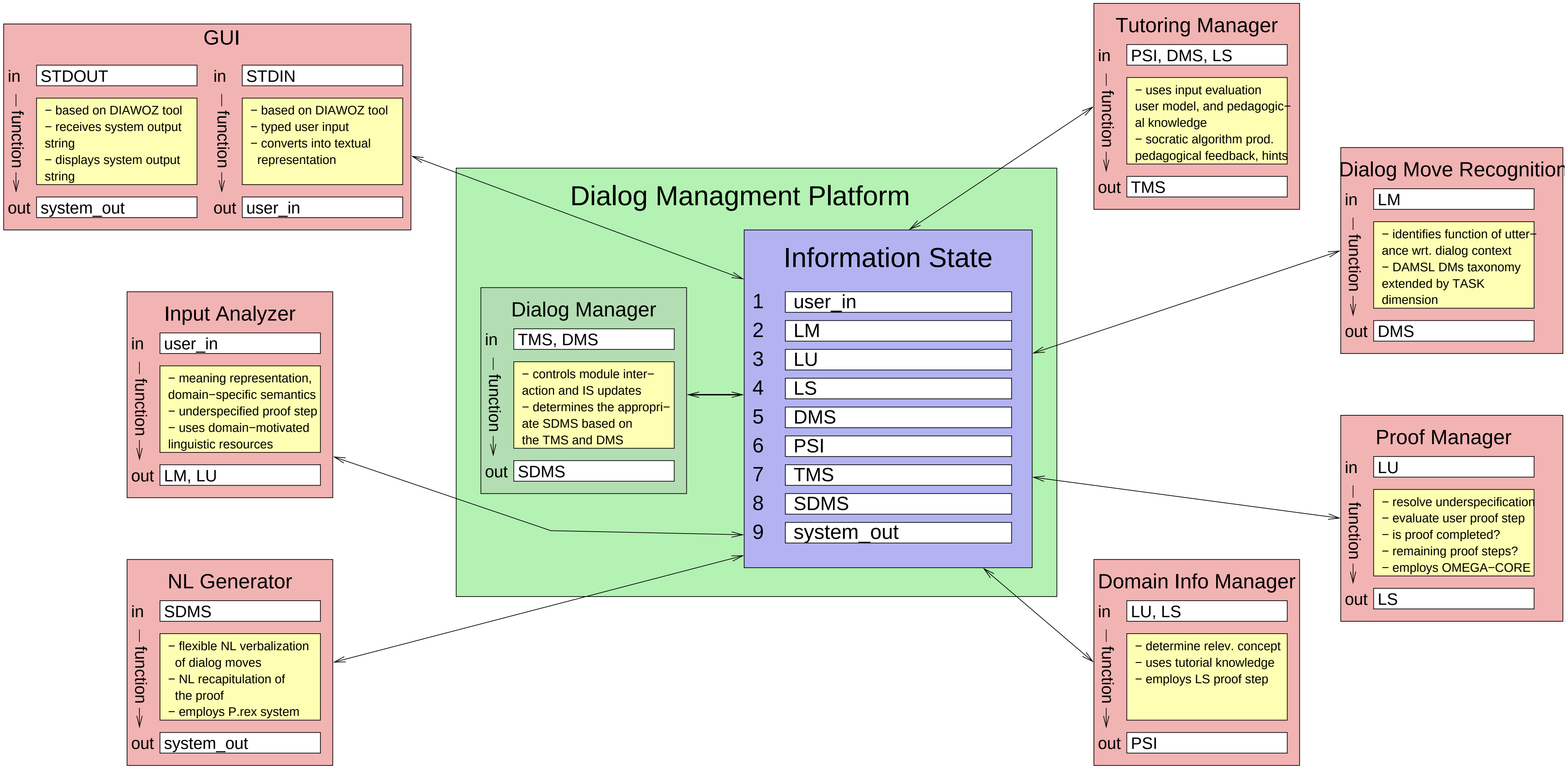


Research Goal

- Can we mechanize and automate the maths teachers tasks?
- What are the main challenges?
- Focus on some these challenges in the DIALOG project



The DIALOG System and Components



Approach Adapt DIAWOZ-II

GUI

Approach Adapt DIAWOZ-II

Input Typed (later spoken) NL & formulas (& later graphics)

DIALOG Project

Approach Adapt DIAWOZ-II

Input Typed (later spoken) NL & formulas (& later graphics)

- Example: $A \cup B$ contains B

GUI

Approach Adapt DIAWOZ-II

Input Typed (later spoken) NL & formulas (& later graphics)

- Example: $A \cup B$ contains B

Output The input converted into an appropriate representation format (here “string”)

GUI

Approach Adapt DIAWOZ-II

Input Typed (later spoken) NL & formulas (& later graphics)

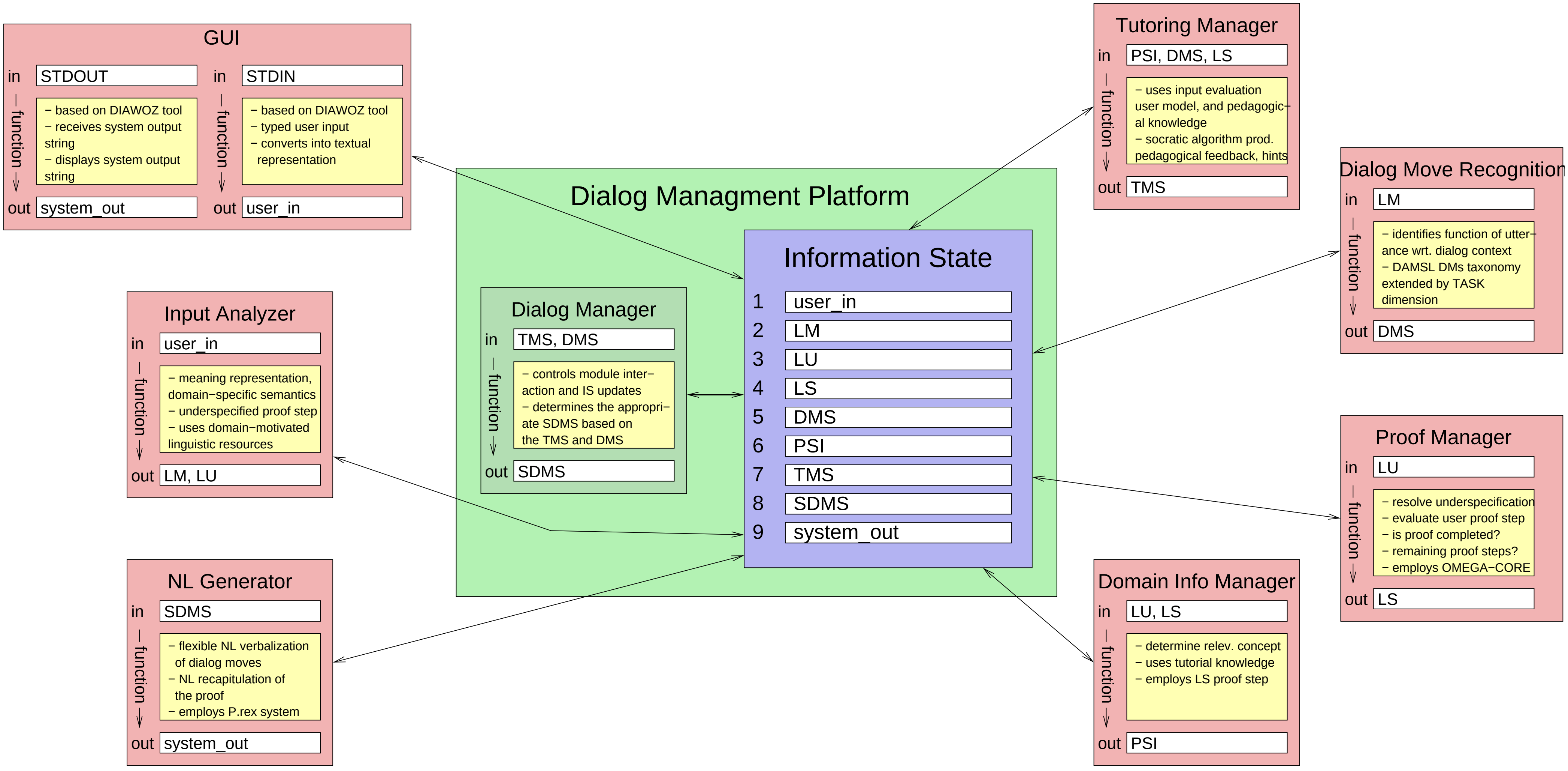
- Example: $A \cup B$ contains B

Output The input converted into an appropriate representation format (here "string")

- Example: "$A \cup B$ contains B"

Input Analyzer

Input Analyzer



Input Analyzer

Approach Own Development → CL task (Magdalena Wolska)

Input Analyzer

Approach Own Development $\longrightarrow$ CL task (Magdalena Wolska)

Input Appropriate representation of input (here we use a “string”)

Input Analyzer

Approach Own Development $\longrightarrow$ CL task (Magdalena Wolska)

Input Appropriate representation of input (here we use a "string")

- Example: "A $\cup$ B contains B"

Input Analyzer

DIALOG Project

Approach Own Development $\longrightarrow$ CL task (Magdalena Wolska)

Input Appropriate representation of input (here we use a "string")

- Example: "A $\cup$ B contains B"

Output LM: The linguistic meaning of the utterance
LU: The domain contribution (uttered proof step) in a semantic representation

Input Analyzer

DIALOG Project

Approach Own Development $\longrightarrow$ CL task (Magdalena Wolska)

Input Appropriate representation of input (here we use a “string”)

- Example: “A $\cup$ B contains B”

Output LM: The linguistic meaning of the utterance

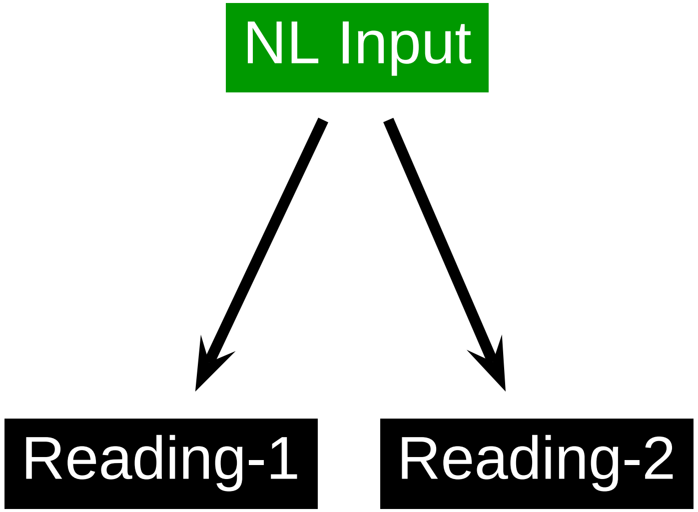
LU: The domain contribution (uttered proof step) in a semantic representation

- Example:

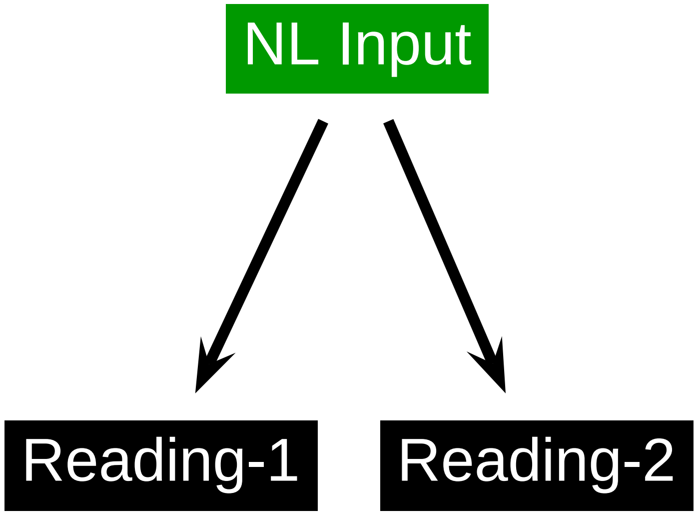
  LM: domain contribution (proof step)

  LU: $B \in A \cup B \parallel B \subseteq A \cup B \parallel B \subset A \cup B$

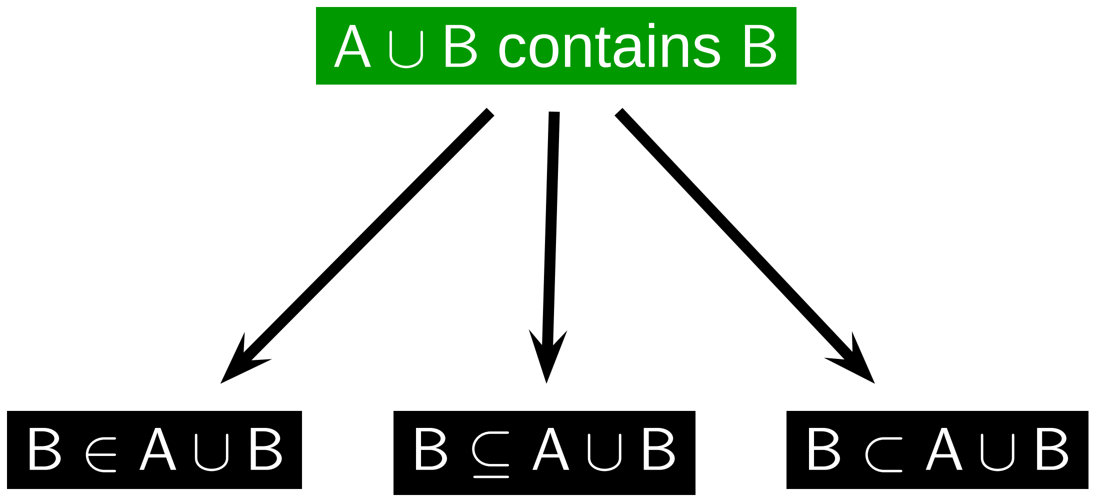
Input Analyzer



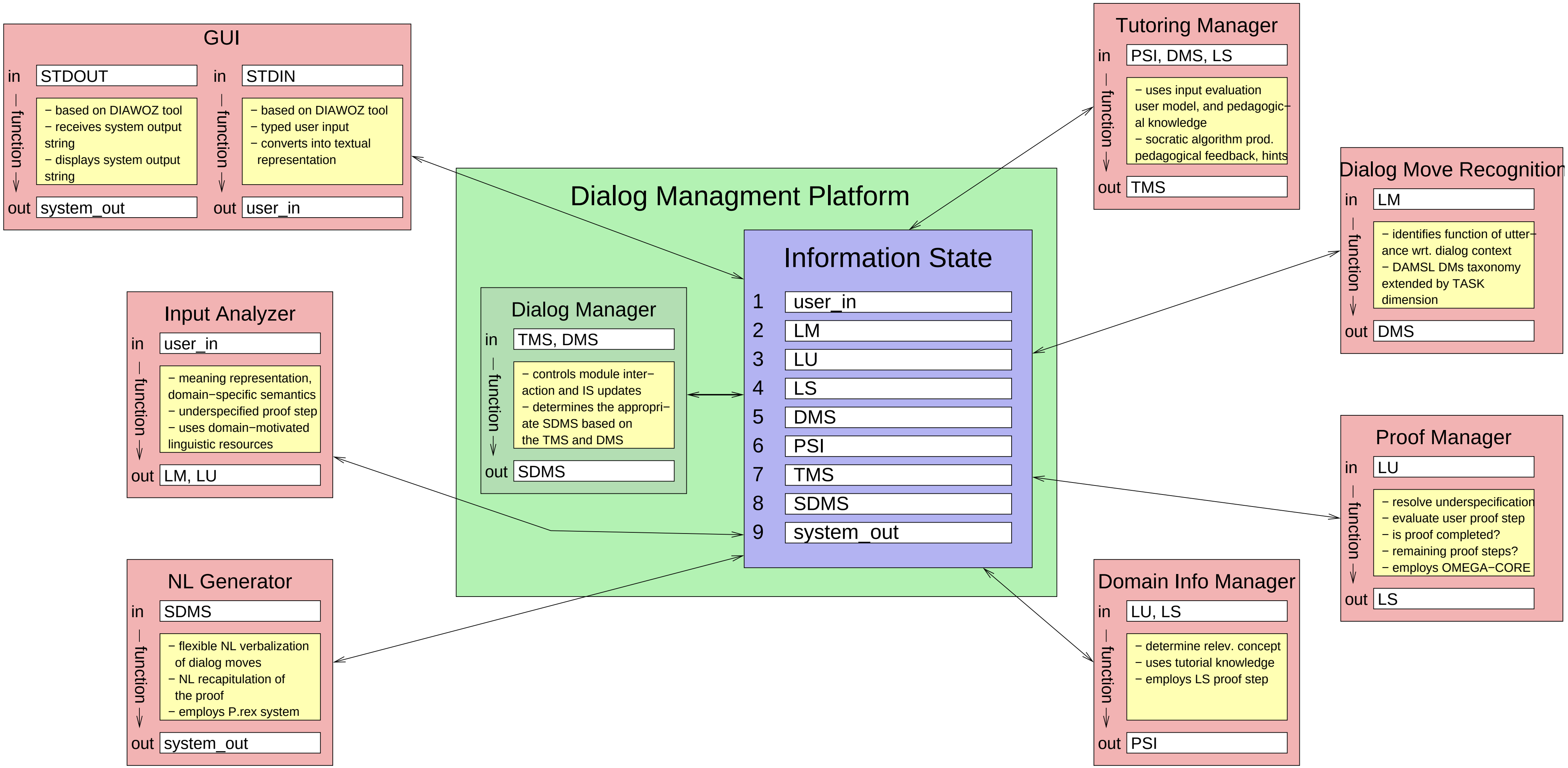
Input Analyzer



In our concrete example:



Dialog Move Recognizer



Dialog Move Recognizer

Dialog Move Recognizer

DIALOG Project

Approach Simulation or highly simplified version (not a research focus)

Dialog Move Recognizer

DIALOG Project

Approach Simulation or highly simplified version (not a research focus)

Input The linguistic meaning of the utterance (+ . . .)

Dialog Move Recognizer

Approach Simulation or highly simplified version (not a research focus)

Input The linguistic meaning of the utterance (+ . . .)

- Example: domain contribution (proof step)

Dialog Move Recognizer

DIALOG Project

Approach Simulation or highly simplified version (not a research focus)

Input The linguistic meaning of the utterance (+ . . .)

- Example: domain contribution (proof step)

Output Suggestion of an appropriate dialog move

UNIVERSITÄT DES SAARLANDES

Dialog Move Recognizer

DIALOG Project

Approach Simulation or highly simplified version (not a research focus)

Input The linguistic meaning of the utterance (+ . . .)

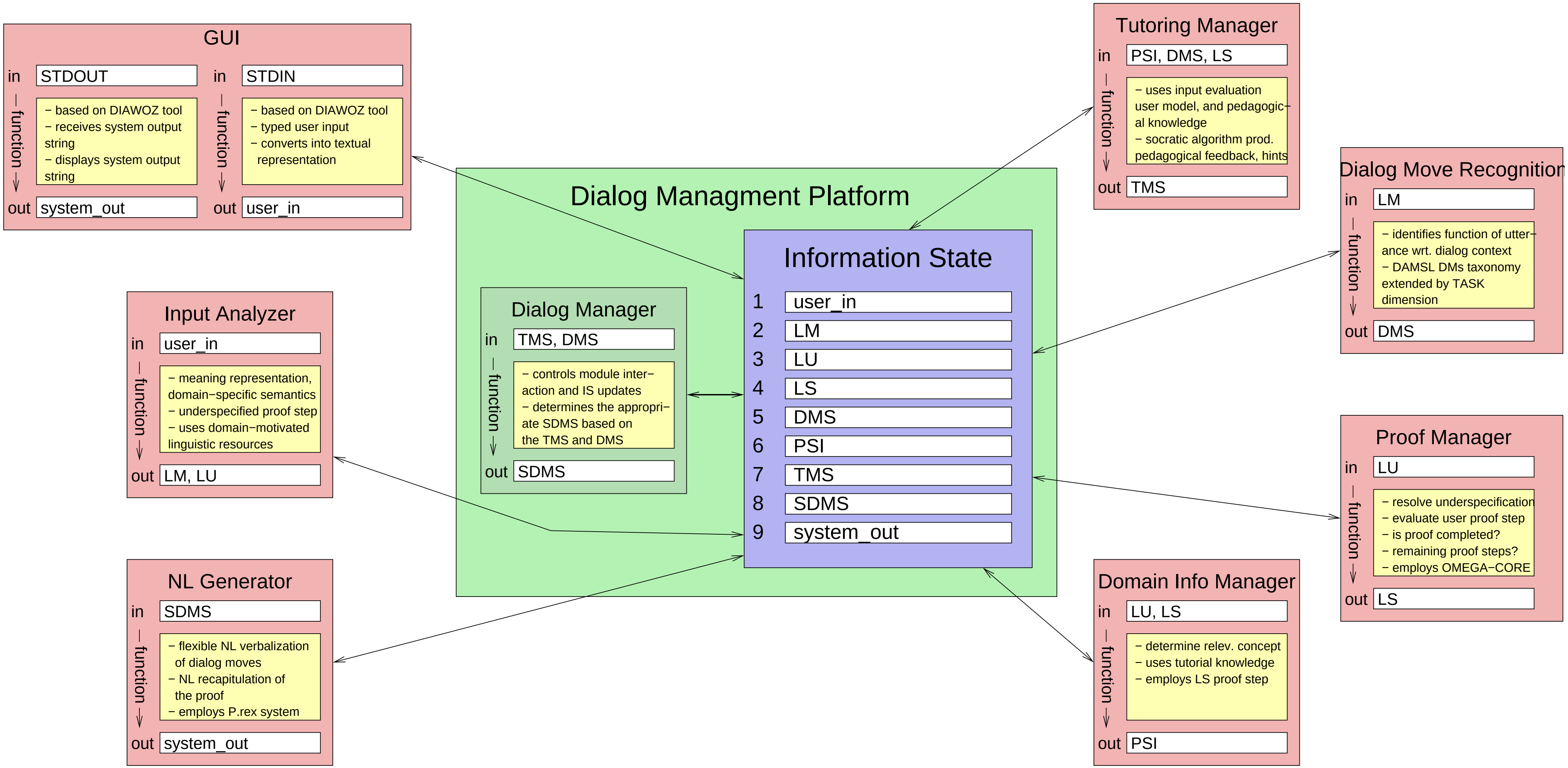
- Example: domain contribution (proof step)

Output Suggestion of an appropriate dialog move

- Example: address the proof step

Mathematical Domain Reasoner

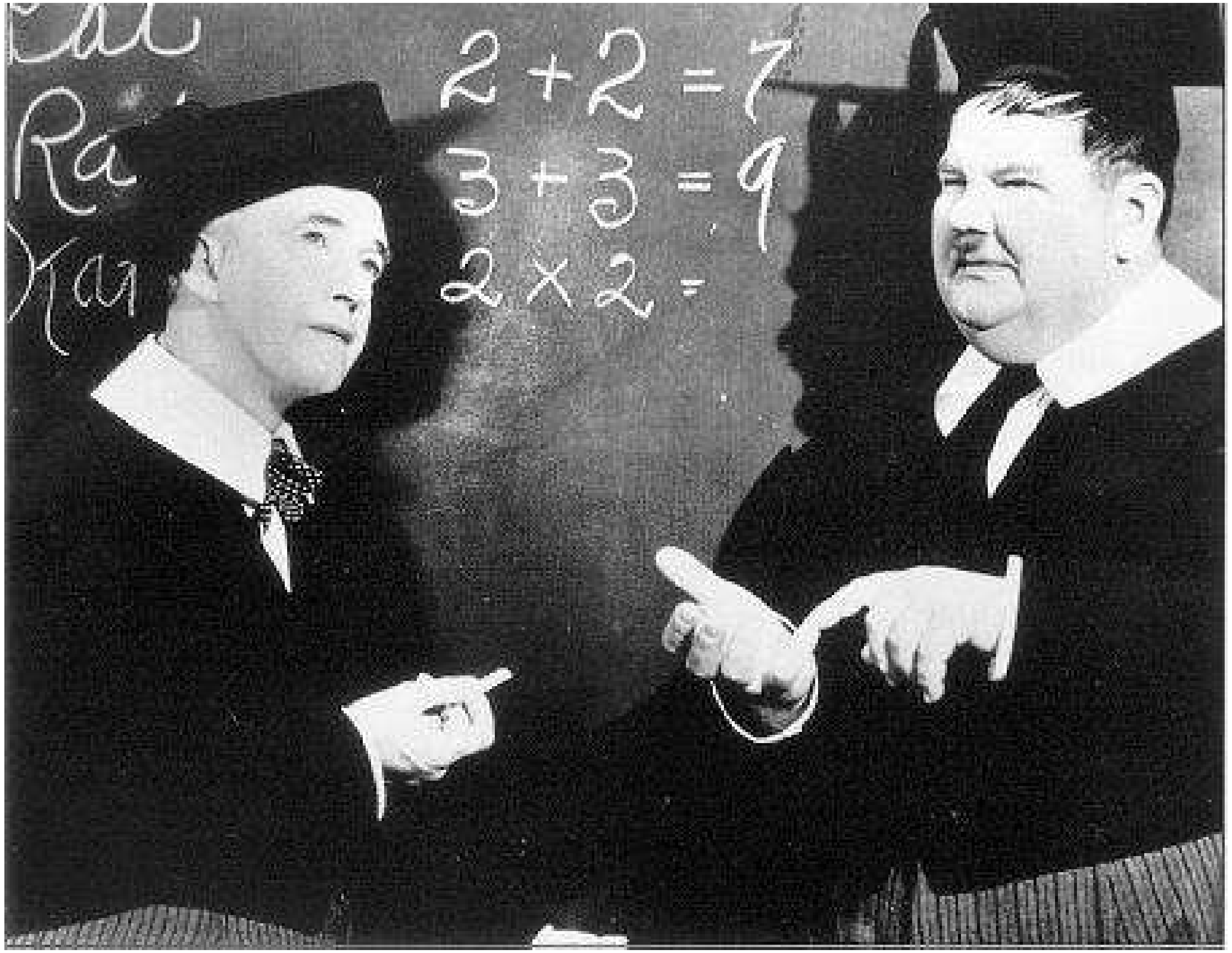
DIALOG Project



UNIVERSITÄT DES SAARLANDES

Mathematical Domain Reasoner

DIALOG Project



A naive first look!

Mathematical Domain Reasoner

DIALOG Project

Approach Extension of proof assistant OMEGA / Development of a new proof manager → CS task (Marvin Schiller, Chad Brown)

Mathematical Domain Reasoner

Approach Extension of proof assistant OMEGA / Development of a new proof manager $\longrightarrow$ CS task (Marvin Schiller, Chad Brown)

Input LU: The domain contribution (uttered proof step) in a semantic representation

Mathematical Domain Reasoner

DIALOG Project

Approach Extension of proof assistant OMEGA / Development of a new proof manager $\longrightarrow$ CS task (Marvin Schiller, Chad Brown)

Input LU: The domain contribution (uttered proof step) in a semantic representation

- Example: LU: $B \in A \cup B \parallel B \subseteq A \cup B \parallel B \subset A \cup B$

Mathematical Domain Reasoner

Approach Extension of proof assistant OMEGA / Development of a new proof manager $\longrightarrow$ CS task (Marvin Schiller, Chad Brown)

Input LU: The domain contribution (uttered proof step) in a semantic representation

- Example: LU: $B \in A \cup B \parallel B \subseteq A \cup B \parallel B \subset A \cup B$

Output An analysis result for each reading in LU; underspecified parts are resolved and made explicit

Mathematical Domain Reasoner

DIALOG Project

Approach Extension of proof assistant OMEGA / Development of a new proof manager $\longrightarrow$ CS task (Marvin Schiller, Chad Brown)

Input LU: The domain contribution (uttered proof step) in a semantic representation

- Example: LU: $B \in A \cup B \parallel B \subseteq A \cup B \parallel B \subset A \cup B$

Output An analysis result for each reading in LU; underspecified parts are resolved and made explicit

- Example: ... see next slides ...

Mathematical Domain Reasoning

$A \cup B$ contains B

$B \in A \cup B$ | $B \subseteq A \cup B$ | $B \subset A \cup B$

Mathematical Domain Reasoning

$A \cup B$ contains B

X

$B \in A \cup B$

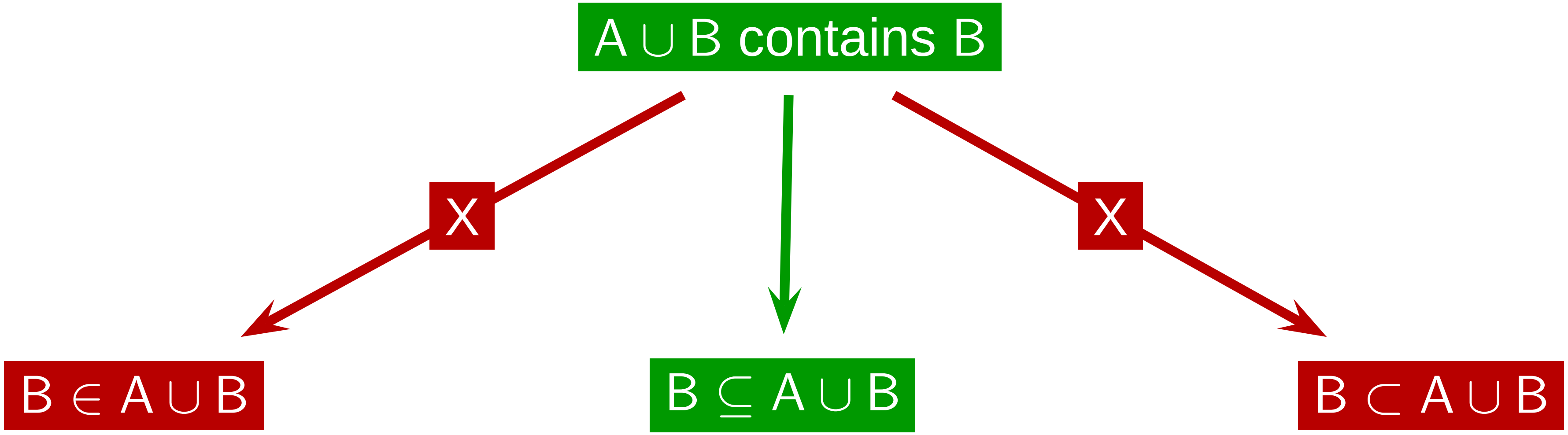
$B \subseteq A \cup B$

$B \subset A \cup B$

type checking

Mathematical Domain Reasoning

DIALOG Project



theorem proving

Mathematical Domain Reasoning

$$
\begin{tikzcd}
& \mathcal{P}((A\cup C)\cap(B\cup C))=\mathcal{P}C \cup(A\cap B) \arrow[dl, "X"'] \arrow[dr] & \\
\mathcal{P}((A\cup C)\cap(B\cup C))=\mathcal{P}(C) \cup(A\cap B) & & \mathcal{P}((A\cup C)\cap(B\cup C))=\mathcal{P}(C \cup(A\cap B))
\end{tikzcd}
$$

type checking

Mathematical Domain Reasoning

DIALOG Project

$\mathcal{K}((A\cup C)\cap(B\cup C))=\mathcal{K}C\cup(A\cap B)$

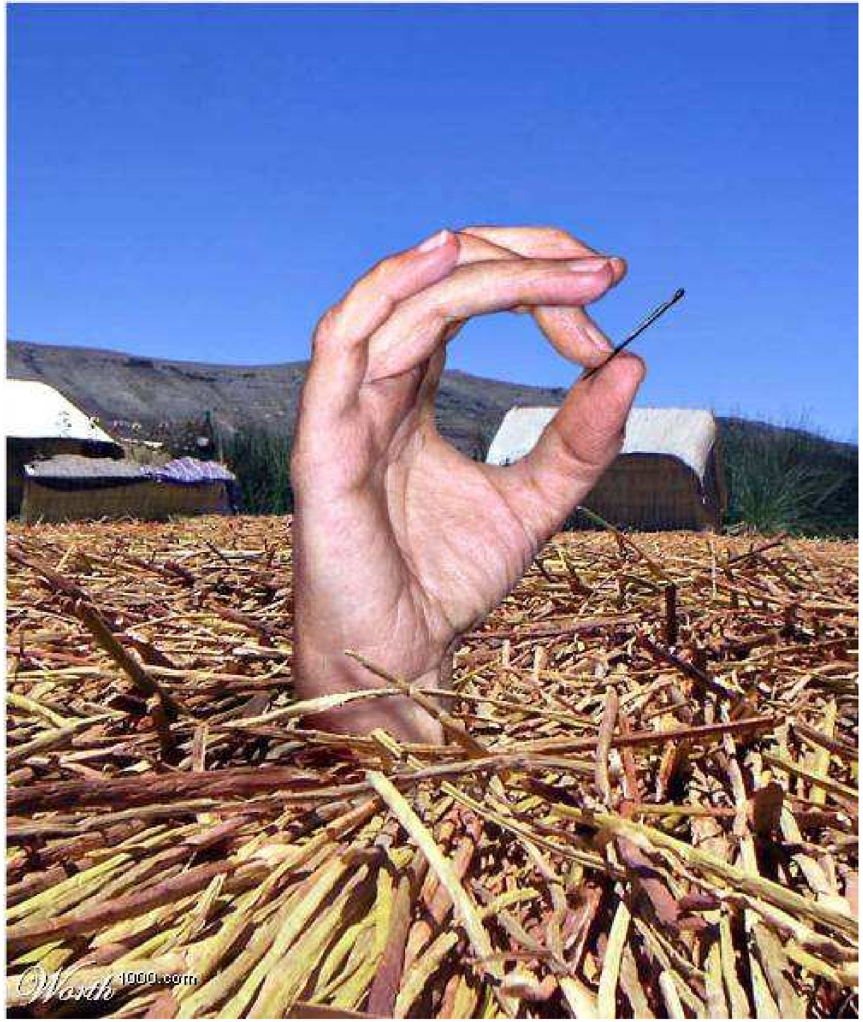
X

$\mathcal{K}((A\cup C)\cap(B\cup C))=\mathcal{K}(C)\cup(A\cap B)$

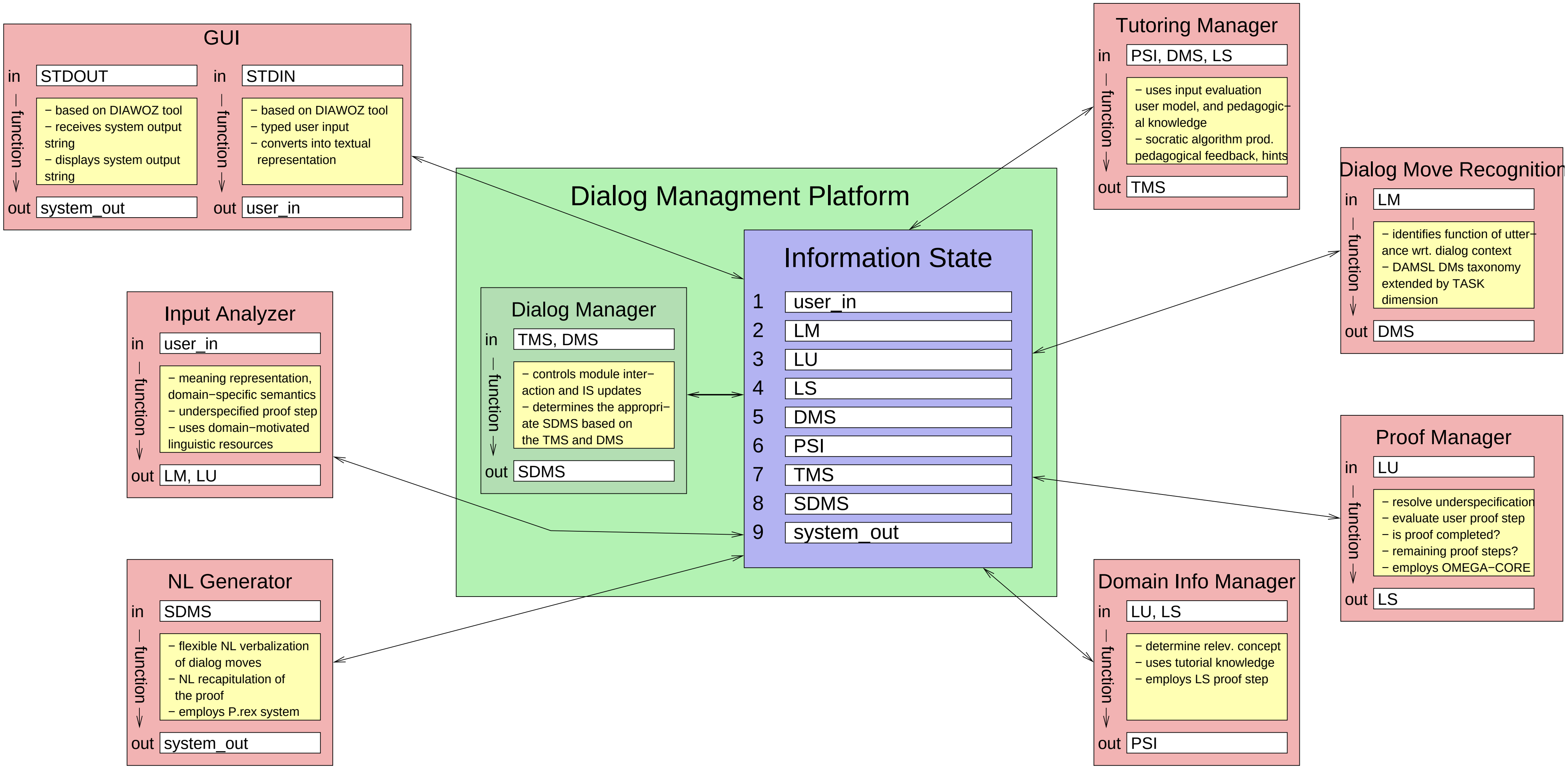
$\mathcal{K}((A\cup C)\cap(B\cup C))=\mathcal{K}(C\cup(A\cap B))$

theorem proving

Domain Information Manager



Domain Information Manager



Domain Information Manager

Approach   Simulation or highly simplified version (not a research focus)

Domain Information Manager

DIALOG Project

Approach Simulation or highly simplified version (not a research focus)

Input LU and its analysis result

Domain Information Manager

DIALOG Project

Approach Simulation or highly simplified version (not a research focus)

Input LU and its analysis result

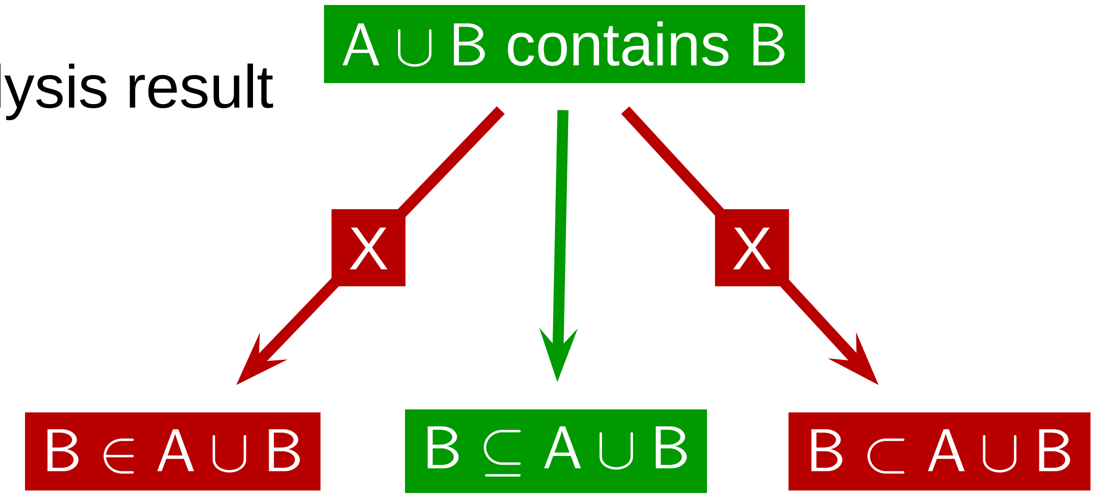
- Example:

Domain Information Manager

Approach Simulation or highly simplified version (not a research focus)

Input LU and its analysis result

- Example:



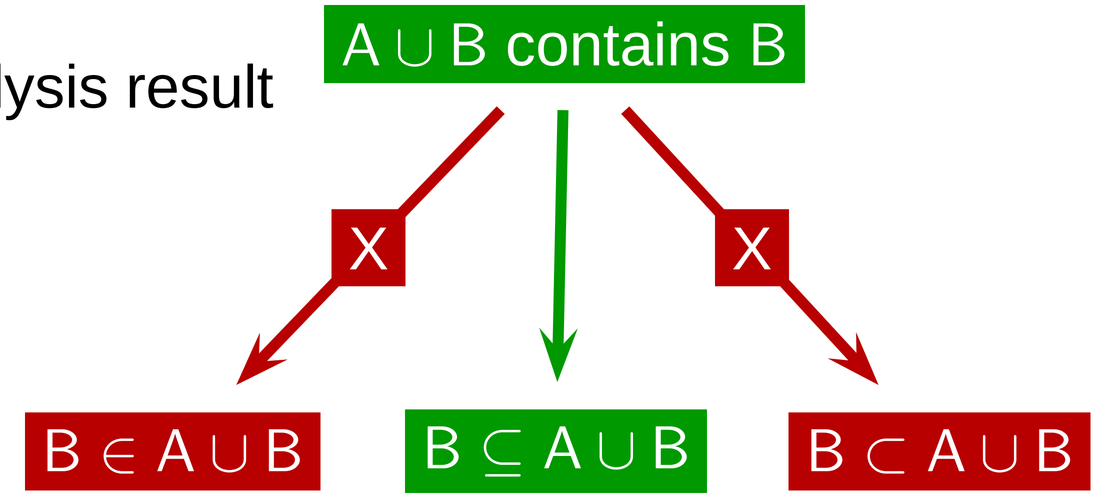
Domain Information Manager

DIALOG Project

Approach Simulation or highly simplified version (not a research focus)

Input LU and its analysis result

- Example:



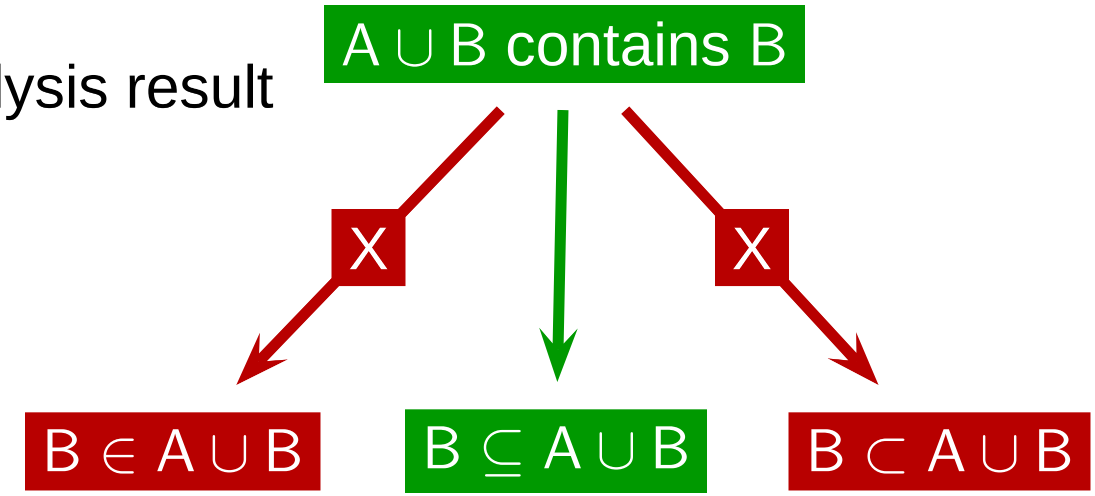
Output The essentially addressed domain concept(s)

Domain Information Manager

Approach Simulation or highly simplified version (not a research focus)

Input LU and its analysis result

- Example:

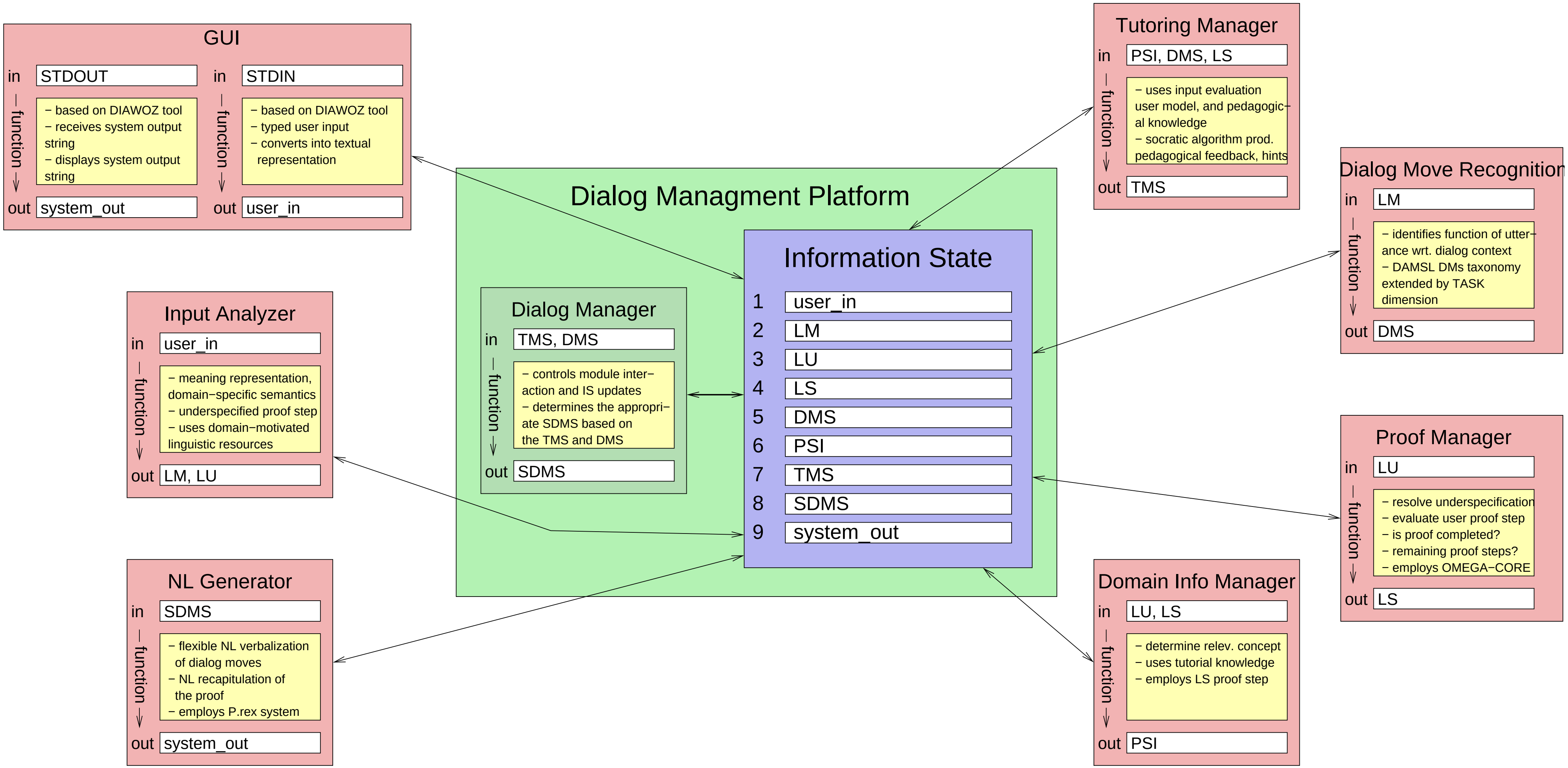


Output The essentially addressed domain concept(s)

- Example: $\subseteq$, $\cup$

Tutoring Manager

Tutoring Manager



Tutoring Manager

Approach Simulation or simple version (Dimitra Tsovaltzis)

Tutoring Manager

Approach Simulation or simple version (Dimitra Tsovaltzis)

Input LU analysis, addressed domain concepts, dialog move suggestion (+ . . .)

Tutoring Manager

Approach Simulation or simple version (Dimitra Tsovaltzis)

Input LU analysis, addressed domain concepts, dialog move suggestion (+ . . .)

- Example: ... as seen before ...

Tutoring Manager

Approach Simulation or simple version (Dimitra Tsovaltzis)

Input LU analysis, addressed domain concepts, dialog move suggestion (+ . . .)

- Example: ... as seen before ...

Output A specification of a tutorial move

Tutoring Manager

Approach Simulation or simple version (Dimitra Tsovaltzis)

Input LU analysis, addressed domain concepts, dialog move suggestion (+ . . .)

- Example: ... as seen before ...

Output A specification of a tutorial move

- Example: signal the correctness of the proof step

Tutoring Manager

Approach Simulation or simple version (Dimitra Tsovaltzis)

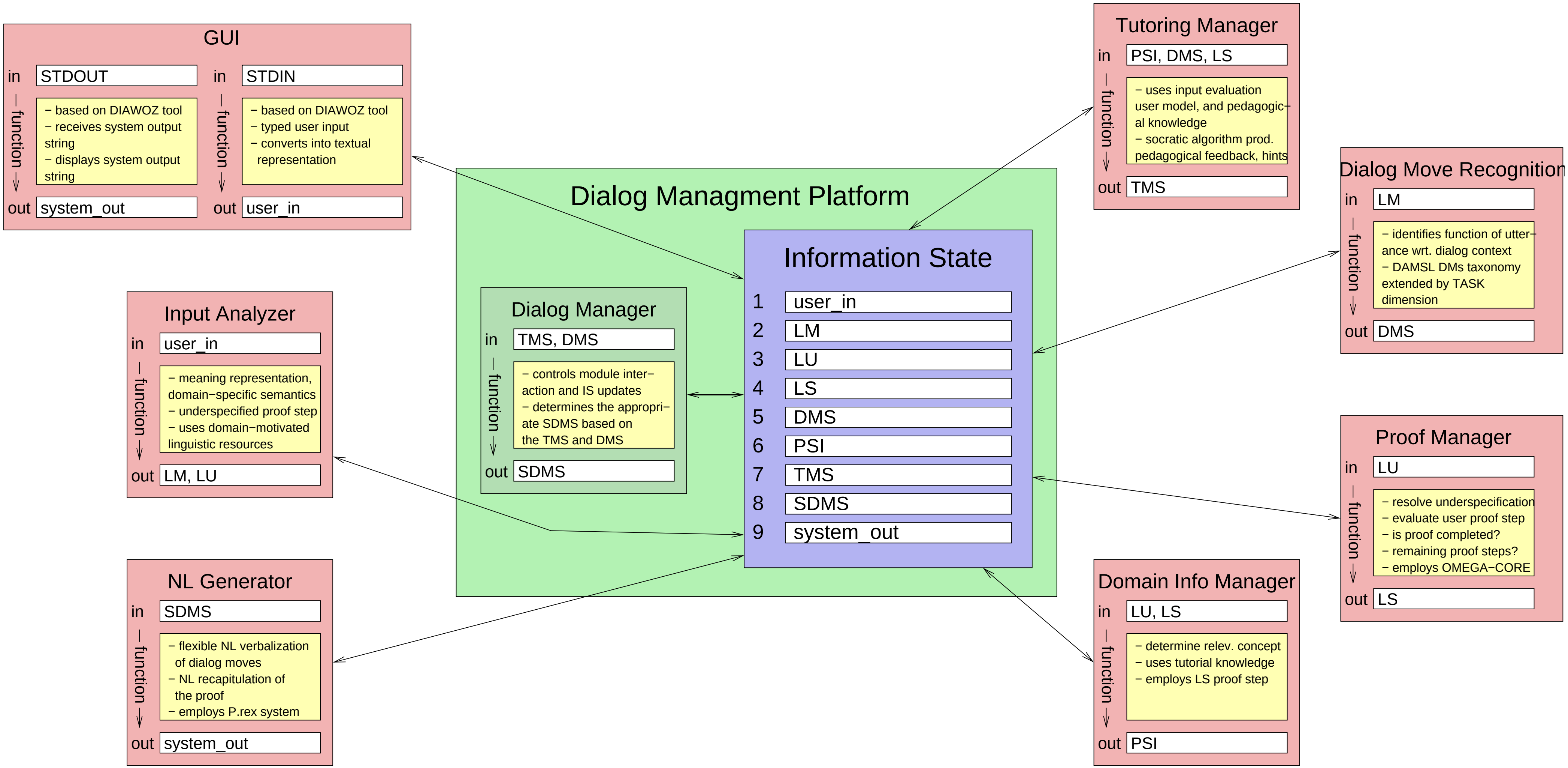
Input LU analysis, addressed domain concepts, dialog move suggestion (+ . . .)

- Example: ... as seen before ...

Output A specification of a tutorial move

- Example: signal the correctness of the proof step

Dialog Manager, NL Generation, GUI, ...



Dialog Manager, NL Generation, GUI, ...

DIALOG Project

. . .

T8: “This is correct! Please continue the proof.”

A Challenge: Proof Step Evaluation

Perspective of Mathematical Domain Reasoning (MDR):

- Support for resolution of Ambiguities and Underspecification

A Challenge: Proof Step Evaluation

Perspective of Mathematical Domain Reasoning (MDR):

- Support for resolution of Ambiguities and Underspecification
- Proof Step Evaluation

A Challenge: Proof Step Evaluation

DIALOG Project

Perspective of Mathematical Domain Reasoning (MDR):

- Support for resolution of Ambiguities and Underspecification
- Proof Step Evaluation
  - Soundness: proof step verifiable by formal system?

A Challenge: Proof Step Evaluation

DIALOG Project

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?

A Challenge: Proof Step Evaluation

DIALOG Project

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?

Research Challenges

DIALOG Project

Perspective of Mathematical Domain Reasoning (MDR):

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

Logical vs Tutorial Dimension

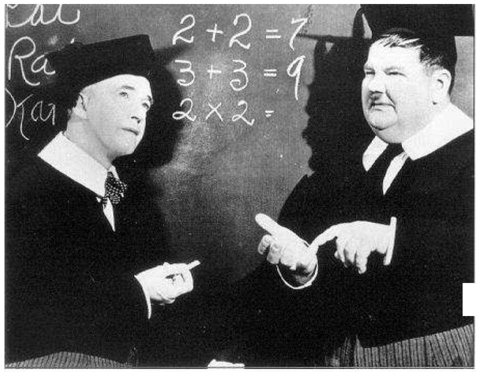
Research Challenges

DIALOG Project

Perspective of Mathematical Domain Reasoning (MDR):

- Support for resolution of Ambiguiti nd Underspecification
- Proof Step Evaluation
  - Soundness: proof erifiable by formal system?
  - Granularity: rgumentative complexity of proof step?
  - Relevanc proof step needed/useful in achieving the goal?

Logical vs Tutorial Dimension



— declarative abstract level sketches ➤

Communication Gap

← procedural calculus level proofs —

Proof Step Evaluation

DIALOG Project

(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

DIALOG Project

(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

UNIVERSITÄT DES SAARLANDES

Proof Step Evaluation: How?

DIALOG Project

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$
?
(G) $D \vee E$

New:

We show E.

(1) . . .
(2) . . .
(3) . . .
(4) . . .
?
(G') E
(G) . . .

PSE:

Soundness

Granularity

Relevance

Proof Step Evaluation: How?

DIALOG Project

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$
?
(G) $D \vee E$

New:

We show E.

(1) ...
(2) ...
(3) ...
(4) ...
?
(G') E
(G) ...

PSE:

Soundness

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

Granularity

Relevance

Proof Step Evaluation: How?

DIALOG Project

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$
?
(G) $D \vee E$

New:

We show E.

(1) ...
(2) ...
(3) ...
(4) ...
?
(G') E
(G) ...

PSE:

Soundness

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

Granularity

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

Relevance

Proof Step Evaluation: How?

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$
?
(G) $D \vee E$

New:

We show E.

(1) ...
(2) ...
(3) ...
(4) ...
?
(G') E
(G) ...

PSE:

Soundness

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

Granularity

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

Relevance

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

Proof Step Evaluation: Claim

DIALOG Project

Granularity and Relevance

call for

cognitively adequate abstract level proofs

+

enumeration of (some) proof alternatives

A closer look at Granularity

DIALOG Project

In the 2005 experiment:

- the wizards additionally had to categorize the contributions

A closer look at Granularity

In the 2005 experiment:

- the wizards additionally had to categorize the contributions

student]	Let $(x, y) \in (R \circ S)^{-1}$
tutor]	Correct. Good start!

correct	appropriate	relevant

A closer look at Granularity

DIALOG Project

In the 2005 experiment:

- the wizards additionally had to categorize the contributions

student]	Let $(x, y) \in (R \circ S)^{-1}$		
tutor]	Correct. Good start!		
	correct	appropriate	relevant

- this data has been recorded and is now available for evaluation

Granularity ... in the 2005 Corpus

Granularity ... the size/argumentative complexity of a proof step

Granularity factors:

- abstraction
- explicitness/underspecification
- cognitive effort

Granularity ... in the 2005 Corpus

- abstraction

student]	let $(x, y) \in (R \cup S) \circ T$, then $(x, z) \in (R \cup S) \wedge (z, y) \in T$
	(...) then $(x, z) \in R \vee (x, z) \in S$
tutor]	This statement is true.
student]	what can be concluded from $(A \vee B) \wedge C$?
tutor]	Then for example it holds $(A \wedge C) \vee (B \wedge C)$
student]	then holds $((x, z) \in R \wedge (z, y) \in T) \vee ((x, z) \in S \wedge (z, y) \in T)$

- explicitness/underspecification
- cognitive effort

Granularity ... in the 2005 Corpus

- abstraction
- explicitness/underspecification

student 13]	$(R \cup S) \circ T$ thus is
...	
student 19]	$(R \circ T) \cup (S \circ T)$ thus is (...)
tutor]	Correct. Can you also indicate according to which law you have transformed input 13 to the current input 19?
student 20]	“distributivity law”

- cognitive effort

Granularity ... in the 2005 Corpus

- abstraction
- explicitness/underspecification
- cognitive effort

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!

Granularity Evaluation

DIALOG Project

Mechanize granularity ratings with thm. proving techniques:

- Use ΩMEGA framework
- Hypothesis: granularity level of a mathematical statement is related to number of inference steps required for its justification
- Calculi: Gentzen's ND (Gentzen 1934) and "Psychology of Proof" (Rips 1994)
- Granularity analysis framework for proofs
- Evaluation: compare mechanical classification to expert's ratings

Granularity Evaluation

What is Mechanized/Automated Theorem Proving?

Example on blackboard:

$$B \subseteq A \cup B$$

Granularity Evaluation

- Example:

student]	$(x, y) \in (S^{-1} \circ R^{-1}) \Leftrightarrow \exists z[(z, x) \in S \wedge (y, z) \in R]$			
tutor]	This is correct!	correct	too coarse-grained	relevant

Granularity Evaluation

DIALOG Project

- Example:

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

tutor] This is correct! | correct | too coarse-grained | relevant |

- generate LU:

...

(equiv
  ((composed M (inverse-1 S) (inverse-1 R)) x y)
  (exists (lam (z a) (and (and (M z) (S z x)) (R y z)))))

...

Granularity Evaluation

- Example:

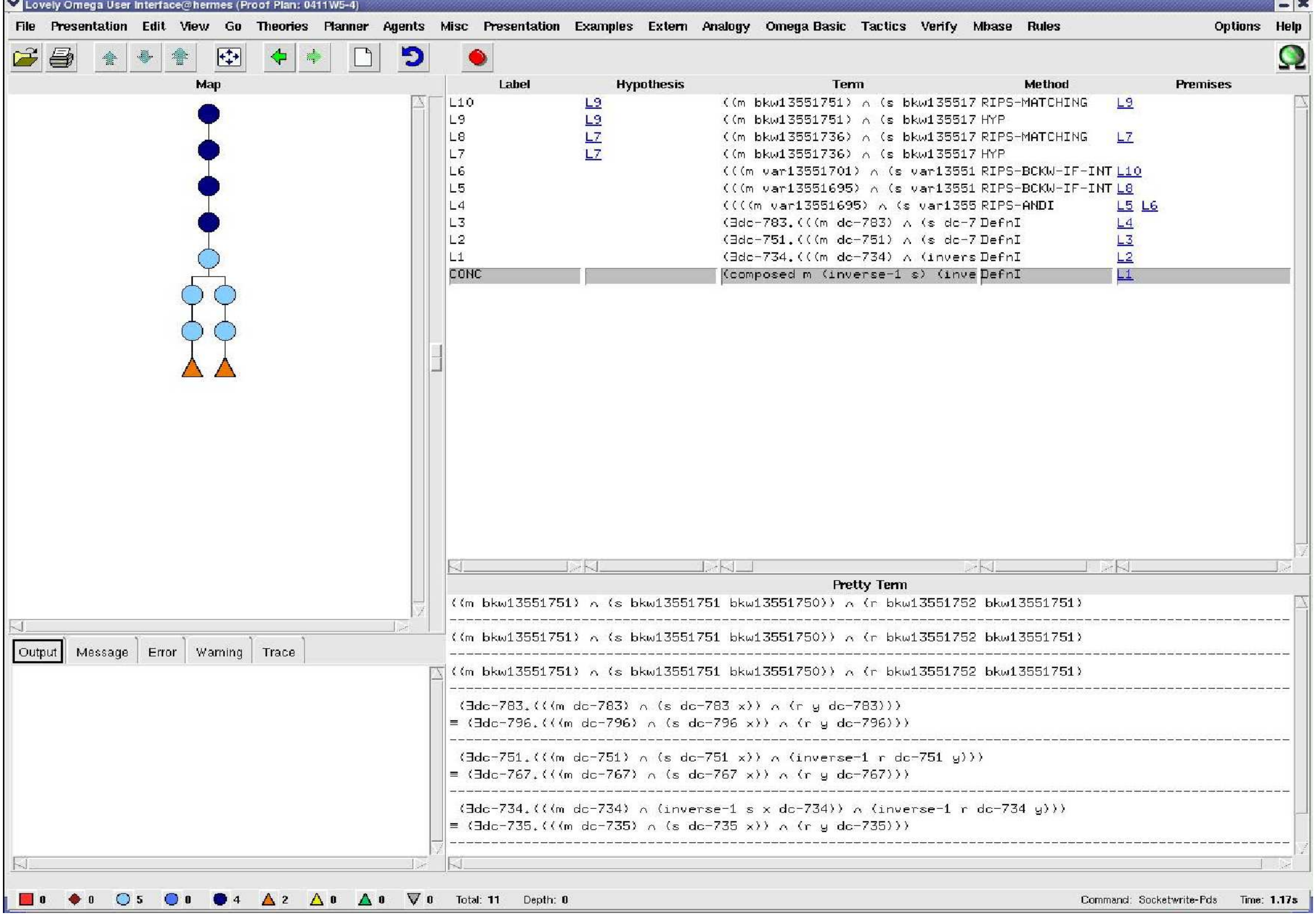
student] $(x, y) \in (S^{-1} \circ R^{-1}) \Leftrightarrow \exists z[(z, x) \in S \wedge (y, z) \in R]$

tutor] This is correct! | correct | too coarse-grained | relevant |

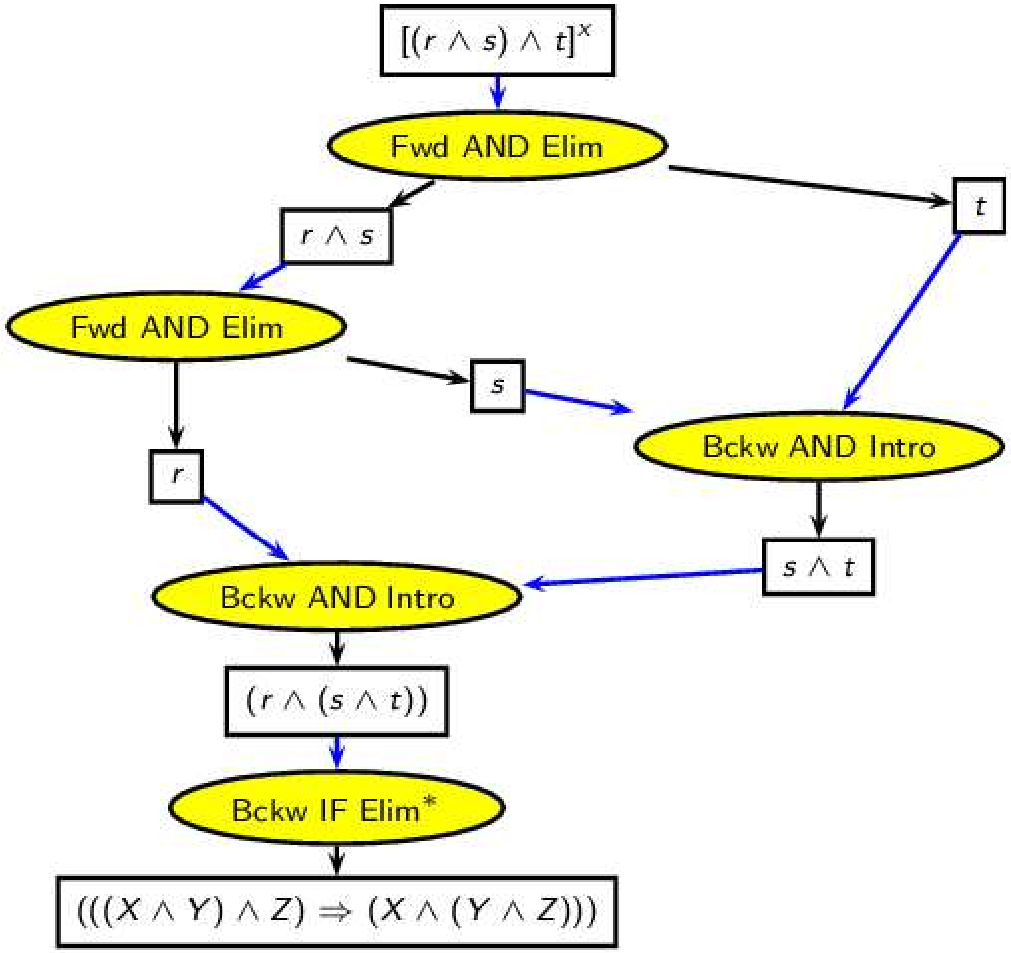
- generate LU:

. . .
(equiv
  ((composed M (inverse-1 S) (inverse-1 R)) x y)
  (exists (lam (z a) (and (and (M z) (S z x)) (R y z)))))
. . .

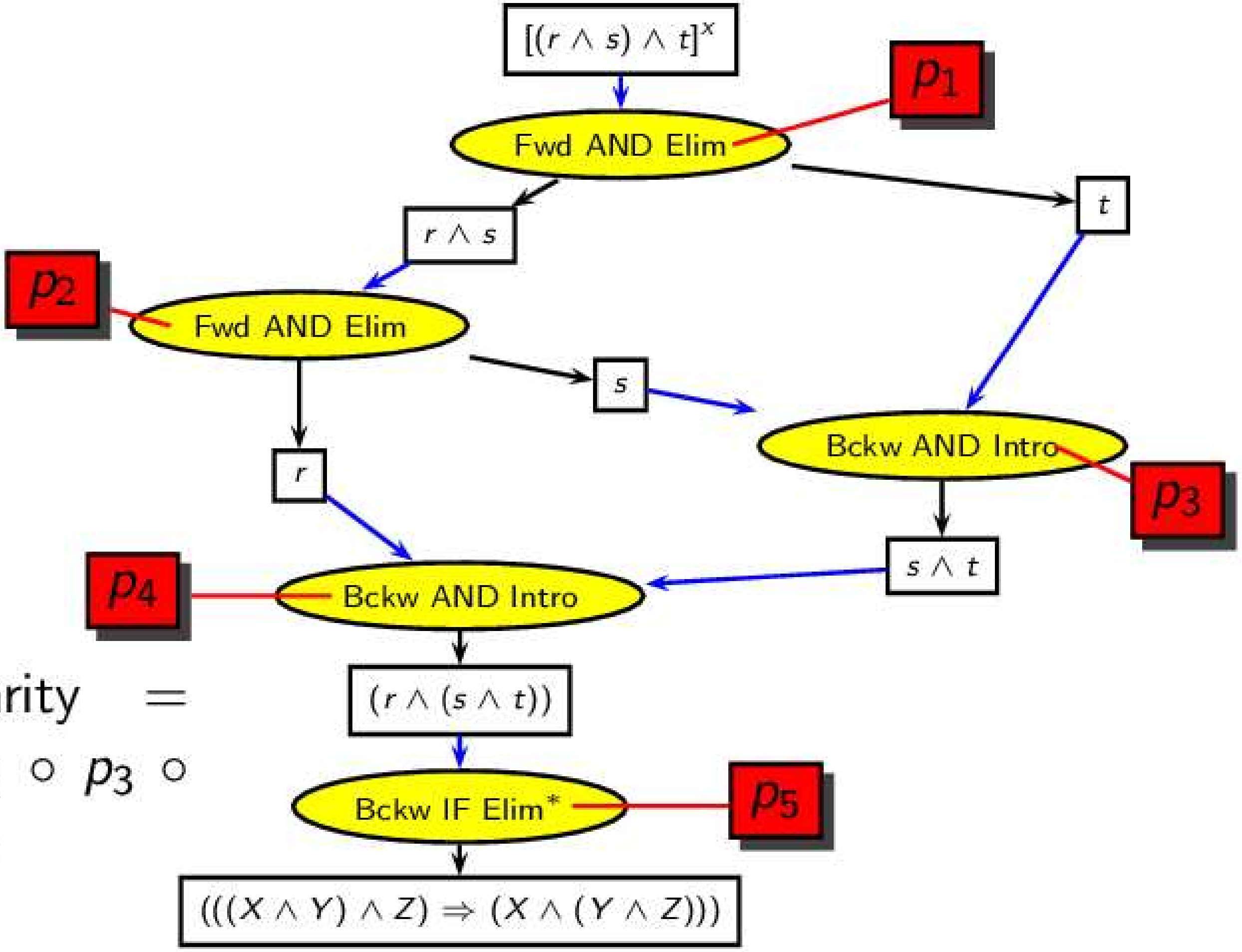
- analyze with domain reasoner . . .


Lovely Omega User Interface@hermes (Proof Plan: 0411W5-4)
File Presentation Edit View Go Theories Planner Agents Misc Presentation Examples Extern Analogy Omega Basic Tactics Verify Mbase Rules Options Help
Map
Label Hypothesis Term Method Premises
L10 L9 ((m bkw13551751) ∧ (s bkw135517 RIPS-MATCHING L9
L9 L9 ((m bkw13551751) ∧ (s bkw135517 HYP
L8 L7 ((m bkw13551736) ∧ (s bkw135517 RIPS-MATCHING L7
L7 L7 ((m bkw13551736) ∧ (s bkw135517 HYP
L6 (((m var13551701) ∧ (s var13551 RIPS-BCKW-IF-INT L10
L5 (((m var13551695) ∧ (s var13551 RIPS-BCKW-IF-INT L8
L4 ((((m var13551695) ∧ (s var1355 RIPS-ANDI L5 L6
L3 (∃dc-783.(((m dc-783) ∧ (s dc-7 DefnI L4
L2 (∃dc-751.(((m dc-751) ∧ (s dc-7 DefnI L3
L1 (∃dc-734.(((m dc-734) ∧ (invers DefnI L2
CONC (composed m (inverse-1 s) (inve DefnI L1
Pretty Term
((m bkw13551751) ∧ (s bkw13551751 bkw13551750)) ∧ (r bkw13551752 bkw13551751)
((m bkw13551751) ∧ (s bkw13551751 bkw13551750)) ∧ (r bkw13551752 bkw13551751)
((m bkw13551751) ∧ (s bkw13551751 bkw13551750)) ∧ (r bkw13551752 bkw13551751)
(∃dc-783.(((m dc-783) ∧ (s dc-783 x)) ∧ (r y dc-783)))
≡ (∃dc-796.(((m dc-796) ∧ (s dc-796 x)) ∧ (r y dc-796)))
(∃dc-751.(((m dc-751) ∧ (s dc-751 x)) ∧ (inverse-1 r dc-751 y)))
≡ (∃dc-767.(((m dc-767) ∧ (s dc-767 x)) ∧ (r y dc-767)))
(∃dc-734.(((m dc-734) ∧ (inverse-1 s x dc-734)) ∧ (inverse-1 r dc-734 y)))
≡ (∃dc-735.(((m dc-735) ∧ (s dc-735 x)) ∧ (r y dc-735)))
Output Message Error Warning Trace
0 0 5 0 4 2 0 0 0 Total: 11 Depth: 0
Command: Socketwrite-Pds Time: 1.17s

Granularity Evaluation

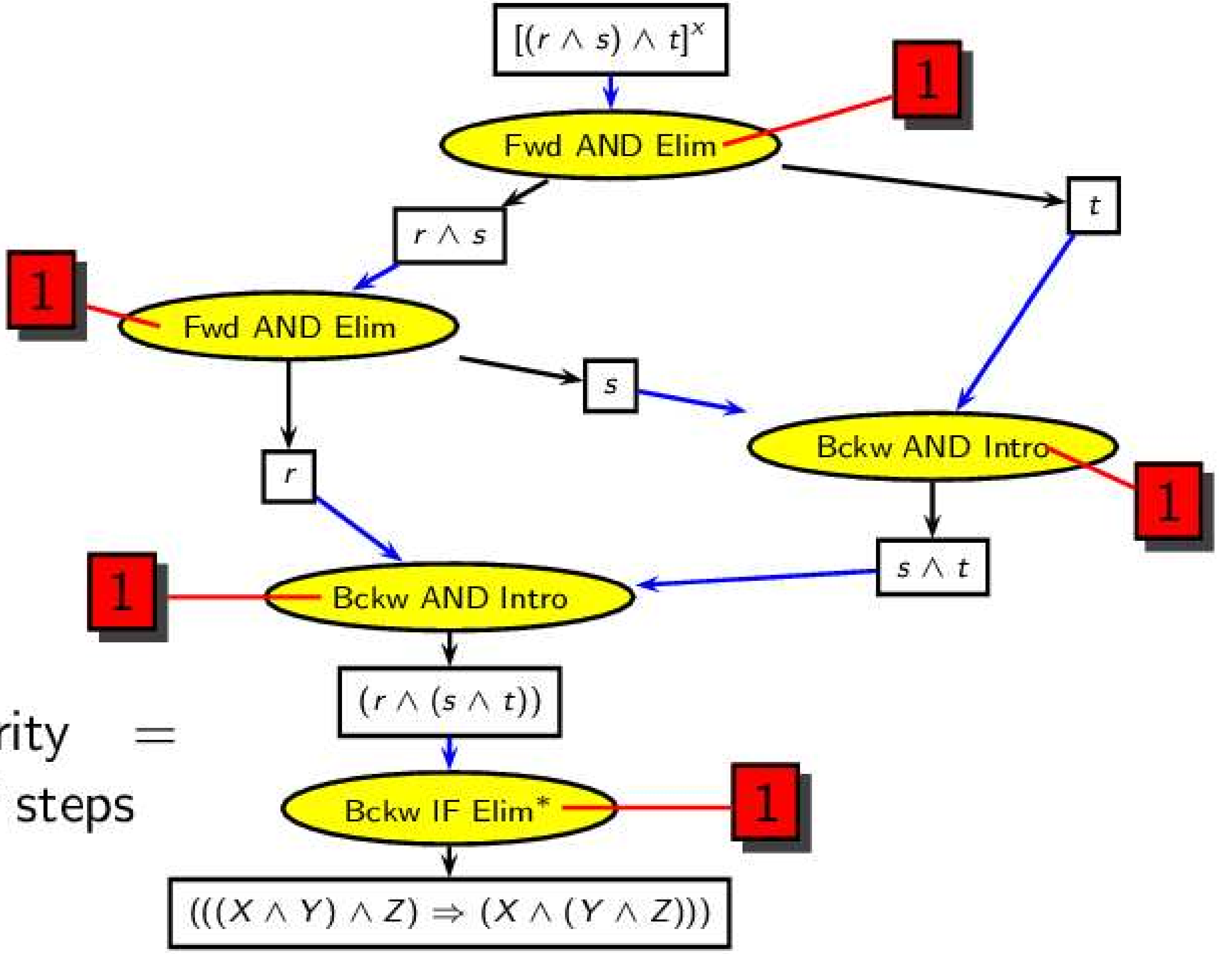


Granularity Evaluation



$$\text{Granularity} = p_1 \circ p_2 \circ p_3 \circ p_4 \circ p_5$$

Granularity Evaluation



Granularity $=$
\# proof steps

Granularity Evaluation

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			
[Gentzen34]			

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

Granularity Evaluation

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]			

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

Granularity Evaluation

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

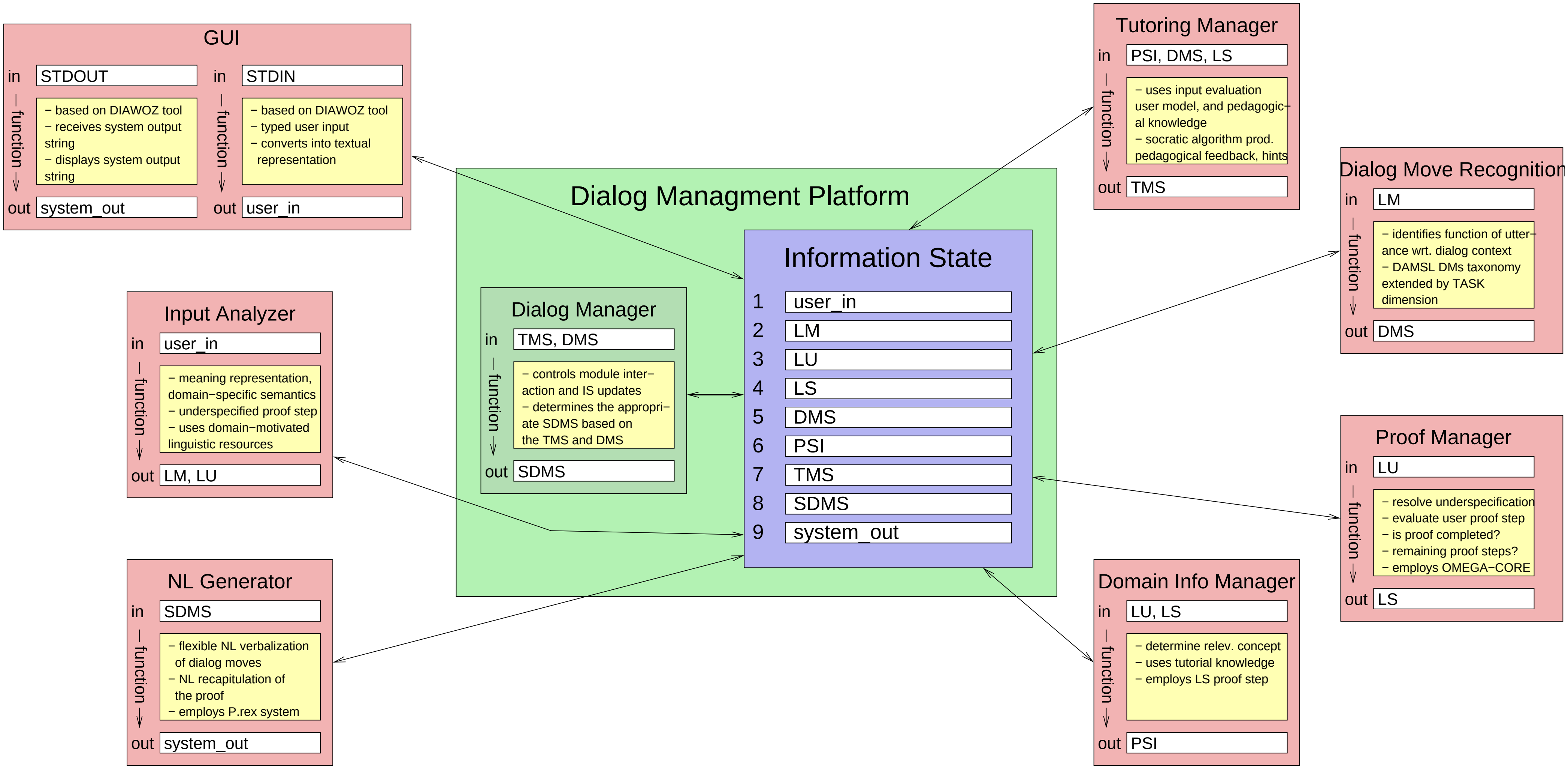
Number of justifying proof steps for PSYCOP and Gentzen’s NK.

Granularity: Further Directions

1. Develop proof analysis mechanisms:
   - Enumeration & analysis of proof alternatives.
   - Develop & investigate complex evaluation hypotheses.
   - Develop & investigate cognitively “realistic” proof systems.
   - Relationship: granularity $\leftrightarrow$ relevance ?
2. Apply techniques and evaluate them empirically .

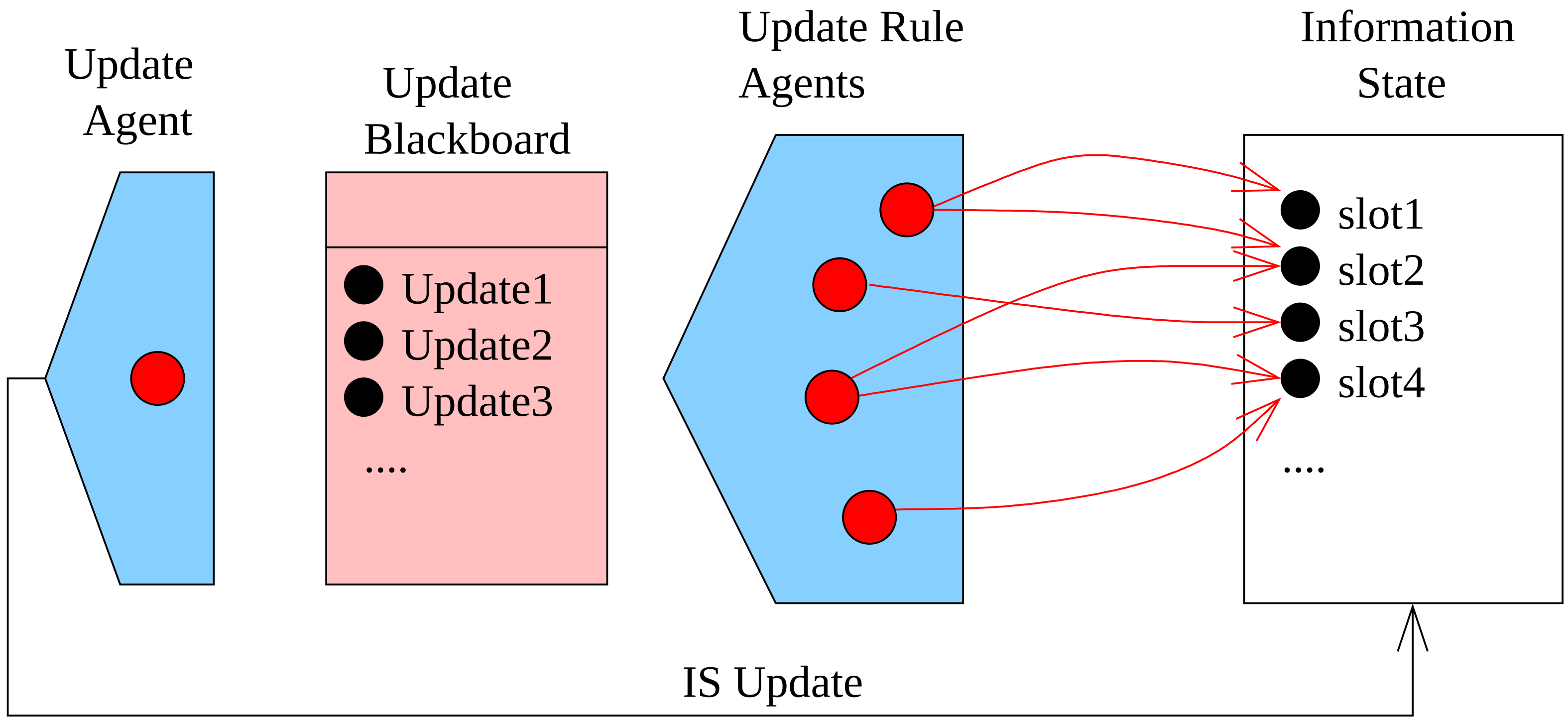
An Agent-based DIALOG Manager

An Agent-based DIALOG Manager



An Agent-based DIALOG Manager

DIALOG Project



Notion of agents here: software agents

Defining a Dialog Manager

- Information State
  - A set of named, typed slots
  - Readable by update rules and writable by the update agent

Defining a Dialog Manager

- Information State
  - A set of named, typed slots
  - Readable by update rules and writable by the update agent
- Update Rules
  - Compute updates of IS slots
  - Consist of preconditions, sideconditions, effects:

$$\frac{\{(s_1, b_1), \ \ldots \ , (s_j, b_j)\}}{\{(s_1, f_1), \ \ldots \ , (s_l, f_l)\}} n \ < (v_1, f_1), \ \ldots \ , (v_k, f_k) >$$

Update Rule Execution

- Current IS satisfies the preconditions $\Rightarrow$ Rule can fire
  - Sidecondition expressions are evaluated and bound to variables
  - Effect expressions evaluated
  - IS update is computed and written to update blackboard

Update Rule Execution

- Current IS satisfies the preconditions $\Rightarrow$ Rule can fire
  - Sidecondition expressions are evaluated and bound to variables
  - Effect expressions evaluated
  - IS update is computed and written to update blackboard

Example:

$$\frac{\{\mathtt{equal}(\mathtt{lm},''\mathtt{domain}-\mathtt{contribution}'')\}\ \{\mathtt{non\_empty}(\mathtt{lu})\}}{\{\mathtt{eval\_lu} \rightarrow \mathtt{r}\}}\ \text{PM} < \mathtt{r} := \mathtt{pm\_analyse}(\mathtt{lu}) >$$

Update Rule Execution

- Current IS satisfies the preconditions $\Rightarrow$ Rule can fire
  - Sidecondition expressions are evaluated and bound to variables
  - Effect expressions evaluated
  - IS update is computed and written to update blackboard

Example:

$$\frac{\{\mathtt{equal}(\mathtt{lm},''\,\mathtt{domain}-\mathtt{contribution}'')\}\ \{\mathtt{non\_empty}(\mathtt{lu})\}}{\{\mathtt{eval\_lu}\rightarrow\mathtt{r}\}}\ \mathrm{PM} < \mathtt{r} := \mathtt{pm\_analyse}(\mathtt{lu}) >$$

- Update agent monitors update blackboard & executes update
- Other agents see change in IS

Benefits of new Dialog Manager

What do we gain from this approach?

- Concurrency, flexibility

Benefits of new Dialog Manager

What do we gain from this approach?

- Concurrency, flexibility
- A natural way to integrate external systems

Benefits of new Dialog Manager

What do we gain from this approach?

- Concurrency, flexibility
- A natural way to integrate external systems
- Application of heuristics in update strategy

Benefits of new Dialog Manager

DIALOG Project

What do we gain from this approach?

- Concurrency, flexibility
- A natural way to integrate external systems
- Application of heuristics in update strategy
- Reasoning on instantiated updates

Benefits of new Dialog Manager

What do we gain from this approach?

- Concurrency, flexibility
- A natural way to integrate external systems
- Application of heuristics in update strategy
- Reasoning on instantiated updates
- Better support for interleaving system modules

Conclusion

- ‘Natural language dialog on mathematical proofs’ is a very ambitious research task

Conclusion

- ‘Natural language dialog on mathematical proofs’ is a very ambitious research task
- Interdisciplinary – solutions and methods from different disciplines need to be combined

Conclusion

- ‘Natural language dialog on mathematical proofs’ is a very ambitious research task
- Interdisciplinary – solutions and methods from different disciplines need to be combined
- Instance of a big integrated AI system

Conclusion

- ‘Natural language dialog on mathematical proofs’ is a very ambitious research task
- Interdisciplinary – solutions and methods from different disciplines need to be combined
- Instance of a big integrated AI system
- Basic research questions – e.g., cognitively adequate proofs

Conclusion

- ‘Natural language dialog on mathematical proofs’ is a very ambitious research task
- Interdisciplinary – solutions and methods from different disciplines need to be combined
- Instance of a big integrated AI system
- Basic research questions – e.g., cognitively adequate proofs
- Technical and software engineering issues

Conclusion

- ‘Natural language dialog on mathematical proofs’ is a very ambitious research task
- Interdisciplinary – solutions and methods from different disciplines need to be combined
- Instance of a big integrated AI system
- Basic research questions – e.g., cognitively adequate proofs
- Technical and software engineering issues

$\Rightarrow$ . . . A fascinating research area

Conclusion

- ‘Natural language dialog on mathematical proofs’ is a very ambitious research task
- Interdisciplinary – solutions and methods from different disciplines need to be combined
- Instance of a big integrated AI system
- Basic research questions – e.g., cognitively adequate proofs
- Technical and software engineering issues

$\Rightarrow$ . . . A fascinating research area

- Several research activities not sufficiently addressed in talk:

Conclusion

- ‘Natural language dialog on mathematical proofs’ is a very ambitious research task
- Interdisciplinary – solutions and methods from different disciplines need to be combined
- Instance of a big integrated AI system
- Basic research questions – e.g., cognitively adequate proofs
- Technical and software engineering issues

$\Rightarrow$ . . . A fascinating research area

- Several research activities not sufficiently addressed in talk:
  - NL Analysis, Tutoring, Dialog Planning, NL Generation, Maths Knowledge Bases, Theorem Proving, . . .

Questions?