

Automata Tutor and what we learned from building an online teaching tool

Rajeev Alur (Penn), Loris D'Antoni (Penn), Sumit Gulwani (MSR), Bjoern Hartmann (Berkeley), Dileep Kini (UIUC), Mahesh Viswanathan (UIUC), Matt Weaver (Penn), Alex Weinert (RWTH)



Why automata?

- Part of any CS curriculum
- Students don't like them
- Hard and tedious to grade

Our goal

Build an online tool for teaching all core topics in formal language theory

DFA's

Pumping lemma

Turing machines

Regular expressions

NFA's

Automata meta-constructions

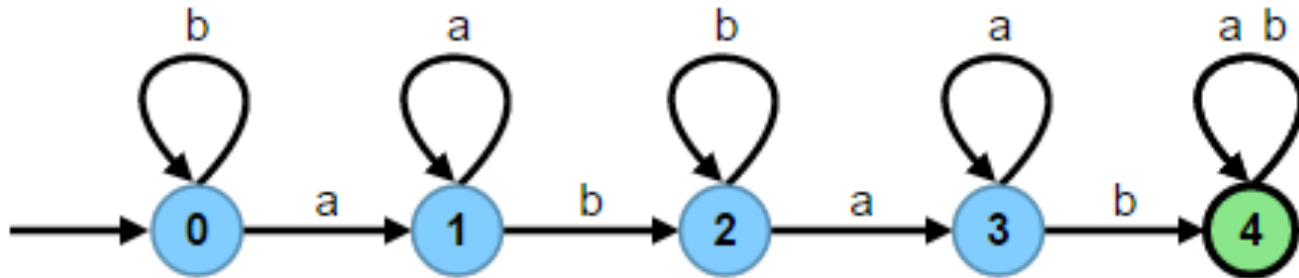
Context-free grammars

HOW DOES IT WORK?

Main idea

- For each type of mistake
 - Create a way to identify it
 - A corresponding metric (for grading)
 - Produce feedback message

Problem syntactic mistake



The problem description was

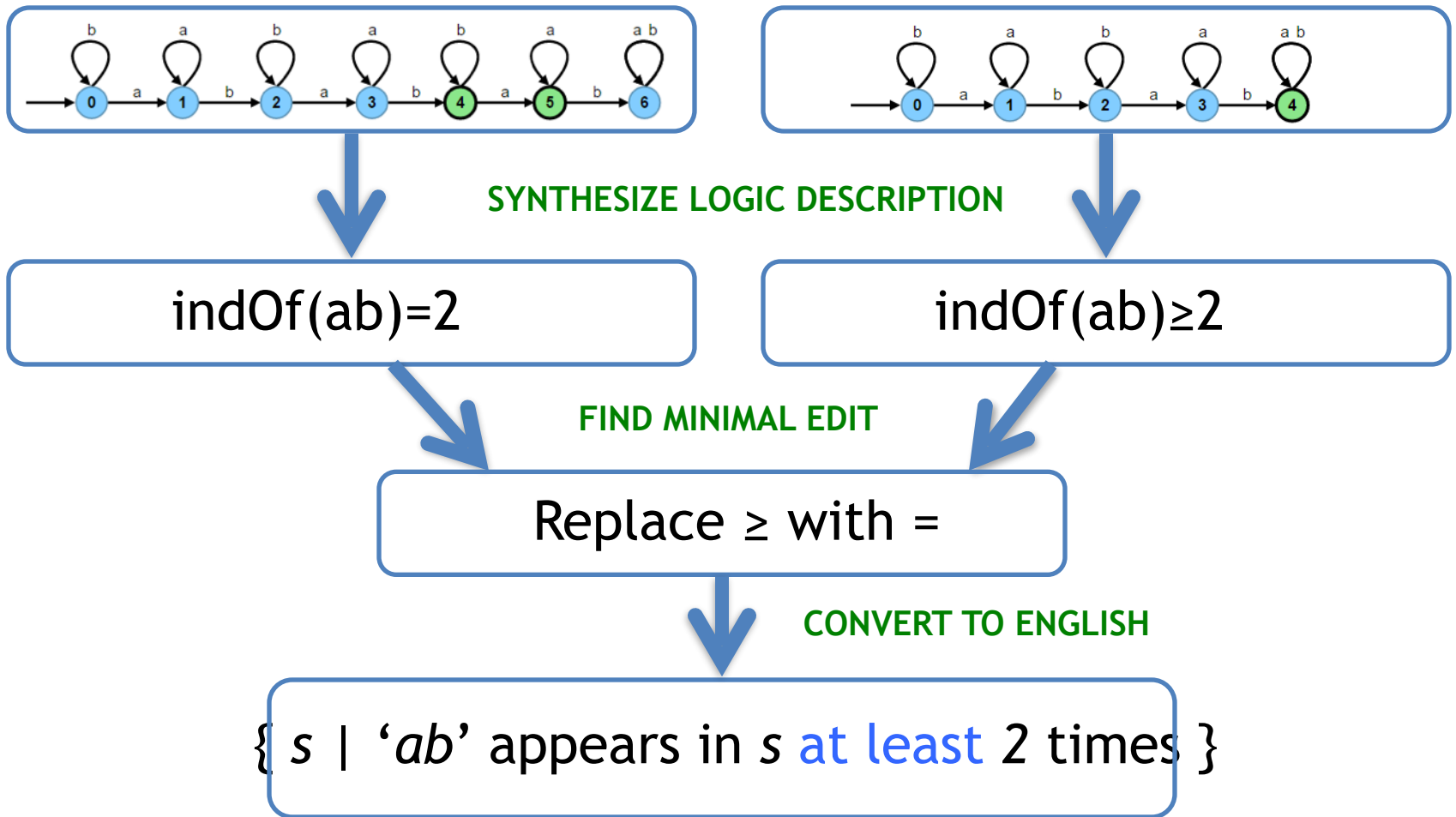
$\{ s \mid \text{'ab' appears in } s \text{ exactly 2 times} \}$

The student instead drew the DFA that accepts

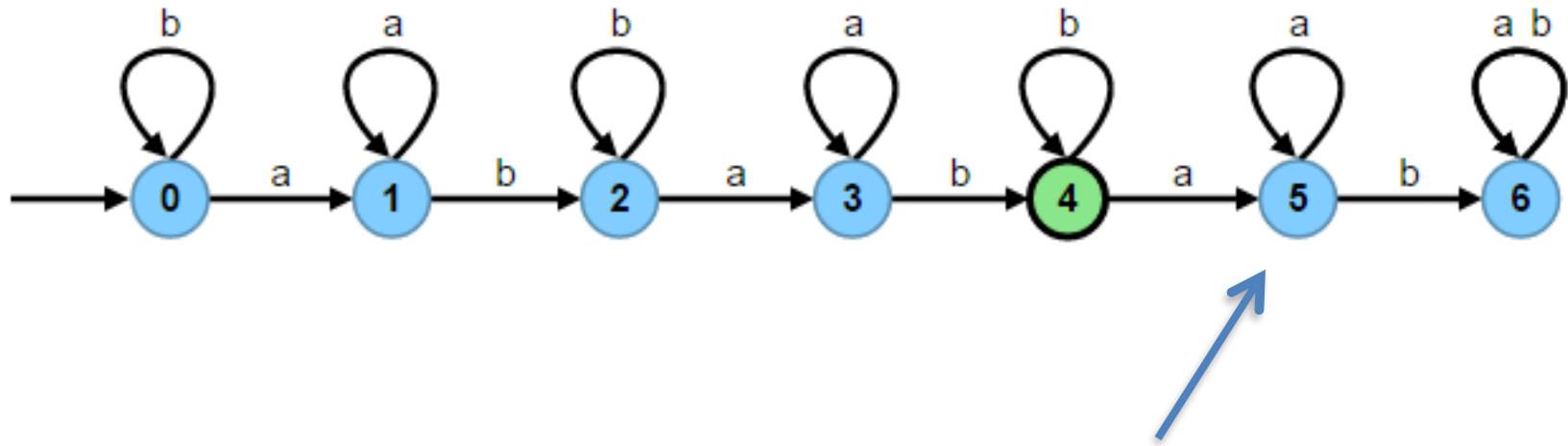
$\{ s \mid \text{'ab' appears in } s \text{ at least 2 times} \}$

INTUITION: find the edit distance between the two language descriptions

Feedback via Synthesis



Solution syntactic mistake



The student
forgot one final
state

INTUITION: find the smallest number of syntactic
modification to fix the DFA

DFA Edit Difference

Compute DFA edit distance:

Number of edits necessary to transform the DFA into a correct one

An edit is

- Make a state (non)final
- Add a new state
- Redirect a transition

HOW DO WE EVALUATE A TEACHING TOOL?

What evaluation tools can we use?

- User studies
 - One for automatic grading
 - One for automatic feedback
- Surveys
 - For grading, feedback, interface, and usability
- Instructors' feedback
 - For features, usability, and suggestions

User study on grading [IJCAI13]

Are the grades computed by the tool representative of human ones?

Setup:

- 250 students in one course (UIUC)
- 6 problems, multiple possible submissions
- ~2500 DFAs submitted
- 2 human graders VS tool

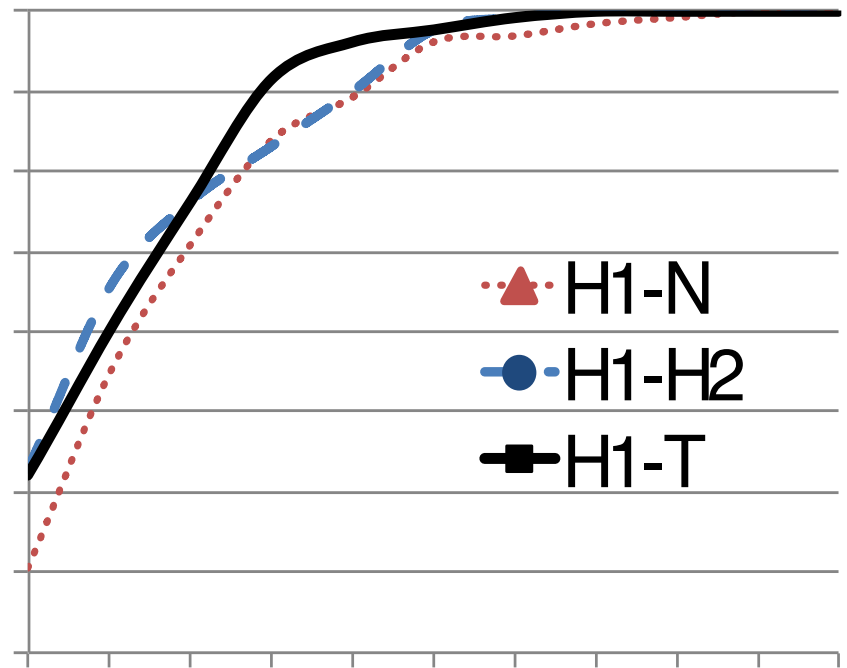
Evaluation of grading

H1, H2 = human graders

N = naïve grader

T = tool

Tool is closer to humans than humans are to each other



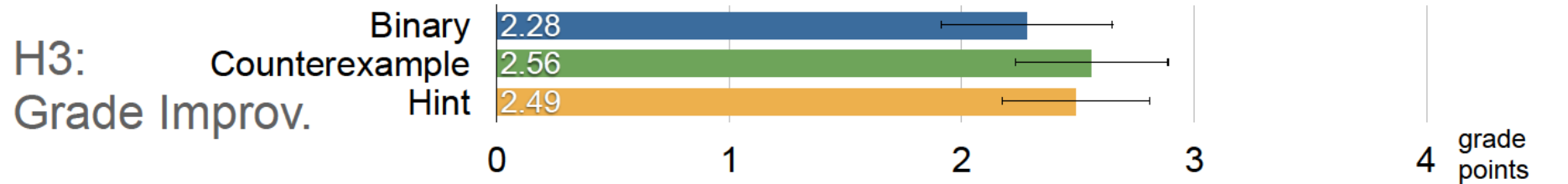
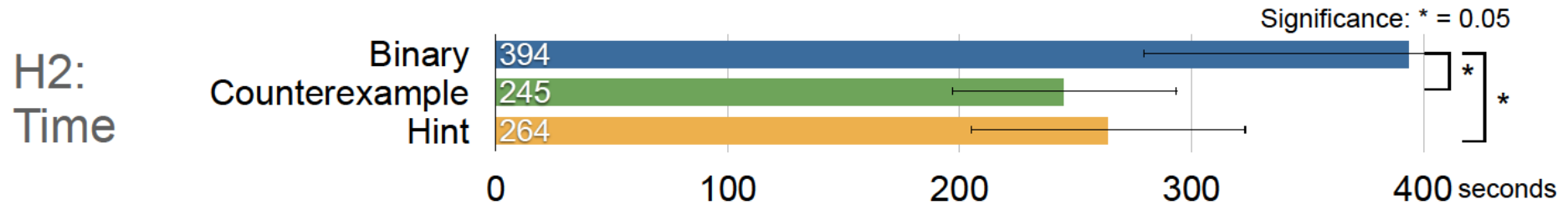
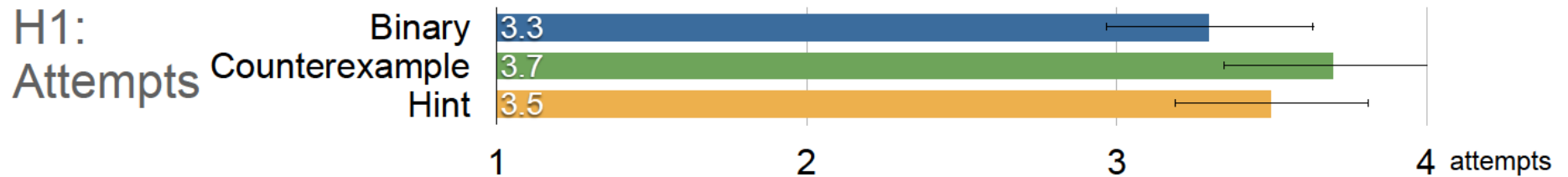
User study on feedback [TOCHI15]

Is the feedback helpful?

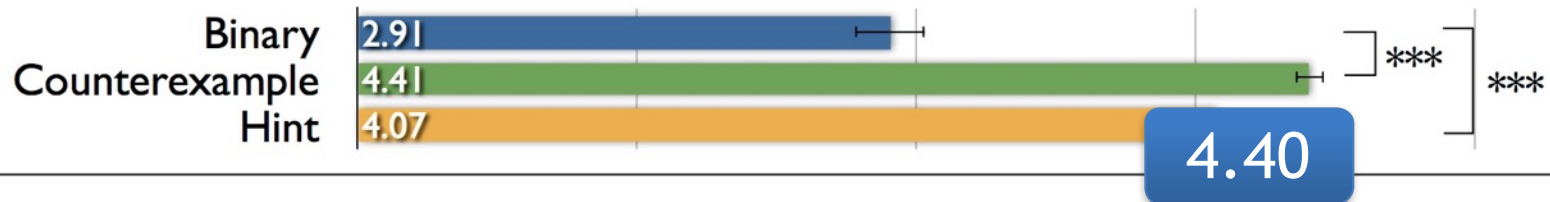
Setup:

- 300 students in 2 courses (Penn, UIUC)
- 4 problems multiple possible submissions
- 3 types of feedback:
 - Binary (yes/no)
 - Counterexample
 - Hints (the ones I described before)

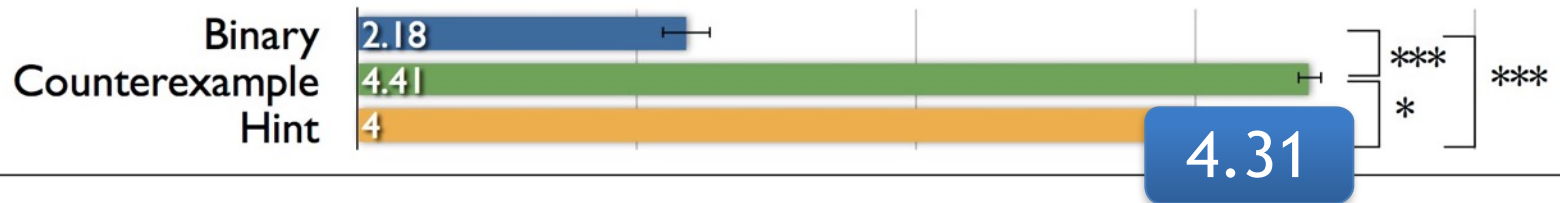
Feedback evaluation



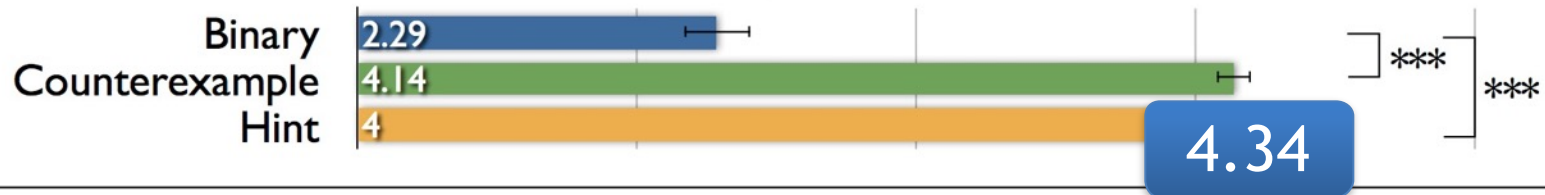
Feedback is useful overall:



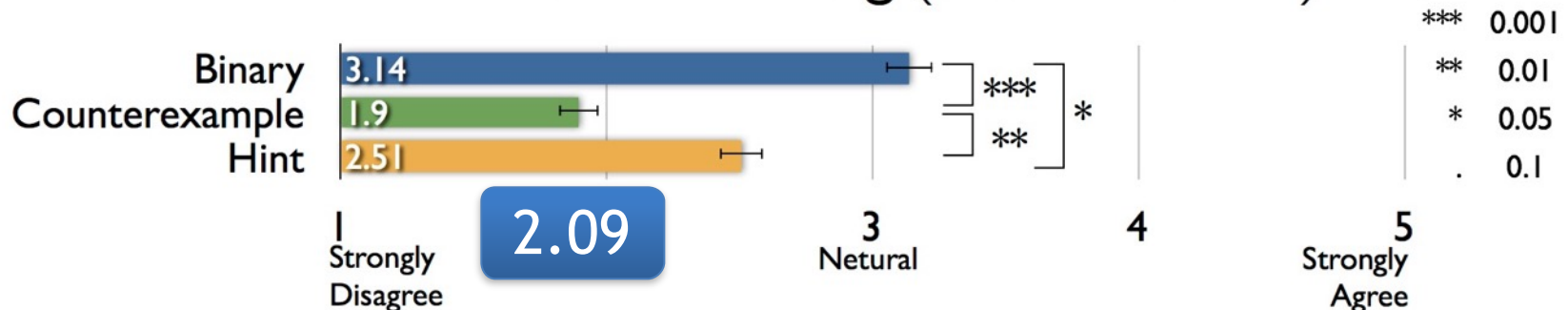
Feedback is helpful for understanding mistakes:



Feedback is helpful for getting to correct DFA:



Feedback is confusing (lower is better):



Is feedback improving students' engagement?

Question:

How often does a student give up on a practice problem based on his type of feedback?

Setup:

4 mandatory homework problems

18 practice problems

Results:

Binary Feedback:

44 % of the time

Counterexample:

27 % of the time

Hint Feedback:

33 % of the time

Survey on interface usability

Is the new drawing interface easier to use?

Setup:

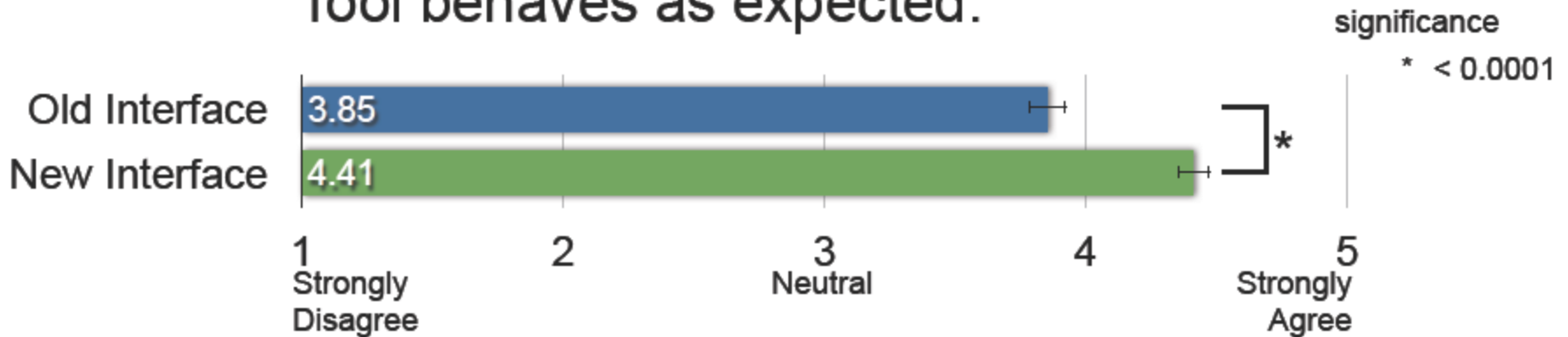
- 2 surveys (old VS new),
- 1 year apart, same course (Reykjavik)

Interface evaluation

Tool is easy to use:



Tool behaves as expected:



Instructors' feedback

- Students seem to be more engaged in the course material
- Average grades are better than in previous offerings of the course

Instructors enjoy the tool

“This is how the construction of finite automata that recognize regular languages should be taught in a modern way! I wish I had similar tools for all the topics I need to cover.”

--- Luca Aceto, Ode to Automata Tutor

WHAT DID WE LEARN?

A brief history of Automata Tutor

Version	Features	Number of users
<i>Version 0. Cerny, Radhakrishna, Zufrey 2012-2013</i>	<i>DFA counterexamples, attempt at non- regularity</i>	<i>~10</i>

What worked

- Good interfaces are important
- Simple feedback is good enough
- Instructors like independence
- Instructors love automated grading
- End-of-course surveys are extremely helpful

What did not work

- Verbose feedback is confusing
 - E.g. Your solution is incorrect on all strings that start with an 'a', end with a 'b', and do not contain a 'c'
- How-to feedback works poorly if the student solutions is too far off
 - E.g. To fix your DFA you need to change the acceptance condition of 3 states and redirect 6 transitions...
- A single crash can cost many users

WHAT'S NEXT?

More features, more users

- Pumping lemma proofs
- Better regular expression feedback
- Meta-constructions
- Minimization
- ...
- Suggestions?

www.automatatutor.com

13 Universities using it

UIUC, Upenn, UC San Diego, Cambridge, Reykjavik University, University of Liverpool, EPFL, Mahidol University International College, University of Louisiana at Monroe, IIT Delhi, John Hopkins, Federal University of Campina Grande, University of Luxembourg

>3000 students

>50000 assignments graded

