



# Improving Information Retrieval-based Concept Location using Contextual Relationships

by  
Tezcan Dilshener



# Maintenance Challenges



- **IR performs concept location based on lexical similarities.**
  - Between search terms and the terms from source code.
- **Some concepts cannot be identified by looking at terms.**
  - Multi word ( $n$ -grams) concepts.
    - e.g. “*standalone risk*”.
  - Compound concepts, e.g. “*covariance*”
    - derived from “*correlation*”, “*volatility*”, “*market index*”.
- **Inadequate relational context of how concepts interrelate;**
  - between program elements.
  - within the application’s business domain.

# Research Method

- **Creation of a contextual model providing strong relational clues by combining:**
  - domain specific ontologies.
  - natural language (NL) approaches.
  - programming semantics e.g. call-graph (CG).
- **Validation by empirical analysis**
  - quantitatively measured by precision and recall.
  - comparing the results against
    - our gold standards on industrial applications.
    - those obtained by the extended approaches.



# Contributions

- Novel concept mapping approach using domain specific ontology relations.
- Detection of all relevant program elements implementing n-gram concepts.
- Ontology supported abbreviation expansion.
- Semi-automatic extraction of concepts from source code.