Graph2step: A System for Knowledge Driven Procedural Step Generation

Pedro Colon-Hernandez

pe25171@mit.edu | Personal Robots

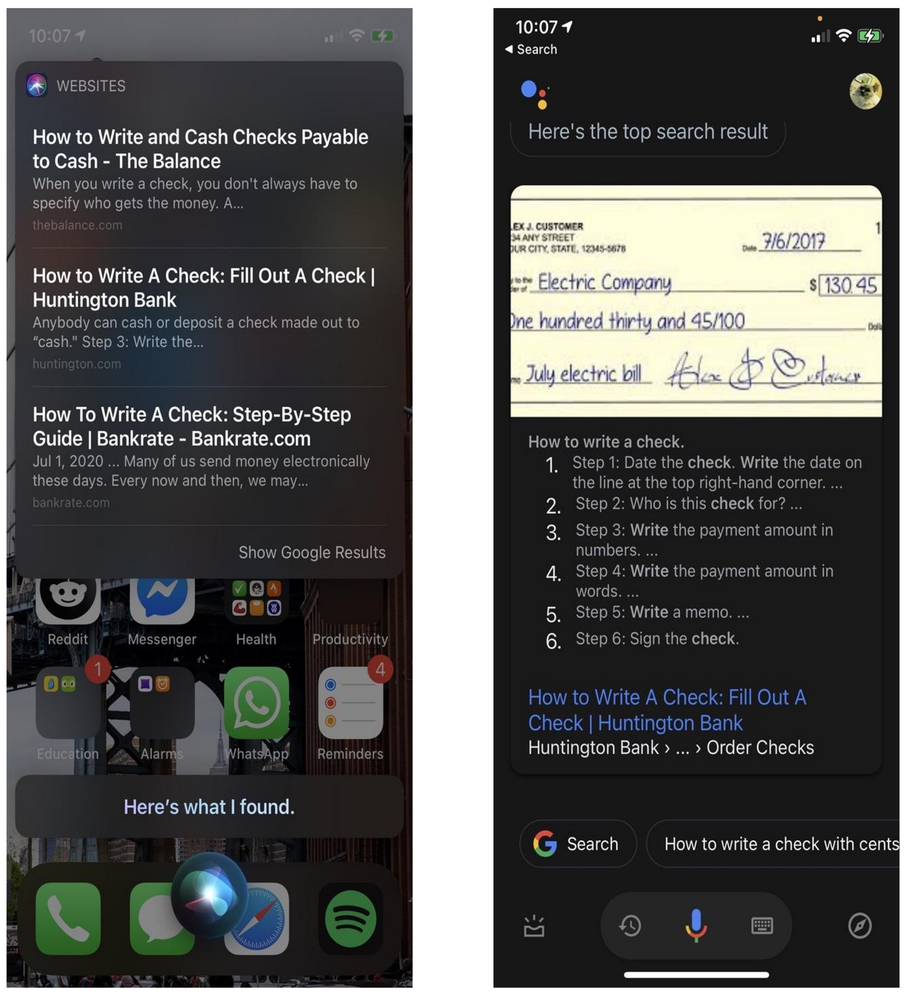

Figure 1. How-to handling of popular assistants. Usually the assistants refer to a web resource for users to perform their how-to procedures.

Conversational Agents for How-Tos

Instructed learning is ever-present throughout our lives. How-To questions, such as "How do I cook rice?", "How do I write a check?", or "How do I send pictures to my family from my iPhone?", are some of the most common queries for search engines and presumably of conversational agents as well. Answers to How-To questions are generally in the form of a procedure; step-by-step instructions that users perform in sequence.

However, people find reading instructions cognitively demanding and often prefer that another person guide them through a procedure. Prior work in automating procedural guidance either concentrates on how to communicate instructions or how to reason about procedural knowledge to extract states of entities. In this work, we present an in-progress end-to-end procedural voice guidance system that would be capable of automatically understanding, generating, and presenting a procedure through a conversational agent.

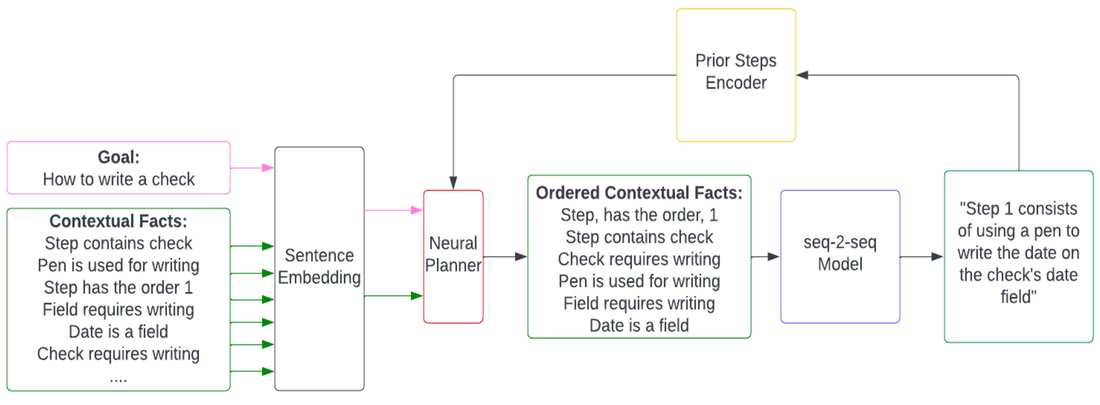

Figure 3. Architecture overview of graph2step system. From left to right, the system is trained by giving a set of contextual facts and a goal, which are embedded and passed into a neural planner along with any prior step encodings, to produce an ordered set of facts which are then translated into a procedural step.

Contextual Knowledge Graph Generation

To understand procedures, we utilize a large language model for contextual commonsense inference. Contextual commonsense inference is the task of generating a commonsense fact, from a given textual context. When these facts are collected, they form a knowledge graph of the given context. We utilize this model to ingest a textual procedural interaction, and generate a knowledge graph for every step of a procedure. The knowledge graph represents all of the possible information necessary to both understand and generate a procedural step. Our model additionally utilizes a mechanism called *hinting* to control the inference process. *Hinting* is a prompting technique that guides a model to generate facts about given entities or types of relations.

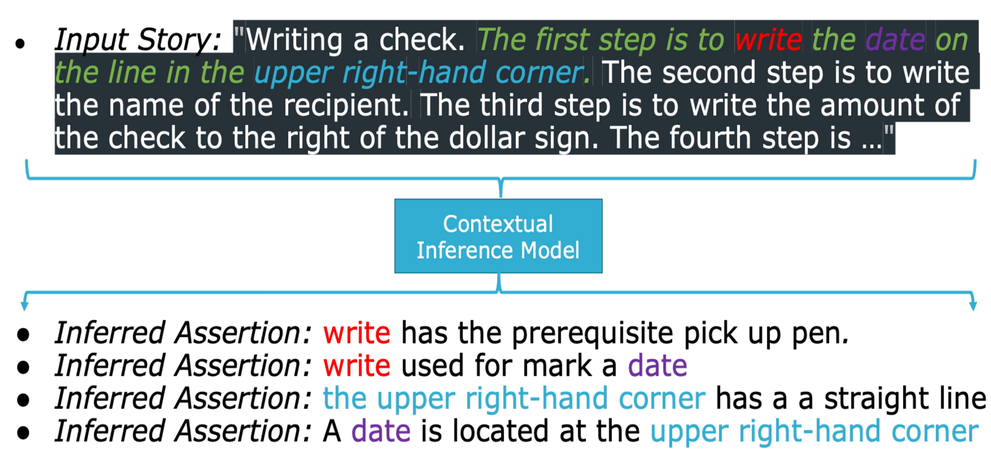

Figure 2. Contextual commonsense inference process. A story (top) is ingested by our model, and for a given sentence (highlighted in green), a set of facts or assertions are inferred.

Graph2Step: Converting a knowledge graph to procedure steps

To produce procedural steps, we built an another system called graph2step that utilizes the graph generated from contextual commonsense inference, to reason about the procedure and deduce an ordered set of facts that can be translated into steps. What makes this abstraction particularly interesting, is that if an error occurs during the procedure, the abstracted knowledge can be changed or updated to handle error-handling sub-procedures. Additionally, the paths or ordering can be recalculated to provide alternative explanations for procedures. When leveraged with natural conversation patterns, we can convey the steps in a procedure through a conversational agent.

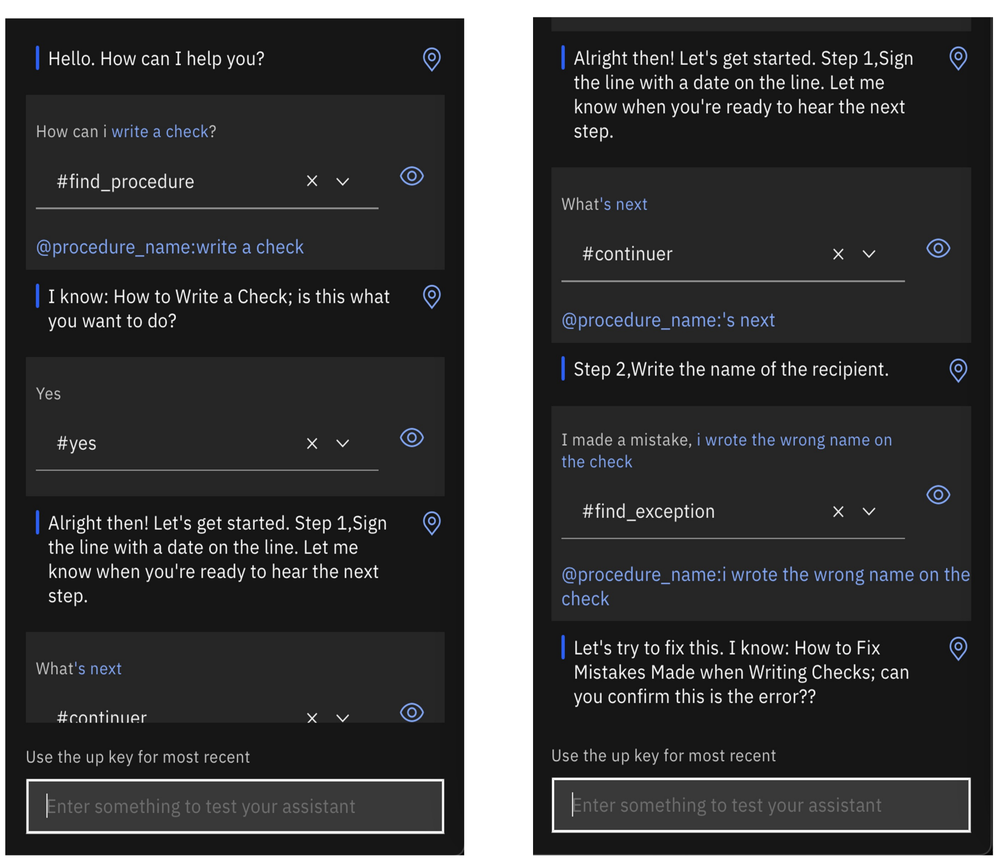

Figure 4. Watson Assistant providing automatically generated steps with conversational patterns.