UMBC

A Review of *Common Sense Reasoning – From Cyc to Intelligent Assistant*

Reece Robertson

Outline of Talk

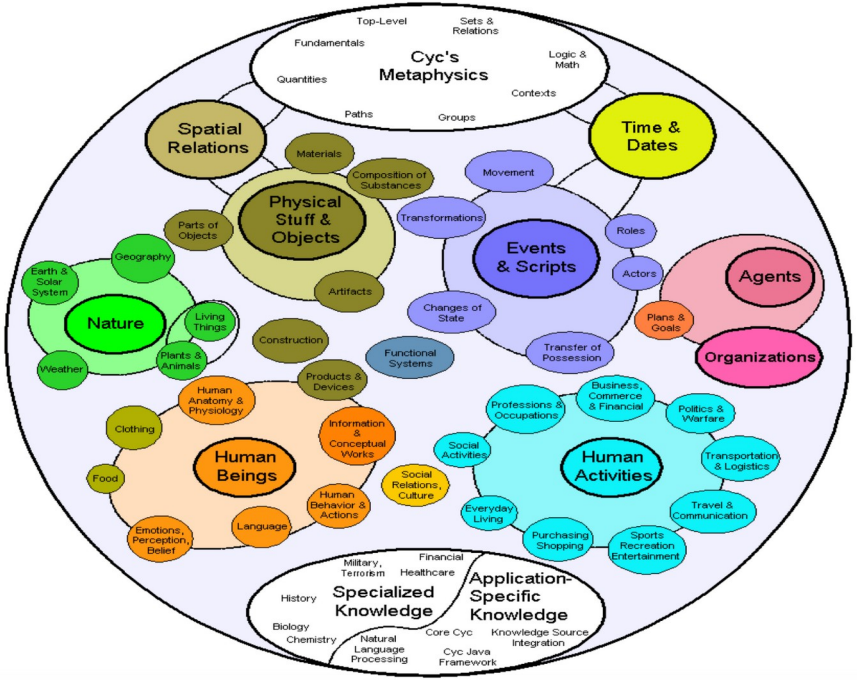
- Summary of Paper
  - General Overview
  - Section 1: The Evolution of Cyc
  - Section 2: The Case for an Ambient Research Assistant
  - Section 3: Natural Language Processing in Cyc
  - Section 4: Automated Learning Within Cyc
  - Section 5: Anticipating Information Needs within Cyc (Developing an Ambient Cyc Application)
- Strengths of the Paper
- Weaknesses of the Paper
- Relation to IF/Story Generation

Summary

- Cycorp Inc.
  - Authors: Kathy Panton, **Cynthia Matuszek**, Douglas Lenat, Dave Schneider, Michael Witbrock, Nick Siegel, and Blake Shepard
- Objectives
  - Research: Create an intelligent software research assistant (called Cyc)
    - Use a general-purpose knowledge base
    - Reduce the time and cognitive overhead for administrative tasks related to research
  - Paper: To present progress made on the Cyc project throughout its ~20 year development history (1984-2006)

The Evolution of Cyc

- Cyc began with the realization that software expert systems would benefit from common-sense reasoning
- Three immediate questions about representing knowledge for common-sense reasoning
  - How much knowledge should be stored?
  - What kind of knowledge should be stored?
  - How should that knowledge be represented formally?
- This lead to the distinction between knowledge and data
  - Knowledge: “underlying heuristics that allow us to reason”
  - Data: “facts or statements about specific items in the world”
- Development of CycL: Cyc’s knowledge representation language



The Case for an Ambient Research Assistant

- The role of an assistant
  - Decide what facts (data) to learn
  - Learn those facts
  - Learn rules (knowledge) related to the facts
  - Generalize the rules
  - Test the rules and revise
- The limitations of human assistants
  - Not always available
  - Can be inflexible in thinking
  - Cannot always anticipate needs
- Components on an intelligent assistant: an intelligent computational assistant can
  - Process natural language
  - Learn both data and knowledge
  - Anticipate needs of the researcher

Natural Language Processing in Cyc

- Components
  - Lexicon containing entries with syntactic properties and semantic links
  - Generator of natural language paraphrases of lexicon entries
  - Parser for reading natural language sentences into lexicographical entries
- Question answering via natural language
  - Parse the question to determine entities
  - Lookup entities in the lexicon–determine relations/facts
  - Return generated natural language

Lexical Information for "tree"

- CycL: `(#$denotation #$Tree-TheWord #$CountNoun 1 #$Tree-ThePlant)`
  Meaning: `#$Tree-TheWord` is a count noun denoting `#$Tree-ThePlant`
- CycL: `(#$singular #$Tree-TheWord "tree")`
  Meaning: One singular form of `#$Tree-TheWord` is the string "tree".

Generation for the predicate `#$hasDiet`

- CycL Template:

```
(genTemplate hasDiet
    (PhraseFormFn NLSentence
        (ConcatenatePhrasesFn
        (BestDetNbarFn-Indefinite (TermParaphraseFn :ARG1))
        (BestNLWordFormOfLexemeFn-Constrained Adverb
TypicalTheWord)
        (BestHeadVerbForInitialSubjectFn Eat-TheWord)
        (BestDetNbarFn-Indefinite (TermParaphraseFn :ARG2)))))
```

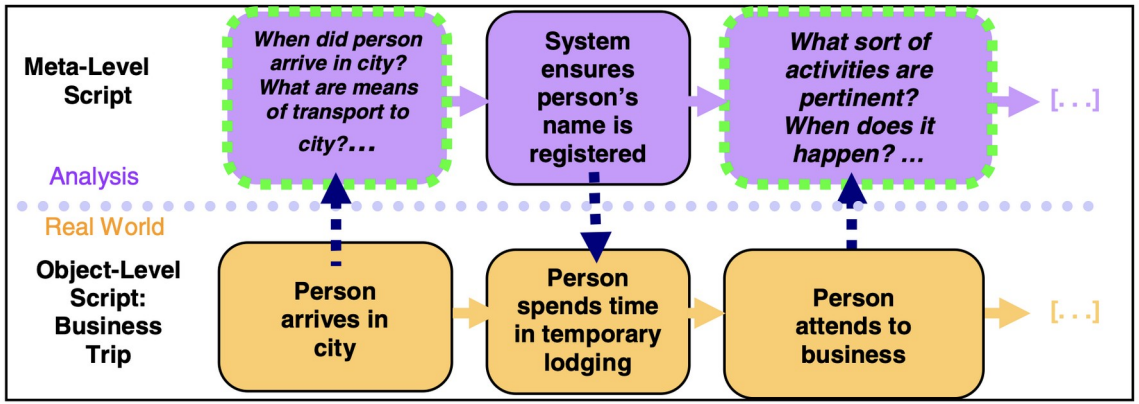
- CycL: `(#$hasDiet #$Termite #$Wood)`
- Generated Text: "Termites typically eat wood."

Automated Learning Within Cyc

- Learn data through gathering facts
  - Via user interaction
  - Via web search
  - Via inference: abduction
- Learn knowledge through rule induction on gathered facts
  - Knowledge (common-sense) often cannot be found in written sources
  - It was originally entered manually
  - With Cyc it can be deduced from gathered facts
    - If all the fish in the dataset live in water, then it stands to reason that all fish live in water

Anticipating Information Needs within Cyc

- Components of the system for anticipating informational needs
  - Object-driven analysis process
  - Script learning
  - Detecting bias
- Hypothesis generation and tracking



Strengths of the Paper

- The paper proposes a system to create a great research assistant
  - It anticipates and fulfills needs more quickly and less intrusive than a human assistant
- The paper outlines the requirements for such an assistant
- The researchers made great strides in knowledge representation and learning while developing Cyc

Weaknesses of the Paper

- Cyc does not fully implement the entire plan laid out by the researchers
- Cyc's learning and common-sense reasoning is still limited

Relation to IF/Story Generation

- Cyc can produce natural language stories in response to queries
  - The lexicon, generator, and parser allows users to ask for and interact with generated stories
  - The automated learning framework allows the system to pull from current data
  - The knowledge inference process produces stories that are consistent with our experience
  - The anticipation of needs may allow the system to produce stories I would like before I ask for them

Questions