Human-Centered AI for Augmented Development of ad hoc Information Structures

Graduate Thesis Proposal

Thurston Sexton

15 December 2022

Abstract

There is an increasing use of NLP and text-mining systems for assistance in the creation of ad hoc information structures. > see: ontology/knowledge-graph mining, automated physics-based process-model extraction from papers, taxonomy embedding and recommendation systems, etc.

There is a need to formalize a design process and more general methodology for these use-cases, since current NLP/ML community practice is often founded on the assumption of human-annotation-as-oracle This conflicts directly with many other communities' long-standing practices in creating and validating these ad hoc structures for wider adoption and repurposing. It is especially problematic when these violating assumptions are not communicated, and the ML/NLP systems are presented *to these communities* as tools for reducing labor intensity when building ad hoc information structures as they have done for a long time.

We address this need by first performing a cognitive task analysis for the task of "developing an ad hoc knowledge structure", specifically in the context of human-AI coordination in the presence of observation data. This will allow us to systematically identify key HCAI system components that would be likely to produce errors, contextualize make explicit the assumed cognitive and computational loads involved, and guide future design strategies to mitigate errors by taking these loads into account.

Then, guided by this foundation, we design solutions to two highly-problematic subtasks in the HCAI system, namely:

- Lack of viable data representation for moving between knowledge structures and mathematical structures used for statistical methods. Structures in use today suffer from a lack of explicit assumptions about the knowledge structures' topology, prevent compatibility and reuse, and encourage human error when developing data pipelines in code.
- Consistent