Conceptual Question Answering

Project Outline for Ning, Ella, Jamie, and possibly others

7-7-2020

The main goal of this project is to get a very crude question answering system working in Conceptual Dependency using available software and tools. These notes propose the steps for getting that to work.

Let's refer to the story "Jack went to the store. Jack got a kite. Jack went home" as STORY1. Ella is currently working on the MicroELI conceptual analyzer (AKA ELI) which parses English into CD. ELI currently has a tiny lexicon, but it can parse the sentences of STORY1.

The steps to follow:

1. Parsing questions: The first task in this project will be to get ELI to parse questions relevant to STORY1, like "who went to the store?", "who got a kite?", "who went home?". These first three are easy b/c they have the same syntax as the sentences in STORY1. You'll just need to add "defword" forms for the question words. I recommend that you set things up so that they get analyzed into CD structures in which the symbol "*?*" is used to represent the unknown/wanted information in the question. For example "who went to the store" => (ptrans (actor (*?*)) (object (*?*)) (to (store))) or something like that.

2. Answering a question: Now, write a function that uses ELI to parse both STORY1 and one of the questions from earlier. Then it performs a match of the CD structure for the question against the three CD structures of the story, and returns the CD structure that is the strongest match. You should be able to do this with the matching functions that are in cd-functions.lisp, which is part of the MicroELI code. (I think Ella is pretty familiar with cd-functions.lisp.)

   If you want some examples of how these matching functions can be used, check out micro-sam.lisp, a micro version of the SAM script applier mechanism (The micro-sam.lisp code is linked below, and so is *Inside Computer Understanding* which describes MicroSAM).

3. Back to parsing questions: Then there are other questions that you'd like to be able to parse with ELI, like "where did jack go", "what did jack get", "what did jack do" into CD. These are a bit harder because the syntax of these English questions differs from the statements. In the first pass, it's OK if you rewrite the ELI lexicon completely so that it

can parse just these questions. Or you can change the ELI parse function so that it has a "mode" flag to tell it whether a question or a statement is coming. (A future challenge will be to get ELI to parse both questions and statements without knowing which one is coming. I guess we can worry about that later.)

4. More complex stories: Try answering the questions in the context of a more sophisticated story which ELI should be able to parse with a few more additions to the ELI lexicon: STORY2 = "Jack went to the store. Bob went to the beach. Jack got a kite. Bob got a seashell..." So now it makes more sense to ask "who" questions b/c the answer isn't always "Jack." And the answers to the "where" and "what" questions depend on which actor is specified.

5. Generating natural language answers (possibly pretty hard): use the BABEL "conceptual generator" to generate the question answers from the previous part. The main function that BABEL uses to generate from a CD structure is EXPRESS. Evaluate/run the PROGN at the top of the file tests.lisp in the BABEL zip folder to see how to load the needed code files and how to call EXPRESS.

   One of the main hurdles you'll need to get over is that the way CD structures are represented in the Lisp code for BABEL is different from the way they are represented in MicroELI. Ella has been working on some code to translate from ELI-generated CD structures to BABEL-friendly CD structures, so you should be able to run Ella's translator and call EXPRESS with its output.

   So now you should be able to modify your question-answering function from part 2 so that it takes a natural language story and a natural language question, and it outputs a natural language answer to the question. EXPRESS may print out multiple paraphrases of the answer, but that's OK for now.

Optional Extensions (More Open-Ended)

6. ProPara: try to add to the ELI and BABEL lexicons so that they can parse some ProPara sentences and paragraphs. Stick to ProPara sentences and paragraphs that we have determined to be “purely physical”. Maybe start with the ones that we have already analyzed by hand. If needed, try adding a primitive like PBUILD to the systems (for BABEL, this will be a major extension, but worth it).

7. Paraphrase generation: Now that you have a path from MicroELI to BABEL, you should also be able to compose a function that takes natural language as input and generates natural language paraphrases as output. All this function has to do is run ELI on the

input, run Ella's ELI-to-BABEL translator on the CD structure from ELI, and then pass that CD structure on to BABEL's EXPRESS for paraphrase generation.

8. Scripts (This one is optional and very open-ended): Try to process a story using the micro version of the script applier mechanism (MicroSAM) after parsing it into CD, and then do question answering on the database that MicroSAM creates. Scripts have knowledge about stereotypical social situations that is often not stated in stories, but is often needed to answer questions. The micro-sam.lisp code is linked below.

Resources

Micro ELI code:
https://drive.google.com/file/d/1vBuQzQP5JbfYaSRlcxxWA4PR1DSLdd8z/view?usp=sharing

BABEL code:
https://drive.google.com/file/d/1SsrQB7jYUwWGGZ4au6CZUvdGJ5yKcDa-/view?usp=sharing

Micro SAM code:
https://drive.google.com/file/d/1wtkHXP3Bq-G5YsoTOy9SHMqSyL3r46z9/view?usp=sharing

Inside Computer Understanding:
https://drive.google.com/file/d/1Wy2ts-KvMpv2c20sFcq__3S9bVidWqae/view?usp=sharing

Wendy Lehnert's book on question answering and CD (for inspiration)
https://drive.google.com/file/d/1-nDSlbbYK7DkipbFKrhVWvGAY2NSe_L6/view?usp=sharing

Lisp resources
https://drive.google.com/file/d/1MMDkF8WlhtBUCAnxc_ddDcjWwAZky3QY/view?usp=sharing
https://drive.google.com/file/d/1MOOreUMBgOnYIbkwYQWHy7NjC3UbN68A/view?usp=sharing
http://www.gigamonkeys.com/book/
https://drive.google.com/file/d/1S2T0UKyJd1CRNZhCb-_SzKsd_93ivy4E/view?usp=sharing
https://www.cs.cmu.edu/Groups/AI/html/cltl/cltl2.html