# Leverage text mining for analyzing accident reports of automated train control systems

Mona Ahmadi Rad, Lianne M. Lefsrud, Michael Hendry

Introduction

Railway occurrence databases and reports are great sources of safety information involving the semi-structured and unstructured textual descriptions of railway incidents or accidents along with numerical data.

Recently, the advent of natural language processing (NLP) technology has offered tremendous possibilities for automatically exploring text corpora and discovering implicit information. This contains everything from information retrieval to text classification and clustering, to the entity, relation, and event extraction.

TSB   FRA

RAIB   ATSB

RODS   NTSB
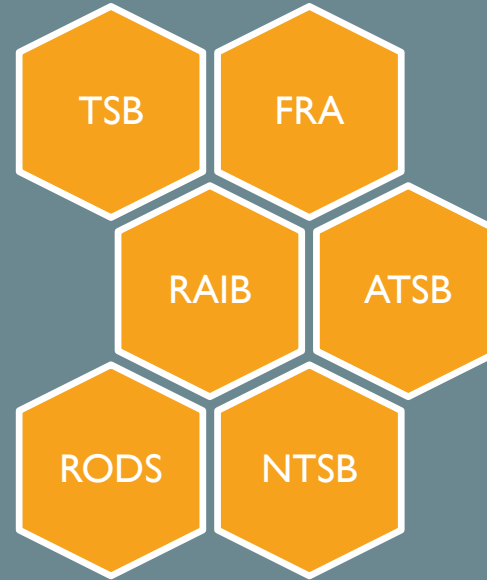
Objectives

- leverage text mining for discovering insights, trends, and patterns of the unstructured and semi-structured textual reports of railway occurrences.
- Extract human errors related to using automated train control systems, human performance shaping factors, and their relationships and interactions.

Research Steps

Task 1

Extract accident factors and entities of interests from the textual accident reports

Task 2

Identify relationships between accident factors and/or entities of interests

Task 3

Perform text classification and clustering