OPPONENT RECORD

Thesis compiled by Max Malmgren & Ulf Åhammar

Title of thesis: Chatterbot with common sense

Opponent: Kristoffer Hallqvist

Was it easy to understand the underlying purpose of the project? Comments.

Yes, to a large extent. I find it clear what ELIZA lacks and what the authors are trying to fix with this project.

Do you consider that the report title justly reflects the contents of the report?

Fairly well. One issue, though, is that the title might suggest that the authors have had access to a chatterbot that actually has real common sense (as humans define it), but judging by the results, that is arguably not the case. Maybe the title should be more clear about the fact that the authors are *investigating the possibilities* of having a chatterbot being perceived as having common sense, rather than having an existing intelligent chatterbot as the starting point.

How did the author describe the project background? Was there an introduction and general survey of this area?

The reader is given an introduction to ELIZA which is easy to understand.

It's very good that the reader is given an explanation of the expression “common sense”, as it does not mean exactly the same thing as what humans usually mean by it, but rather a simulation of it.

The external source about *semantic networks* is good, but maybe it would be preferable with a slightly more detailed explanation in the report (maybe an own subchapter, and preferably a picture), because it is such a central concept, and not all readers will bother following the links immediately.

To what degree did the author justify his/her choice of method of tackling the problem?

I can't find any explanations for why semantic networks were used instead of other methods, but I assume that the authors want to focus on evaluating that particular method, and not *just* improve the chatterbot in general regardless of how. The possibilities a semantic networks opens up are mentioned clearly, but there is no information about alternative methods and how they compare.

Did the author discuss the extent to which the prerequisites for the application of such a method are fulfilled?

The report mentions the effort required to build a semantic network, but the authors are using an existing one for this project, which should mean there are no other non-obvious requirements worth mentioning. They also have access to a pre-existing ELIZA-like implementation.

Is the method adequately described?

It's easy to get an overview, but the explanation of the specifics could be more detailed.

Has the author set out his/her results clearly and concisely?

The human-bot conversations are easy to read and understand. Due to their limited size, it would have been nice to see them in the report itself and not just in an appendix, to save the reader the trouble of having to jump around in the text.

The ConceptNet search results will surely be confusing to many readers, but I assume that they are represented “as is”, and it's generally a good thing to include such a representation. A complementing description of the results would have been appropriate, though.

Do you consider the author’s conclusions to be credible?

Yes. It seems logical in general that computers have a harder time classifying feelings than more well-defined things, such as physical objects. Also, the chatterbot didn't make sense most of the time, but when it did, it managed to do so in a way that was significantly different from that of the original ELIZA, which hints at possibilities of extending the project further.

What is your opinion of the bibliography? What types of literature are included? Do you feel they are relevant?

There are references to sources about ELIZA, semantic networks, ConceptNet and Python NLTK, all of which are highly relevant to this project.

Which sections of the report were difficult to understand?

The semantic networks are not easy to quickly understand, including their inner workings, the way they are queried, and for example the “start” and “end” concepts.

Other comments on the report and its structure.

The authors correctly identify the fact that whether a response makes sense or not is to a certain extent subject to the person involved. They claim that some of the responses were insightful, and I am curious about which responses the authors found the most interesting, and why.

In chapter 4, the authors alternate between describing pre-existing concepts, and own made implementations, which harms the structure.

The report contains some grammatical errors.

What are the stronger features of the work/report?

It's very easy to understand the background and the problem the authors are working with.

What are the weaker features of the work/report?

It's difficult to quickly get a good understanding of the inner mechanics of the semantic networks and the way they are queried.

What is your estimation of the news value of the work?

Interesting, but not revolutionary by any means. I could probably have made an educated guess that would turn out to be close to the conclusions of this report, but considering the scope and time constraints, one cannot really demand much more.

Summarize the work in a few lines.

It's an interesting report that successfully explains the context and the aim of the project, but lacks a bit of detail in the procedure part. The conclusions are not overwhelmingly interesting, which is to be expected considering the time frame, but a future extension could very well be of great interest.

Questions to author:

1. How is a query executed in the semantic network?

2. Which specific conversation responses from the results contribute the most to the understanding of how successful the method is, and why?

3. Are there any other methods (other than semantic networks) that can be used to achieve similar effects, and how do they compare to semantic networks?

4. Is ELIZA designed specifically for a psychologist-patient conversation environment, or is it a general purpose program that just happens to be used that way?

5. Regarding the bad quality of many replies, how much do you think is down to the semantic network?

6. What are the differences between the original ELIZA implementation and the one used in this project? Are they relevant?