AutomataTutor: one year later

Rajeev Alur (Penn), Loris D’Antoni (Penn), Sumit Gulwani (MSR), Bjoern Hartmann (Berkeley), Dileep Kini (UIUC), Mahesh Viswanathan (UIUC)
What is AutomataTutor?

Draw the DFA accepting the language:

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

Solution:
Your DFA accepts the language

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

Grade: 6/10
Feedback via Synthesis

\[ \text{indOf(}ab\text{)} = 2 \]

\[ \text{indOf(}ab\text{)} \geq 2 \]

Replace \( \geq \) with \( = \)

\[ \{ s \mid 'ab' \text{ appears in } s \text{ at least } 2 \text{ times} \} \]
You need to change the acceptance condition of one state

Grade: 9/10
Evaluation Questions

Are the computed grades fair?
- YES [IJCAI13]

Is the computed feedback helpful?
- YES (in some sense) [will submit to TOCHI14]

Is anyone else (beside me) going to use the tool?
- It seems like a YES (Penn, UIUC, Reykjavik)
Time to Complete a Problem After First Feedback

<table>
<thead>
<tr>
<th></th>
<th>Binary</th>
<th>Counterexample</th>
<th>Hint</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>166</td>
<td>105</td>
<td>163</td>
</tr>
<tr>
<td>A2</td>
<td>500</td>
<td>235</td>
<td>257</td>
</tr>
<tr>
<td>A3</td>
<td>537</td>
<td>415</td>
<td>493</td>
</tr>
<tr>
<td>A4</td>
<td>269</td>
<td>77</td>
<td>119</td>
</tr>
</tbody>
</table>

Significance:
- *** 0.001
- ** 0.01
- * 0.05
- . 0.1

0 100 200 300 400 500 seconds
Setup:
   4 mandatory homework problems
   18 practice problems

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

Results:
   Binary Feedback: 44 % of the time
   Counterexample: 27 % of the time
   Hint Feedback: 33 % of the time
Feedback is useful overall:

- Binary: 2.91
- Counterexample: 4.41
- Hint: 4.07

Feedback is helpful for understanding mistakes:

- Binary: 2.18
- Counterexample: 4.41
- Hint: 4

Feedback is helpful for getting to correct DFA:

- Binary: 2.29
- Counterexample: 4.14
- Hint: 4

Feedback is confusing (lower is better):

- Binary: 3.14
- Counterexample: 1.9
- Hint: 2.51

Significance:
- *** 0.001
- ** 0.01
- * 0.05
- . 0.1
Descriptive Feedback Better

**PROBLEM:** strings that contain \texttt{ba} exactly twice

<table>
<thead>
<tr>
<th>Group</th>
<th>Tot. Students</th>
<th>% Giveup</th>
<th>% Full Score</th>
<th>Avg Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct/Incorrect</td>
<td>27</td>
<td>22%</td>
<td>45%</td>
<td>7.77</td>
</tr>
<tr>
<td>Counterexample</td>
<td>22</td>
<td>10%</td>
<td>45%</td>
<td>7.2</td>
</tr>
<tr>
<td>Hint</td>
<td>24</td>
<td>4%</td>
<td>50%</td>
<td>7.82</td>
</tr>
</tbody>
</table>

**COUNTEREXAMPLE FEEDBACK:**
Your DFA accepts the string \texttt{bababa} but the correct answer does not!

**HINT FEEDBACK:**
Your solution accepts the following set of strings:
\{ s \mid \texttt{ba} appears in s at least twice \}
We used the previous experiment results to improve the tool and...

It gave one simple example for why my solution didn't work so I had to think about what was wrong and how I could find the solution right away. This is a far superior way to learn new things rather than read about something, solve random examples and not knowing what is wrong with one's solution immediately. It doesn't matter how many examples I solve if I can't be certain what the right answer is. I hope this way of teaching will be implemented in all schools.
Univ. of Reykjavik test

I thought the feedback was absolutely Excellent.

It helped me solve a couple of exercises, helped me with the extreme/end cases.

It was short, simple and to the point!

I liked its subtle hints.

Sometimes the feedback was confusing.
Conclusions

AutomataTutor.com: a tool that grades DFA constructions fully automatically and provides students with personalized feedback

We will fully deploy it by Fall14.