Feasibility Study

ELIZA Plus - Long-Term Memory

by

Pritam Sangani

Supervised by: Dr. James O'Shea

School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University

A Feasibility Study document submitted to
Manchester Metropolitan University, as part of a final
year dissertation, in accordance with the requirements
of the degree of Bachelor of Science in the
Faculty of Science and Engineering.

October 2018

Course-Specific Learning Outcomes

Below are the Course-Specific Learning Outcomes for BSc (Hons) in Computer Science. Please note that only those outcomes which are relevant to this dissertation have been listed.

On completion of the course students will be able to:

- use knowledge, abilities and skills for further study and for a range of employment in areas related to scientific and technical computing;
- interpret legislation appropriate to computer professionals and also be aware of relevant ethical issues and the role of professional bodies;
- analyse, design, and implement algorithms using a range of appropriate languages and/or methodologies;
- demonstrate effective communication, decision making and creative problem solving skills, and identify appropriate practices within a professional, legal and ethical framework;
- critically appraise and apply suitable artificial intelligence techniques for a variety of software systems.

Background

In 1950, Alan Turing published a paper called 'Computing Machinery and Intelligence' (Turing 1950), in which he posed the question, "Can machines think?". Alongside this question, he also proposed the Turing Test - a test that would use conversation to answer this question. Since then, Computer Scientists have been researching, for decades, how to develop chatterbots that converse convincingly with humans. Research, in this field, intensified when in the mid-1960s, Joseph Weizenbaum, developed ELIZA (Weizenbaum 1966) - known to be the first chatter bot. ELIZA used an early-form of Natural Language Processing, where it would match patterns in the text input and substitute it with phrases, to create the illusion of understanding. In recent years, chatterbot developers have been trying to win the Loebner Prize - a modern day version of the Turing Test. This contest has been held since 1991 in which judges converse with chatterbots not knowing whether they are talking to a human or a chatterbot (Zdenek 2001). Chatterbots have also become very popular commercially, as businesses look to make their customer services more efficient, innovative and, most importantly, more personal, due to the rising demand of customers looking for fast resolutions to their problems and the increase in time they are spending online (Bakhshi et al. 2018). This report by Deloitte, also found that one of the market forces driving chatterbot development is the "technological advances in AI and NLP".

Figure 1 below shows an infographic, also from the Deloitte report, displaying the "different functions of the human brain" that chatterbots try to "mimic". For my thesis, I will be concentrating on the Natural Language Processing (NLP) and Entity Recognition parts of the infographic for Dialog Management. NLP is a branch of AI concerned with the research of interactions between humans and computers through natural language. Natural Language has been defined in a white paper on NLP as being "the most natural means of communication between humans, and the mode of expression of choice for most of the documents they produce" (Weischedel et al. 1989). Entity Recognition (also known as Named Entity Recognition[NER]) is a branch of NLP concerned with labelling "sequences of words in a text which are the names of things, such as person and company names, or gene and protein names." (*Stanford Named Entity Recognizer (NER)* n.d.)

A big part of developing chatterbots is actually evaluating the quality of existing chatterbots and

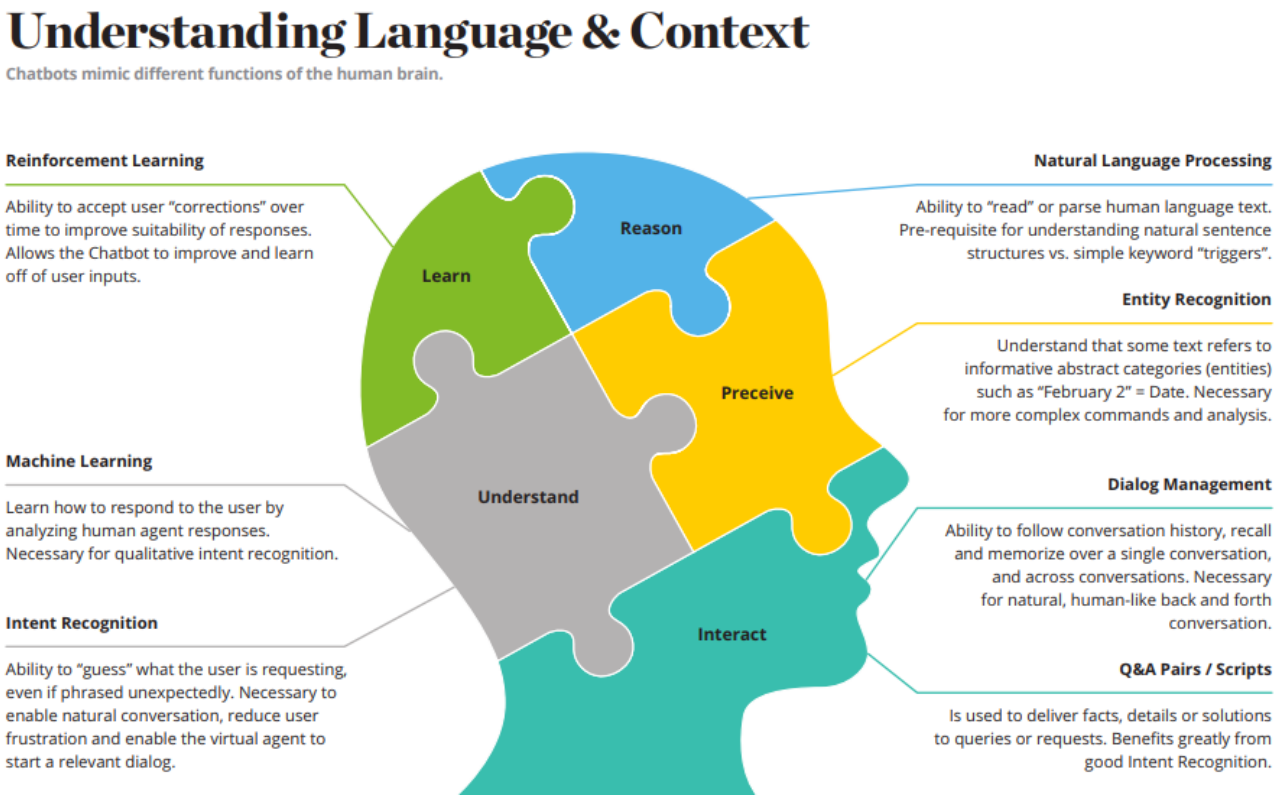


Figure 1: Infographic showing the different functions of a chatterbot architecture (Bakhshi et al. 2018)

how convincing they are at mimicking the functions of the human brain. Two recent studies have tried to answer two questions that must be answered when evaluating the quality of a chatterbot. Firstly, the current uses of chatterbots in society must be examined. Brandtzaeg and Følstad have published their study of the uses of chatterbots on various platforms and across various categories of uses. (Brandtzaeg & Følstad 2017) The vast majority of participants in the study reported using chatterbots to increase their productivity, by quickly retrieving information or accessing assistance. Rather surprisingly, the study also found that 12% of the participants reported using chatterbots for social or relational use. The study found that the human nature of chatterbots drove people to using them for this purpose - the responses stated that chatterbots were a way of "avoiding loneliness" and "improving their social and conversational skills". Another study tried to gather together all the attributes and features that can be used to assess the quality of chatbots. (Radziwill & Benton 2017) The study found that two attributes that measures the effectiveness of the chatterbot are the ability to "maintain themed discussion" and to deliver "convincing, satisfying and natural interaction". The study also found that one of the attributes that measures the satisfaction of the chatterbot is the ability to "detect meaning or intent".

A big part of this project is to add a "long-term memory" mechanism to the core of the ELIZA chatterbot. Long-term memory is the ability to refer back to earlier conversations and to bring the information back at relevant points in the current conversation. This would make the chatterbot more human-like and, as discussed, would improve the quality of the chatterbot. I plan to achieve this using NER and NLP to store relevant and linked information into a database to retrieve at a later point in the conversation.

Research Question

The question I propose to answer with my research is as follows:

Is it possible to implement convincing long-term memory into an existing chatterbot, such as Weizenbaum's ELIZA?

Hypotheses

From the research question above, I propose that the following hypotheses could be derived from my research:

- **H_0 - it is not possible to add long-term memory to a chatterbot, at this stage, which is convincing enough to be comparable to human memory mechanism.**
- **H_1 - it is possible to add convincing long-term memory at a significant statistical level, if the number of participants in the research is high enough.**

Note: The degree of convincingness would be determined by the analysis of conversation log files, to see how often the memory mechanism appeared to access what the user was referring to, and by a subjective analysis of a questionnaire put to the participants of a research experiment.

Aim

The aims of this thesis are as follows:

- **Integrate Long-Term Memory mechanism to the ELIZA chatterbot core.**
- **Evaluate how convincing the long-term memory mechanism is, by way of research methods determined by research done in literature review.**

Objectives

The problem is a complex one to be solved and so I have broken the problem down into small steps and approaches that I will take to solve the problem. These are as follows:

1. Complete a Literature Review by reading research papers and articles on the topic of Natural Language Processing and Named Entity Recognition from the Internet and textbooks.
2. Find source code for the ELIZA chatterbot in Python and start understanding how the code works.
3. Compare the different tools for Named Entity Recognition readily available for use in Python.
4. Design an overview of the software and a plan to implement.
5. Implement the software according to the plan, whilst incrementally testing where appropriate.
6. Carry out a final testing of the software.
7. Evaluate the quality of the software by carrying out a study with participants.
8. Write up the research findings in a report.

As well as the objectives listed above there are a number of interim deliverables that need to be met. These are as follows:

1. Prototype Report
2. Prototype Software
3. Final Software
4. Report Outline
5. Showcase Event
6. Final Report

Problems

As with any project, problems can arise and it is important to resolve them quickly. To mitigate the effect of problems during this project, I am considering the problems that could arise and how I would overcome each problem.

Firstly, a problem that is likely to occur is in the implementation of the product. As I have not worked with Natural Language Processing before, I am likely going to have some trouble, initially, implementing the Named Entity Recognition algorithms. To resolve any issues in implementation, I will first look at the documentation for any modules I use. If this does not solve the problem, I can use sites, such as Stack Overflow, to search for possible solutions.

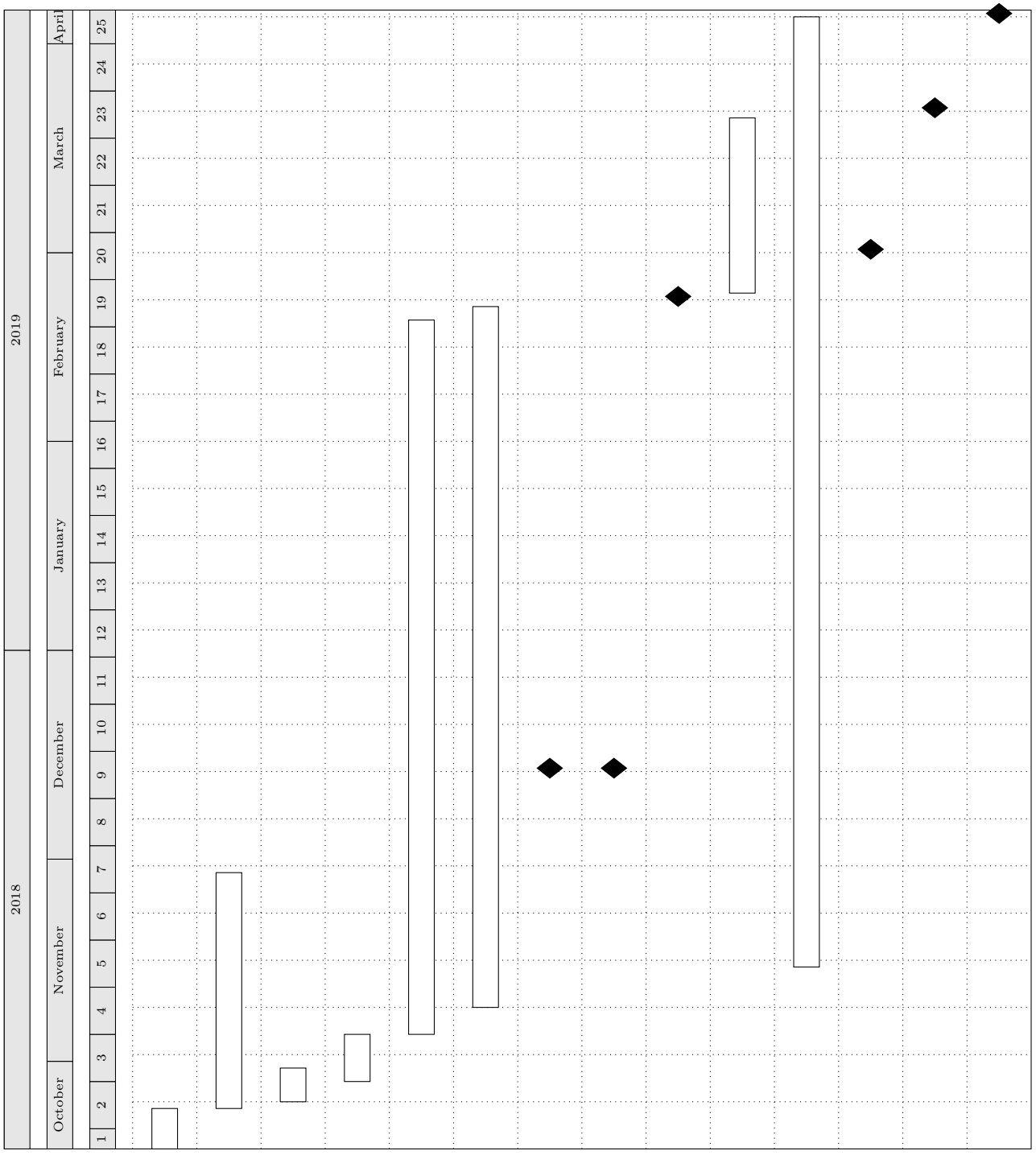
A second problem, that is not likely to occur but could still occur, is that my code could be lost, either by my laptop becoming corrupt or stop working. To mitigate this problem, I will use a version control system called Git and I will host my code on GitHub, which is a cloud-based storage service for code repositories.

Evaluation

As with all projects of this nature, an evaluation of the produced software must be undertaken. I intend to do this by running an experiment where participants would converse with the chatterbot for two sessions. There will be two sessions as this project involves implementing long-term memory. Therefore, in order for the participant to properly evaluate the quality of the chatterbot, in particular its ability to retrieve information from previous conversations, I would need the participant to have two sessions with the chatterbot. Once the participant has concluded the experiment, I would have them answer a questionnaire to gauge their responses and feelings about the chatterbot's quality. I would then subjectively analyse the responses, as well as, look at the conversation logs to determine the degree of convincingness of the chatterbot.

Schedule

Below is the proposed week-by-week schedule for the project as outlined by the tasks/objectives outlined in the Objectives section. I have displayed the objectives as bars, whilst the deliverables are classed as milestones, which are displayed using diamond symbols.


2018
2019
October
November
December
January
February
March
April
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Literature Review
Understand Source Code
Compare available NER tools
Design Software Overview & Plan
Implementation of Software
Testing of Software
Prototype Report
Prototype Software
Final Software
Evaluation of Software
Report Write-Up
Report Outline
Showcase Event
Final Report

Required Resources

The resources required to carry out this research project is as follows:

- A laptop or PC capable of smoothly running:
  - Python 3.x
  - Google Chrome
  - PyCharm - an IDE, by JetBrains, to write and compile Python code
  - Visual Studio Code - an IDE to write code in any programming language (HTML, CSS and JavaScript for this project)

All resources listed above are all available to download for free. (Note: PyCharm has a paid version, but, for the scope of this project, the free community edition will suffice.) Furthermore, I will not be relying on the University Laboratory PCs as I can run all required resources on my personal laptop.

References

Bakhshi, N., van den Berg, H., Broersen, S., de Vries, D., El Bouazzaoui, H. & Michels, B. (2018), Chatbots point of view, Technical report, Deloitte The Netherlands.

Brandtzaeg, P. B. & Følstad, A. (2017), Why people use chatbots, *in* 'International Conference on Internet Science', Springer, pp. 377–392.

Radziwill, N. M. & Benton, M. C. (2017), 'Evaluating quality of chatbots and intelligent conversational agents', *arXiv preprint arXiv:1704.04579* .

Stanford Named Entity Recognizer (NER) (n.d.), Available at `https://nlp.stanford.edu/software/CRF-NER.shtml` (Accessed: 17/10/2018).

Turing, A. M. (1950), 'I.—computing machinery and intelligence', *Mind* **LIX**(236), 433–460.

Weischedel, R., Carbonell, J., Grosz, B., Lehnert, W., Marcus, M., Perrault, R. & Wilensky, R. (1989), White paper on natural language processing, *in* 'Proceedings of the Workshop on Speech and Natural Language', HLT '89, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 481–493.

Weizenbaum, J. (1966), 'Eliza - a computer program for the study of natural language communication between man and machine', *Commun. ACM* **9**(1), 36–45.

Zdenek, S. (2001), 'Passing loebner's turing test: A case of conflicting discourse functions', *Minds and Machines* **11**(1), 53–76.