Representing Human Knowledge For Common Types

Project Pitch – CS294S Fall 2020
Background & Problem Statement

- To understand sentences, virtual assistants must map them to a formal representation with defined semantics
  - In Lecture 2, we’ve seen ThingTalk for Question Answering
- Natural language concept $\leftrightarrow$ corresponding formal language construct
  - Example: to express “I am 6ft 5in tall” we need the concept of addition of measurements

- Some natural language concept are so generic that they apply to any skill
- Most important: **Time, Location**
- What are the formal constructs to support how humans refer to time & loc.?
- Can we define time & loc. once for all skills?
Why You Should Work on This Project

● Practically grounded: time and location are ubiquitous and improvements immediate

● No domain knowledge required
● Well-explored in literature
● Yet, commercial assistants don’t do so well!
  ○ E.g. Alexa supports limited set of absolute and relative dates

● Anytime algorithm: continuous improvement
  ○ You get to choose how much progress you make
  ○ You can stop any time if you run out of weeks in the quarter
Concept 1: Time

- Absolute: “on Tuesday Sep 22 2020 at 10:30 am PDT”
- Partial: “on Tuesday”
- Relative: “now”, “this week”
- Range: “in 2020”, “back in the 90s”
- Interval: “every 2 hours”, “every month”
- Frequency: “twice a day”
- Recurrency: “Monday through Friday at 9 am”
- ...

- Use cases: Alarms, Reminders, Restaurant Opening Hours, Weather Forecast
Concept 2: Location

- Absolute: “37.4249531° N, 122.1882103° W”
- Address: “353 Jane Stanford Way, Stanford 94305”
- Name: “Gates Computer Science”
- Personal: “My workplace”
- City: “Stanford”
- Administrative Area: “California”, “Santa Clara County”
- Zip Code: “94305”
- ...

- Use cases: Restaurants, Weather
High-level Project Plan

● Part 1: Need-Finding
   a. Collect use cases
   b. Collect sentences (from surveys and/or from HCI/NLP literature)
   c. Classify them by concept / feature

● Part 2: System Architecture
   a. Design in-memory representation
   b. Design syntax
   c. Implement in ThingTalk

● Part 3: Evaluation
   a. Add templates to Genie
   b. Generate dataset and train model
   c. Evaluate

● Part 4: Profit!
References: Time

- Conceptual and Quantitative Representations of Time Expressions, Mizobuchi et al. IJCPL 2000 https://doi.org/10.1142/S0219427900000223
References: Location

- GeoParsing Web Queries, Guillen,
  https://link.springer.com/chapter/10.1007/978-3-540-85760-0_98

- Detecting geographical references in the form of place names and associated spatial natural language, Leidner and Lieberman, SIGSPATIAL 2011
  https://dl.acm.org/doi/abs/10.1145/2047296.2047298

- Evaluation of NER systems for the recognition of place mentions in French thematic corpora, Brando et al. GIR 2016
  https://dl.acm.org/doi/abs/10.1145/3003464.3003471

- Alexa SDK: