

SKATE: A Natural Language Interface for Encoding Structured Knowledge

Clifton McFate, Aditya Kalyanpur, Dave Ferrucci,
Andrea Bradshaw, Ariel Dierani, David Melville, Lori Moon

Elemental Cognition
New York, NY, USA

{cjm, adityak, davef, andreab, arield, davidm, lorim}@elementalcognition.com

Abstract

In Natural Language (NL) applications, there is often a mismatch between what the NL interface is capable of interpreting and what a lay user knows how to express. This work describes a novel natural language interface that reduces this mismatch by refining natural language input through successive, automatically generated semi-structured templates. In this paper we describe how our approach, called SKATE, uses a neural semantic parser to parse NL input and suggest semi-structured templates, which are recursively filled to produce fully structured interpretations. We also show how SKATE integrates with a neural rule-generation model to interactively suggest and acquire commonsense knowledge. We provide a preliminary coverage analysis of SKATE for the task of story understanding, and then describe a current business use-case of the tool in a specific domain: COVID-19 policy design.

Introduction

Interactive natural language applications typically require mapping spoken or written language to a semi-formal structure, often represented using semantic frames with fillable slots. This approach has been used in popular commercial spoken dialogue systems (e.g., Google’s Dialogflow and Amazon’s Alexa skills) through the developer-defined “intents.” Frame semantic parsing more broadly (e.g., Gildea and Jurafsky 2002) has demonstrated benefit in a number of downstream applications including dialogue systems (Chen, Wang, and Rudnicky 2013) and question answering (Shen and Lapata 2007).

Despite advances in frame semantic parsing (e.g., Swayamdipta et al. 2017), no semantic parser is perfect. Accordingly, developers of natural language interfaces must carefully curate correction dialogues to avoid frustrating interactions. This sort of mismatch between system and user expectations is what we aim to resolve with SKATE (Structured Knowledge Acquisition and Extraction).

In SKATE, a user’s text is parsed in real time as they type. The resulting partial semantic structures can be completed with additional required slots and fillers, and are then recursively refined by the user through micro-dialogues. At any point, the user can continue to give structured interpretations for a slot filler (e.g., a complex noun phrase), or they

can leave it in unstructured form for the system to interpret later.

In the following sections, we first walk through the SKATE architecture using an exemplar interaction from an open-domain rule learning task. We then summarize our implementation of the core SKATE components. We demonstrate how SKATE has been integrated with a natural language rule generation model to interactively acquire structured rules for story understanding, and conclude with a current application that uses SKATE to build COVID-19 policy diagrams.

SKATE Overview

The SKATE architecture (Figure 1) is built around an interaction model of recursively: recognizing a concept, producing a partially interpreted template to instantiate the concept, and allowing a user to refine the template. The result is text annotated with semantic frames. These frames are processed by the downstream application.

Each interaction begins by selecting a top-level, application-specific semantic template. As an example, a top-level frame for rule acquisition may be an “If/Then” construction as in the first pane of Figure 2. These top-level templates provide the initial scope of interaction, and can be used to apply additional application-specific semantics as needed (e.g. If/Then could produce a causal rule while After/Then might only imply temporal sequence).

As a user fills a slot, the concept recognizer processes their text, selects a lexical trigger, and instantiates possible semantic frames for that trigger. The frames are in the style of FrameNet (Ruppenhofer et al. 2016): each defines a concept, a set of possible trigger phrases, and semantic arguments that may be instantiated as text spans. At each interaction, we use syntactic heuristics to select the trigger with the widest syntactic scope.

For each instantiated frame, the template renderer receives the interpretations (frame predicates and optionally argument labels/spans) from the concept recognizer and decorates that information (e.g., by adding display texts, examples etc) to send to the front-end UI. It also presents the user with options for what frame to assign as the word sense of the trigger. For example, in the second pane of Figure 2, the template generator has built frame assignment options for the word “take.” The resulting micro-dialogue is presented

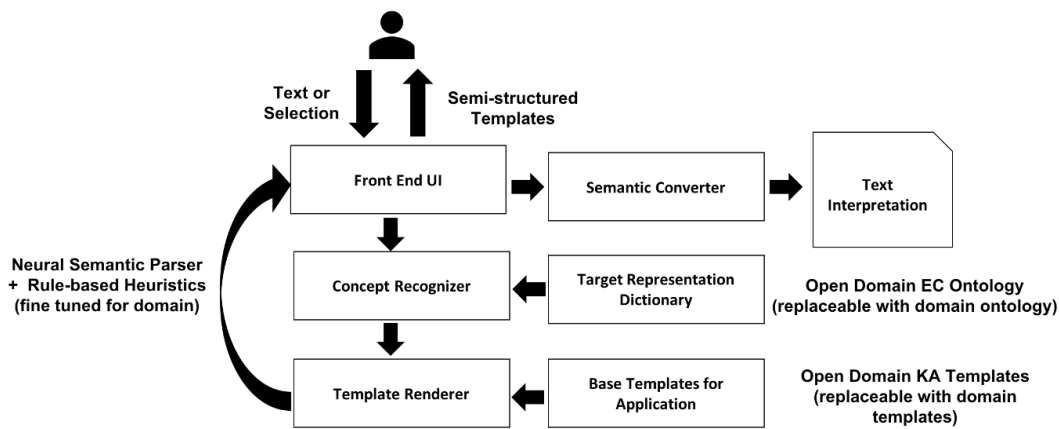


Figure 1: SKATE Architecture Diagram

to the user. Once an option is selected, the corresponding template is displayed, and the user can recursively refine unstructured slot fillers as in the third pane of Figure 2. The user can also choose to leave slots as unstructured text. For instance, in Figure 2, the user may not need to specify the desired sense of “cookie,,” and the entry can be submitted without full specification (in which case, uninterpreted tokens become placeholders/variables in the underlying semantic representation, and can be refined at a later point).

Note that in many domains, it is necessary to solicit extra information from a user given an evoked frame. When instantiating templates, required roles that remain unfilled can be added to the template to appear as blank slots for the user to specify (and must be filled in before the user submits). Additionally, likely roles suggested by context can be added and optionally deleted.

Once a user is satisfied with what they have typed, they can submit the entry. The set of composed frames can be further processed by the application-specific semantic converter if necessary. For example, in the rule application and COVID-19 policy builder, the resulting frames are turned into a set of Horn clause-like statements.

SKATE Components

In this section, we briefly describe the core SKATE components¹.

Target Representation Vocabulary

Our concept vocabulary is organized such that each predicate corresponds to a frame. All frames minimally possess the “focal” role which corresponds to the lexical trigger for the frame, though they may have additional optional and required roles. Frames are stored in an inheritance hierarchy, allowing multiple inheritance.

As a domain general starting point, we have created a frame ontology called Hector, derived from FrameNet (Ruppenhofer et al. 2016) and the New Oxford American Dictio-

¹Due to lack of space, we omit details for the Template Renderer which primarily sets up the UI look & feel.

Teach the system by providing knowledge

- If _____ then _____
antecedent consequent

Often when _____ then _____
antecedent consequent

Before _____ it is required that _____
outcome requirement
- If the child takes the cookie from the jar then _____
antecedent consequent

In “the child takes the cookie from the jar” did you mean **REMOVE, TAKE** take like remove as in :
 Someone removes an object from a container. Score: 1.00
He put the sandwich in the lunchbox.

In “the child takes the cookie from the jar” did you mean **TAKE** take as in :
 an agent consumes a substance: food, drink, medicine, etc. Score: 0.88
The party-goers consumed the spread.
- If the child takes the cookie from the jar then _____
agent object container consequent

By **COOKIE** cookie I mean :
 a small sweet cake, typically round and flat and having a crisp or chewy texture: Score: 1.00

By **COOKIE** cookie I mean any **biscuit** more generally. For example :
 biscuit, hardtack, cookie, cracker, wafer Score: 1.00

Figure 2: Example SKATE flow for entering “The child takes the cookie from the jar.”

nary (Stevenson and Lindberg 2010, NOAD). These two resources are complementary: FrameNet has broad coverage for multi-arity relations, while NOAD has a large library of lexical concepts (entities, attributes, etc.). The Hector ontology can easily be pruned into subsets for specific domains and/or expanded with novel concepts. Defining a new frame requires, minimally, defining its roles, writing a short definition or example, and optionally positioning it in the existing frame hierarchy. SKATE’s performance improves with annotated examples, but they are not required, and as discussed in the next subsection, SKATE can generate its own training data as a new frame is selected by the user and elaborated upon in SKATE interactions.

Concept Recognizer

The concept recognizer component consists of two semantic parsers. The first, SPINDLE (Kalyanpur et al. 2020a), is a transformer-based neural semantic parser. This model can be fine-tuned using a corpus, but requires annotated data. The second parser acts as a fallback and is used when Spindle returns no results or low confidence frame interpretations. It is based on an unsupervised approach that retrieves k nearest frames based on an embedding match between the sentence typed so far and potential frame embeddings, the latter being generated from minimal frame annotated examples pre-specified by a domain author². As SKATE is used in an application, the corrected output of the second parser becomes training data to improve the first. Thus, SKATE is able to improve with use.

Supervised Frame Semantic Parsing (SPINDLE) We have developed a neural semantic parser called SPINDLE that treats frame parsing as a multi-task problem involving related classification and generation tasks. Given a sentence and a frame-triggering span, the model decomposes parsing into frame-sense disambiguation (multi-label classification), argument span detection (generation), and role-labeling (classification). Since these tasks are related, SPINDLE uses a joint multi-task encoder-decoder architecture (see Figure 3), where the encoder layer is shared among the various tasks, with different decoders used depending on the task type.

The model is trained on 500K annotated frame sentences (available in FrameNet and NOAD) by fine-tuning a pre-trained, transformer-based language model such as GPT2 (Radford et al. 2019) or T5 (Raffel et al. 2020). The SPINDLE model achieved the best results using T5 as the base encoder/decoder, with a frame sense disambiguation accuracy of 91% and a span detection/role labeling F1 of 84%. Even though the parser was trained on full sentences, we have found that it returns results with high accuracy when run on partial sentences like those typed in SKATE. Moreover, as the user continues to type text, the parsing results change to consider the additional context, which helps to disambiguate the correct frame sense.

²This feature supports quick domain adaptation when there isn’t sufficient training data for a supervised semantic parsing model.

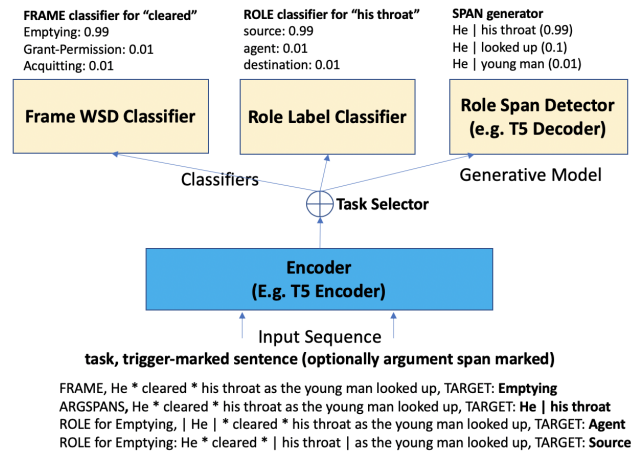


Figure 3: SPINDLE: Joint multi-task encoder-decoder model for semantic parsing

Embedding-Based Heuristic Parsing To complement the neural semantic parser, which needs many annotated examples for training, we have developed an unsupervised, k -NN-based approach for frame parsing that can work with a handful of examples per frame. The approach first computes a frame embedding by aggregating GLoVe (Pennington, Socher, and Manning 2014) embeddings for trigger lemmas (which are specified in the frame definition) and content words in frame examples. Our tool then sums the GLoVe embeddings for all words in the sentence to produce a sentence embedding, and computes the similarity between the frame and sentence embeddings. The algorithm also detects argument spans using syntactic heuristics based on a dependency parse of the sentence. Finally, it assigns a role for each span by considering how well the type of the span phrase matches the expected role type as inferred from frame examples (type similarity checking is also done using embeddings).

Semantic Converter

The result of a submitted entry is a possibly incomplete frame-semantic parse of the input text. The semantic converter can also contain domain-specific logic to further convert the frame semantic interpretation into usable data for a downstream application (as described in the Domain Adaptation section).

Commonsense Knowledge Acquisition for Story Understanding

SKATE has been applied for open-domain structured rule acquisition. The task is: given a short story and a question, provide a rule or set of rules with which the answer can be derived from the story. Using SKATE, we can collect structured formal rules usable by a downstream reasoning engine.

As described above, SKATE templates are meant to guide the user both with explicit structure (e.g., slots) and, optionally, with unstructured slot-fillers. These unstructured fillers

POSSIBLE CAUSES »	EVENT / STATE	» POSSIBLE EFFECTS
<ul style="list-style-type: none"> Olivia feel(s) competitive » Olivia want(s) competition <li style="padding-left: 20px;">Olivia want(s) fun Olivia feel(s) competition Olivia want(s) to play soccer Olivia goes to a soccer game Olivia wants to play soccer <li style="padding-left: 20px;">Olivia plays soccer <li style="padding-left: 20px;">Olivia likes soccer Olivia likes to play soccer 	<div style="background-color: #e0e0e0; padding: 5px; display: inline-block;">Olivia plays in a soccer game</div>	<ul style="list-style-type: none"> » Olivia runs toward the goal Olivia want(s) success Olivia feel(s) competitive Olivia want(s) to win Olivia has the ball Olivia feel(s) happy Olivia feel(s) excited Olivia gets the ball She has the ball Olivia scores a goal

Figure 4: GLUCOSE suggestions generated from a children’s story about soccer

can be used to guide the user to submissions with high-confidence semantic parses or towards prototypical examples. For this task we integrate SKATE with a neural unstructured rule prediction system to guide the user towards general, syntactically simple, rules.

GLUCOSE: Contextual Rule Prediction

GLUCOSE (GeneraLized and COntextualized Story Explanations; Mostafazadeh et al. 2020) is a crowd-sourced dataset of common-sense explanatory knowledge. GLUCOSE defines ten dimensions of causal explanation, focusing on events, states, motivations, emotions, and naive psychology. The GLUCOSE dataset consists of both general and specific semi-structured inference rules that apply to short children’s stories. These rules were acquired via crowd-sourcing, and Mostafazadeh et al. (2020) demonstrated that neural models trained on these semi-structured rules could be used to produce human-like inferences for story understanding.

Following Mostafazadeh et al. (2020), we train an encoder-decoder rule generation model. For each sentence in a story, we use the GLUCOSE trained model to predict unstructured textual causal inferences. These uninterpreted inferences are then used to seed slots in SKATE rule templates, guiding the user towards high-likelihood story-relevant rules (see Figure 5).

Teach the system by providing knowledge

If Olivia plays in a soccer game then

antecedent
Olivia wants to win
consequent

In "olivia plays in a soccer game" did you mean **play**

like **compete** as in :

Competitors participate in a competition.

The kids play softball.

COMPETE, PLAY

Score: 0.96

Figure 5: GLUCOSE suggestion for a story about soccer.

The GLUCOSE-trained model can also be used for *auto-complete* suggestions in SKATE. As the user types text in

one of the structured template slots, we run the model on the text typed so far (i.e., in earlier slots of the template) and generate potential completions. A novel feature of SKATE is that we use the already specified frame semantics to filter out incompatible language model suggestions. For example, say the user is providing knowledge about a soccer story, and starts typing: “*If a player gets*” and specifies the interpretation for the verb “get” as the frame *arriving-at-a-location*. At this point, the frame template has an unfilled slot for “destination”. Suppose the user continues by typing text in this slot, and we use the GLUCOSE model to generate completions, it may produce the following alternatives: “*..a ball*”, “*..to the goal*”, “*..into trouble*” given the prior text “*If a player gets*”. However, because the user has specified the frame semantics for “get” and the active slot is “destination”, the only compatible suggestion is “*..to the goal*”. To identify compatible suggestions, we run the SPINDLE semantic parser on the full generated completion (including the prior text) and filter out suggestions where the frame doesn’t match the prior specified frame. In the above example, we would throw out “*gets a ball*” (where get means *acquire*), and “*gets into trouble*” (where get means *transition-to-state*), since it does not match the earlier specified interpretation of “get” (*arrive-at-location*). We believe that suggesting text completions that are consistent valid semantic interpretations given prior context is unique to SKATE.

Coverage Analysis

As a preliminary coverage evaluation, we asked domain experts to use the tool to encode knowledge needed to answer and explain commonsense questions generated from children’s stories. The questions and required rules were generated in English as a part of several manually created story understanding rubrics (Dunietz et al. 2020). Our data set consisted of 340 target natural language rules from 11 children’s stories. Rules could range in complexity from simple attributive statements (e.g. “often, a house has a yard.”) to complex script-like statements (e.g. “If a person plays soccer and the person belongs to a team and the person moves the ball to the goal then the team gets a point.”).

To test declarative statements (factoids), we additionally asked the annotators to enter, exactly or as a paraphrase, 67 sentences from an additional 4 stories.

Annotators were trained on how to use the SKATE interface and then, for each rule or statement, they rated how close in intended meaning their resulting entry was to the original NL expression on a scale from 0-3 (0 = not close; 1 = substantial deviation; 2 = minor deviation; 3 = paraphrase).

Results were promising, with 85% of entries scoring 2 or higher, including several complex constructions involving nested clauses, conjunctions and negation. Some high scoring examples are shown in Table 1. The main gaps were missing frames from the target ontology (Hector).

Knowledge Target	SKATE Input (Score)
People generally want to eat food that is tasty	Often people want to eat tasty food (3)
When a larger animal approaches a smaller animal, the smaller animal might get afraid	Often when animal1 approaches animal2 and size of animal1 is greater than size of animal2, animal2 feels fear (3)
When one person helps another, the person being helped thanks the helper	Often when person1 helps person2, then person2 thanks person1 (3)
If something is not obscured behind another object, it can be seen	If object1 does not cover object2, then someone can see object2. (2)
If someone doesn't know something, and someone else tells them, then they know what it is	If person1 does not know a fact and person2 tells person1 the fact, then person1 learns the fact (3)

Table 1: Commonsense rules acquired via SKATE

Annotation Results and Feedback

SKATE was used by 5 domain / subject matter experts (SMEs) to specify knowledge on the story understanding task, and as mentioned above, we collected 400+ structured logical statements (rules and facts) via the tool. Note that the alternate to using SKATE, and the prior methodology used by the SMEs, was to hand-code the formal knowledge in a verbose representational syntax (similar to Prolog), which was tedious, error-prone and time-consuming.

The SMEs impressions can be summarized as follows:

- The tool is dramatically faster than writing rules in logical syntax not just because the user is writing in natural language, but also because the tool actively coaxes the desired interpretation while moving along (instead of landing up in a undesired/confused state and backtracking).
- The user does not need to worry about special character formatting (commas, parentheticals etc) which need to be specified correctly when encoding logical formulae.
- The dynamic, organized nature of the information-flow to the user, such as options regarding word senses and phrase interpretations, ensures that the rules the user produces are compatible with the resources with which they are intended to interact. In traditional approaches, writing the same rules would often require a user to check the syntax against a set of rules and the word meanings against a sense ontology as separate steps.

Path to Deployment

As a next step for SKATE on story understanding, we plan to use it to collect formal causal explanatory knowledge via crowdsourcing, as an extension of our previous work, GLUCOSE. The current version has a “SKATE-lite” UI with subject-verb-object (SVO) templates (see Figure 6).

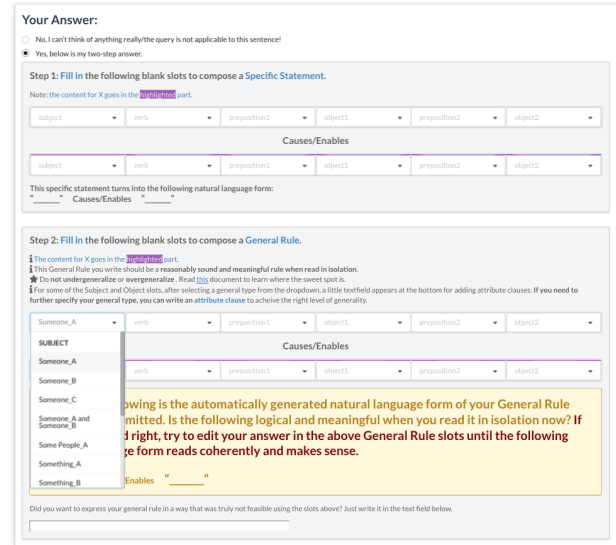


Figure 6: Current GLUCOSE UI for collecting causal explanatory rules

By using SKATE, we should be able to collect richer, more expressive knowledge, and our coverage analysis, done on the same domain and a similar/related task, has shown that the tool is more than adequate for this problem. We are keen on this deployment as it would expose SKATE to hundreds of our trained turkers.

Domain Adaptation: COVID-19 Policy Diagram

As the world recovers from COVID-19, many institutions have been required to define robust facility access policies. These policies can be complicated, often with many branching conditions (e.g. lists of symptoms) and potential actions for a user to complete (e.g. various policy compliant COVID tests).

Automated systems can help guide users through these policies, but the policies must first be formalized. In the following section, we present an application of the SKATE NLI for building domain-specific policy diagrams around access to school facilities.

Task: Defining Compliance Paths

A policy diagram is defined by

- Compliance states: terminal actions, whether a person returns or quarantines.

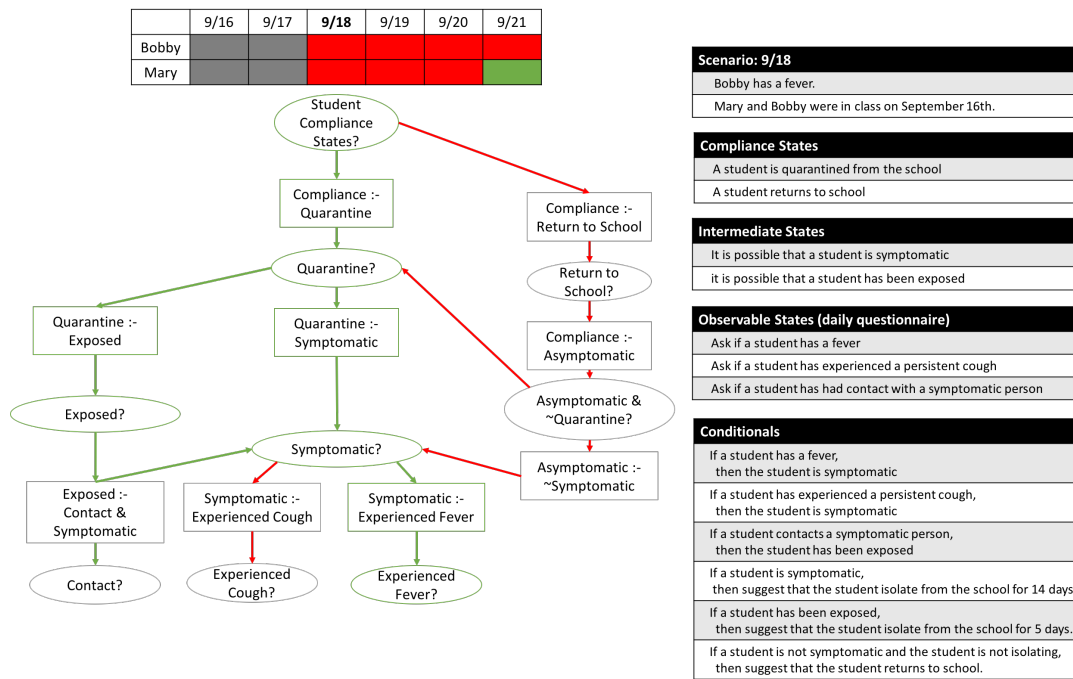


Figure 7: An example policy pertaining to school access. Queries can be issued against the graph given a world state to determine compliance. Here, a query for compliance states reveals two students currently under quarantine

- Intermediate states: States that lead to compliance states or further modify them, e.g. quarantining because a student is symptomatic.
- Scenarios: observable states that lead to an intermediate state, e.g. A student experienced a cough and a fever.
- Variables: observable from the world, e.g. a person marked on a questionnaire that they experienced a cough.

Together these form a flow chart (policy diagram). Nodes in the diagram are states, and each type of state (above) is assigned a top-level template to allow a user to define them.

Compliance and intermediate states can be mapped to a unique frame instance or combination of frames which allows for compositionality (e.g. quarantine for 14 days / quarantine at home). Variables are also compositional (e.g. has a persistent cough) and can be inferred by the system using rules or observed directly through an end-user questionnaire. An example of acquiring and applying a policy diagram is shown below.

Example

Figure 7 shows a simplified COVID policy for returning to school along with the SKATE statements used to construct the policy. In this example, *quarantining (from school)* and *returning (to school)* have been defined as compliance states. Other rules append *adjuncts* (optional roles) to a state (e.g. duration) when it is evoked to further specify it. Thus, we can define conditions that lead to 5 or 14 day quarantines based on whether the student was exposed or symptomatic respectively.

Figure 8 is an example entry defining an intermediate state. Later conditionals can elaborate these states (e.g. symptomatic if feverish) and target them.

It is possible that a student is symptomatic
symptomatic
entity

Figure 8: Defining an intermediate state

Figure 9 is an example entry defining a suggested compliance state given an intermediate state. The state is compositional, specifying a population to quarantine from. These conditionals form rules usable by a reasoning system.

If a student is symptomatic then suggest that
the student isolate from school
agent population

Figure 9: Defining a branch of a policy diagram

We also define intermediate states (*exposed* and *symptomatic*) to intuitively provide reasons for suggesting a quarantine. These hold given combinations of observable facts, which can be set through a daily questionnaire.

In our representation, *duration* adjuncts on states map to counters which can align to a calendar. Thus, we can chart when a student will be able to return to school.

The world state (i.e. a specific scenario) is also defined in

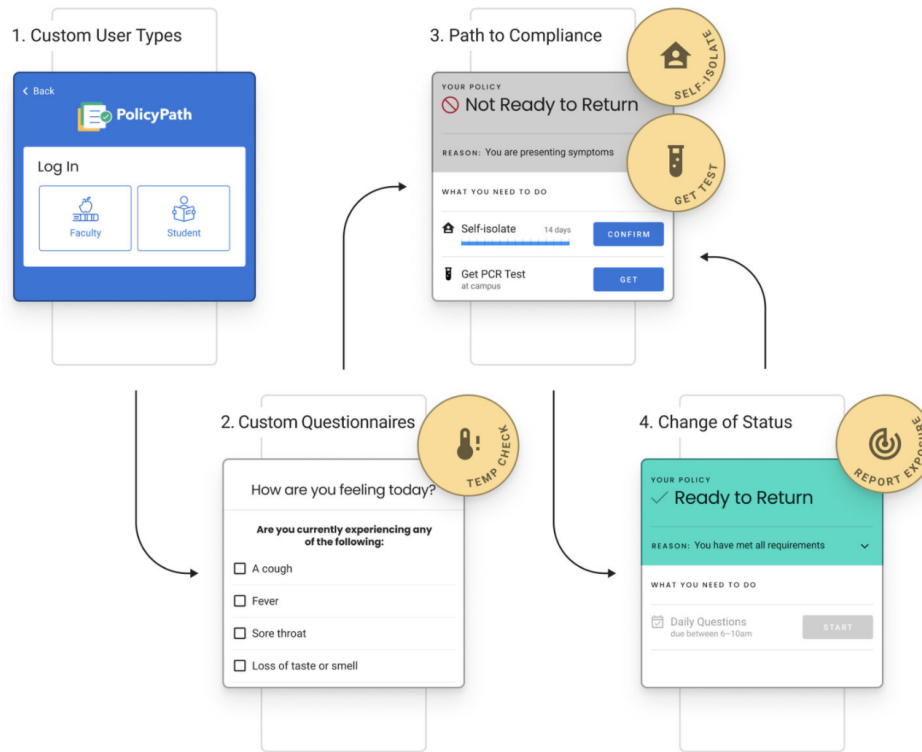


Figure 10: Policy-Path Mobile App for COVID-19

SKATE (as shown in the Figure). Interestingly, in this example, "Mary and Bobby were in class.." is interpreted as *co-location* in SKATE, which is used to infer *contact* between the two via background knowledge in the ontology.

Given a world state (defined in SKATE), an administrator can query the graph to determine which students are in which compliance states. In the example, the system correctly infers that Bobby has 14 days left to quarantine on 9/18, while Mary only has 3 days left.

Path to Deployment

A team at our company, Elemental Cognition, has built a mobile app called *PolicyPath* for customized and interactive COVID policy guidance. Details of the app are available at <https://www.policypath.app/covid-19>.

The application, currently in use by several customers, allows an organization to define and tailor their COVID policy, which specifies the conditions that employees must satisfy before they are allowed to return to the workplace. It supports the ability to specify a custom questionnaire, testing requirements, quarantine periods, temperature checks, and more, while also providing real-time monitoring and alerts (see Figure 10).

In the current setup, the policy definition is typically done by the organization's management (e.g. HR department), often in a Word document, and subsequently, the logic is captured in procedural code by knowledge engineers and software developers. Our goal is to replace this process by us-

ing SKATE, as described above. The policy writer can use SKATE to express the policy in natural language, which is automatically converted to a declarative logic form under the covers. We can then use our logical reasoning and constraint checking engine called Braid (Kalyanpur et al. 2020b) to enforce the semantics of the policy, when applied to a specific individual (or a group of individuals). We believe it is much easier to define, alter and test the policy when done in this "no code" declarative paradigm.

Related Work

Natural language knowledge capture has long been a goal in AI, and interest has only grown with the advent of crowd sourcing platforms like Amazon's MTurk. Our approach draws inspiration from and improves upon this research.

ConceptNet (Speer, Chin, and Havasi 2016) started as the Open Mind Common Sense crowd-sourcing effort (Singh et al. 2002) which solicited natural language common sense statements. While the OMCS knowledge acquisition interface could make use of semi-structured templates, their captured knowledge remains as natural language and they do not further decompose an entry into semantic forms. Their approach additionally used generated natural language inferences for user feedback. This plays a similar role to our auto-complete feature, though their feedback is presented after the fact rather than as inline guidance.

LEARNER (Chklovski 2003) uses cumulative analogies to gather new information from ConceptNet like statements

(e.g. newspapers have pages) via answerable questions (e.g. do books also have pages?). LEARNER2 builds on that design by adding templates with slots for a small set of target top-level relations (Chklovski 2005). They also generate slots to enumerate an entry, however, much like OMCS, they do not further refine input text with templates.

Our approach leverages recent advances in language modeling to generate templates from user text and to provide unstructured guidance. Recently (Gopinath et al. 2020) presented a “contextual auto complete” approach for clinical documentation which used a completion mechanism to disambiguate clinical concepts and create annotated notes. In contrast, our completion mechanism (templates and unstructured text) is far broader in scope (interpreting the full text) and depth of representation (compositional frames).

Conclusion and Future Work

While great advances will continue to be made in the field of semantic parsing, it is highly unlikely that any parser will always perform perfectly. As such, even when a natural language application is capable of a desired behavior, lay users face uncertainty and obstruction when their requests are wrongly interpreted.

SKATE is a Natural Language Interface that reduces the mismatch between system ability and lay user expectation by interactively guiding them towards a structured representation. Our approach combines frame-based KR with a hybrid semantic parsing approach to construct interpretations with both structured (i.e. template slots) and unstructured (i.e. textual slot fillers) content. A novel aspect of the hybrid parsing approach is its potential to automatically improve with use, since the unsupervised embedding based parser acts as a vehicle to collect training data for the supervised model. Furthermore, the use of a neural rule generation model to produce semantically valid auto-completions is a novel and significant feature from a usability standpoint.

We have demonstrated the utility of the SKATE NLI in both an open domain task (story understanding) and in a highly specialized domain (building policy diagrams). We plan to host a public endpoint demonstrating SKATE shortly.

Many challenges still remain as we integrate SKATE into an end-user application, e.g., we are exploring ways to allow users to create new frames and/or slots on the fly, when the pre-defined vocabulary is insufficient.

References

Chen, Y.-N.; Wang, W. Y.; and Rudnicky, A. I. 2013. Unsupervised induction and filling of semantic slots for spoken dialogue systems using frame-semantic parsing. In *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, 120–125. IEEE.

Chklovski, T. 2003. Learner: a system for acquiring commonsense knowledge by analogy. In *Proc. K-CAP '03*, 4–12.

Chklovski, T. 2005. Designing interfaces for guided collection of knowledge about everyday objects from volunteers. In *Proc. IUI '05*, 311–313.

Dunietz, J.; Burnham, G.; Bharadwaj, A.; Chu-Carroll, J.; Rambow, O.; and Ferrucci, D. 2020. To Test Machine Comprehension, Start by Defining Comprehension. *ACL2020 Theme Track abs/2005.01525*.

Gildea, D.; and Jurafsky, D. 2002. Automatic labeling of semantic roles. *Computational linguistics* 28(3): 245–288.

Gopinath, D.; Agrawal, M.; Murray, L.; Horng, S.; Karger, D.; and Sontag, D. 2020. Fast, Structured Clinical Documentation via Contextual Autocomplete. In *Machine Learning for Healthcare Conference*, 842–870. PMLR.

Kalyanpur, A.; Biran, O.; Breloff, T.; Chu-Carroll, J.; Dieriani, A.; Rambow, O.; and Sammons, M. 2020a. Open-Domain Frame Semantic Parsing Using Transformers. *arXiv preprint arXiv:2010.10998*.

Kalyanpur, A.; Breloff, T.; Ferrucci, D.; Lally, A.; and Jantos, J. 2020b. Braid: Weaving Symbolic and Statistical Knowledge into Coherent Logical Explanations. *arXiv preprint arXiv:2011.13354*.

Mostafazadeh, N.; Kalyanpur, A.; Moon, L.; David Buchanan, L. B.; Biran, O.; and Chu-Carroll, J. 2020. GLUCOSE: Generalized and Contextualized Story Explanations. *EMNLP 2020*.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *Proc. EMNLP '14*, 1532–1543.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language Models are Unsupervised Multitask Learners. *OpenAI Blog*.

Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. J. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *JMLR* 21(140): 1–67.

Ruppenhofer, J.; Ellsworth, M.; Schwarzer-Petruck, M.; Johnson, C. R.; Baker, C.; and Scheffczyk, J. 2016. *FrameNet II: Extended theory and practice*. Berkeley, CA: International Computer Science Institute.

Shen, D.; and Lapata, M. 2007. Using semantic roles to improve question answering. In *Proc. EMNLP-CoNLL '07*, 12–21.

Singh, P.; Lin, T.; Mueller, E. T.; Lim, G.; Perkins, T.; and Zhu, W. L. 2002. Open mind common sense: Knowledge acquisition from the general public. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, 1223–1237. Springer.

Speer, R.; Chin, J.; and Havasi, C. 2016. Conceptnet 5.5: An open multilingual graph of general knowledge. *arXiv preprint arXiv:1612.03975*.

Stevenson, A.; and Lindberg, C. A., eds. 2010. *New Oxford American Dictionary, Third Edition*. Oxford University Press.

Swayamdipta, S.; Thomson, S.; Dyer, C.; and Smith, N. A. 2017. Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold. *arXiv preprint arXiv:1706.09528*.