**Manuscript title:** NLP-based approach to classifying heterogeneous terms for unambiguous exchange of roadway data

**Manuscript #**: CPENG-1952

**Summary of research contribution**

The data consistency among project phases, stakeholders, or geographic regions (counties, states, etc.) is a major barrier to digital exchange of life cycle project data in the highway industry. However, the current manual process of identifying the semantic relations among heterogeneous in the convention ontology development methods for the highway sector is laborious and time-consuming due to the lack of an effective automated method. This paper proposes an NLP based methodology to assist professionals in extracting roadway terms and their semantic relations from text documents. A key contribution to the body of knowledge is the novel framework with a new algorithm that allows for automated detection of technical terms and their relations from roadway text documents without reliance on existing hand-coded dictionaries. The present framework is not to completely eliminate the human interfere, but is expected to become an enabling tool that can help researchers in the domain quickly develop supporting ontologies and other forms of semantic resources for their specific use cases. With respects to the facilitation of semantic interoperability for the infrastructure sector, the implications of this study would accelerate the process of removing the current bottleneck in extensive machine readable dictionaries which are required for an unambiguous data sharing, integration or exchange.