

Project specification for 'Chatbot using common sense knowledge'

by Gustaf Lindstedt & Martin Runelöv

Introduction

The aim of this project is to develop a chatbot which utilizes a common sense database when interacting with a user. A common sense database includes such statements as "Something you find in the ocean is a fish". This could possibly enable more complex word associations than previously achieved, and could prove to be a good start for a chatbot which utilizes machine-learning.

Rationale:

Computers lack common sense. It would be interesting to see if a computer could mimic the association patterns humans have using common sense knowledge.

Scope:

We will try to build a chatbot that can incorporate a common sense database to provide simple but relevant (non-gibberish) responses while chatting. We need to successfully parse human input and find the most relevant keywords with some understanding of the context.

A very basic response would be simply stating a fact that is related, for example:

Human: *"I live by the ocean".*

Chatbot: *"Fish live in the ocean".*

However, we hope to achieve a more dynamic response:

Human: *"I live by the ocean".*

Chatbot: *"Do you fish?"*

Success criteria:

The chatbot can use the database to generate responses that are relevant and (relatively) grammatically correct.

Risks:

The parsing of the human input need to be very good for this to work well. The chatbot needs to understand what keywords are relevant. It will also be hard to use the database to generate responses that seem natural.

Constraints

We have a very limited amount of time.

We have no experience with natural language processing.

Assumptions

We are assuming that there is a well defined api for interfacing with the common sense database, which if not true would severely disturb our time-plan. We are also assuming that the NLTK package is easy to use for our purposes.

Quality

This is a research project which will examine the possibilities of increasing the quality of chatbot-responses. We are therefore striving for as much quality as possible in regards to the chatbot's responses. We will mainly measure this quality by subjective assessments of the responses the chatbot gives. These assessments will be provided by us and other test-subjects.

Problem Statement

Develop a method of integrating a common sense database with a chatbot such that it would be able to form relatively coherent messages based on words associated with the user's input.

Approach

We will start by producing a skeleton which accepts input and processes the input using Natural Language Processing (using the NLTK python package). Then we will start querying a common sense database with different input, observing the results and refining a method for obtaining the relevant data. Then when we have a sufficiently developed model for common sense queries, we can begin to construct a response model for the chatbot, utilizing the common sense database.

References

NLTK:

<http://nltk.org/>

Common sense database:

<http://openmind.media.mit.edu/> (which is a part of: <http://conceptnet5.media.mit.edu/>)

Tutorial on making an Artificial Intelligence Chatbot

<http://www.codeproject.com/Articles/36106/Chatbot-Tutorial#context>

Time Plan

Things to have ready for the half-way meeting:

- Halfway draft of the essay.
- We have parsing of the input using NLTK that can select keywords.
- We can query the database with the parsed data and print relevant (but not coherent) results.

1-2 Weeks before deadline:

- Stop developing chatbot.
- Start compiling and analyzing the results and finalizing the essay.