

Weekly report

Aug. 5 – Aug. 9

Progress

- Build labeling specifications for text2scene data
- Analyses of the current model, in terms of runtime, performance and feature contributions
- Preparation for the next version of text2scene, regarding feature engineering and the model design

Skills

- Learned about sophisticated textual embedding techniques
- Related papers
 - Top-k Exploration of Query Candidates for Efficient Keyword Search on Graph-Shaped (RDF) Data
 - A Semantic Approach to Recommending Text Advertisements for Images
- Related tools
 - [BabelNet](#)
 - [ConceptNet](#)
 - [Wikipedia-based Explicit Semantic Analysis](#)

To do

- A feature encoder that properly tackle the curse of dimensionality
- Synonym replacement to handle unseen tokens
- Make use of babelNet or ConceptNet to encode the similarity between tokens with different POS tags
- Optimization on the runtime efficiency