Improving and evaluating eMMA’s communication skills
A chatbot for managing medication

Gabriel Hess, Guillaume Fricker, Kerstin Denecke

Healthcare of the Future, Biel 05.04.2019
Introducing eMMA

- electronic Medication Management Assistant
- Bachelor Thesis 2017
- eHealth connected medication app
Introducing eMMA

- electronic Medication Management Assistant
- Bachelor Thesis 2017
- eHealth connected medication app

Source: M. Lehmann
What did we do?

- „Improve eMMA’s communication skills“
  - Limited original conversational interface
  - Implement a real chatbot
  - For better understanding of the users utterances
  - With better knowledge of context
  - Longer attention span

- Evaluate the improvements
  - Three step evaluation
Improve eMMA’s communication skills

- RiveScript as Chatbot service
  - Rule-based
  - Runs on device
  - Rules in .rive-Files
- Service class as a black-box
  - Load context information at launch
  - Instruction interface for controlling application
Evaluating eMMA

- Is eMMA 2.0 better than eMMA 1.0?
  - What means better?
  - Literature analysis of 17 publications about medical chatbots

- Three stage evaluation
  - Theoretical
  - Quantitative
  - Qualitative
Theoretical: TRINDI checklist

- 16 questions in 3 categories
  - Flexibility of the dialogue system (9 questions)
  - Overall functionality of the system (5 questions)
  - Context awareness (2 questions)
- Yes / Partially / Theoretically / No / (Unknown)*

<table>
<thead>
<tr>
<th></th>
<th>Flexibility</th>
<th>Overall functionality</th>
<th>Context awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Partially</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Theoretically</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

* according to Beveridge / Fox: Automatic generation of spoken dialogue from medical plans and ontologies, in Journal of Biomedical Informatics 39 (2006)
Quantitative: Analyzing chatlogs with PARADISE*

- Task success
  - Based on use case: dialog with multiple steps
  - e.g. „add new medication“
  - Success / fail
- Efficiency
  - Based on use case: dialog with multiple steps
  - Number of steps for completing use case
- Dialog quality
  - Based on request / answer-pairs
  - Accuracy: does the answer fit the question?
  - Speed: How fast is eMMA answering?
- User satisfaction
  - Questionnaire with 12 questions

## Quantitative: Analyzing chatlogs with PARADISE

<table>
<thead>
<tr>
<th></th>
<th>eMMA 1.0</th>
<th>eMMA 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of tests</strong></td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td><strong>Task success (use case)</strong></td>
<td>3 / 39</td>
<td>53 / 85</td>
</tr>
<tr>
<td></td>
<td>7.7%</td>
<td>62.4%</td>
</tr>
<tr>
<td><strong>Dialog efficiency (use case)</strong></td>
<td>5.3 steps</td>
<td>9.7 steps</td>
</tr>
<tr>
<td></td>
<td>successful: 5.8 steps</td>
<td>successful: 9.7 steps</td>
</tr>
<tr>
<td><strong>Dialog quality (accuracy) (utterance)</strong></td>
<td>10 / 59</td>
<td>213 / 359</td>
</tr>
<tr>
<td></td>
<td>16.9%</td>
<td>59.3%</td>
</tr>
<tr>
<td><strong>Dialog quality (speed) (utterance)</strong></td>
<td>20 ms</td>
<td>20 - 50 ms</td>
</tr>
<tr>
<td><strong>User satisfaction</strong></td>
<td>2.6 pt</td>
<td>-3.25 pt</td>
</tr>
<tr>
<td></td>
<td>55.4%</td>
<td>43.2%</td>
</tr>
</tbody>
</table>
Qualitative: User tests with eMMA 2.0

- Focussed on usability, no clinical outcome - thus no „real patients“

- Insights
  - Users are able to interact with eMMA, even with complex tasks
  - Some users are more efficient using GUI
  - Users expect deeper knowledge
  - Higher expectations with eMMA 2.0 than eMMA 1.0?
  - Software is still in prototype stage, with plenty of bugs
Conclusion

- Big leap compared to eMMA 1.0
- Not yet ready for the real world
- RiveScript as a technical foundation set for further improvement
Outlook

- Better context management (improve RiveScript scripts)
- Bigger knowledge base
- eMMA needs to „know“ if she understood the users utterance
- Include more graphical UI elements directly into conversational UI
Outlook

- Better context management (improve RiveScript scripts)
- Bigger knowledge base
- eMMA needs to „know“ if she understood the users utterance
- Include more graphical UI elements directly into conversational UI
Thank you for your attention!

- And special thanks to:
  - Guillaume Fricker, for doing the projects with me
  - Prof. Kerstin Denecke, for advising us
  - Tim Dorner and Mauro Tschanz, for inventing eMMA*
  - the people testing eMMA, for their help

- Any questions?

- hessg1@bhf.ch