Location: Online (Internship from Home)

Task: Knowledge Graph based Text Entailment in

Software Engineering context

In the context of specific domains such as automotive, this internship aims towards addressing the following fundamental research questions:

* 1. What are the formal machineries for ensuring entailment between short texts and domain corpus that also remain sparse in nature? Can Knowledge Graphs be leveraged to confirm this entailment?
  2. How do we inference “knowledge gaps” in automatically extracted domain knowledge graphs from sparse corpus? Can Graph Completion techniques be developed and applied towards this?
  3. How do we evaluate domain knowledge graphs using graphical entropy measures and quantify the gaps/uncertainty?
  4. How do we translate information from knowledge graphs to generate meaningful context specific texts without losing compositional semantics

# Domain for Illustration – Automotive Domain Software Development

Requirements specification documents generally involve uncertainty and conflicting stakeholders’ objectives. Goal Models as formal representations are useful in prioritizing such objectives by highlighting conflicts and variations of required features. Goal Models are also pivotal in model driven development of systems, leading to generation of process models, use case models, data models and operational models. Currently the goal models are constructed through a human intensive exercise that are error prone and are often repeated with changing requirements.

Therefore, we propose automated generation of goal models from software requirements using generative neural networks as a research problem statement. We believe this can be formalized as a NLP driven graph generation problem by subscribing to standard goal model notations such as I\* or KAOS. As part of this proposed internship work, one can explore generation of Goal models as Hyper Graphs, lattice models from requirement specification documents. Employing LSA and word embedding techniques for variability analysis to formally differentiate AND /OR refinements of a given requirement need to be considered.

# Deliverables

The internship is expected to result in following deliverables

1. Tool for Confirming Text Entailment between the pair of short text and sparse domain corpus
2. A formal machinery with corresponding list of algorithms to generate graphical representations from sparse domain corpus. The graphical representations need to be optimal in terms of information-theoretic centrality measures.
3. A formal machinery with graph based inferencing techniques to reduce noise in constructed knowledge graphs in terms of defined entropy measures
4. Identification of entropy measures that can be used for evaluating and comparing knowledge graphs for specific domains.

# Internship Request Details:

1. Number of Interns : 2
2. Preferred Educational Qualifications : Final year Masters Students or Early PhD Students
3. Expected Skills : NLP, Knowledge Engineering, Symbolic Reasoning, Graph Theroy, Python, Tensor Flow