A Contextualized Vocabulary Model for Identifying Technical Debt in Code Comments.

Mário André, André Batista, Manoel Mendonça, Rodrigo Spínola
Agenda.

• Introduction
  – Motivation
• A Contextualized Vocabulary Model for Identifying Technical Debt (CVM-TD)
• Exploratory Study
• Conclusion and Future Works
Motivation.

• Different indicators have been used by automated approaches to identify TD[2];
  – These indicators use software metrics.
• The problem with these approaches...
• In order to complement these quantitative analyses, we developed the CVM-TD
Agenda.

• Introduction
  – Motivation
• A Contextualized Vocabulary Model for Identifying Technical Debt (CVM-TD)
• Exploratory Study
• Conclusion and Future Works
CVM-TD

- CVM-TD is a contextualized collection of terms that focuses on using word classes and code tags.
To generate a NP, we combine [Verb] + [Adj] + [SE Noun]

SE noun is the central element in a NP.

e.g. bad (Adj) + code (Noun).
remove (Verb) + method (Noun).

NP is a composed terms that provide high context information
We combine:

Expression = [Adv] + [Tag] + [NP]

Expression = [Adv] + [Tag] + SE Noun

e.g.: “Todo: should fix the bad code”

Here, we have a contextualized expression.
CVM-TD.

- CVM-TD provides a set of TD vocabulary that may be used to filter comments that need more attention because they may indicate a TD.

- The contextualized vocabulary can be used to support TD identification.
Agenda.

• Introduction
  – Motivation
• A Contextualized Vocabulary Model for Identifying Technical Debt (CVM-TD)
• Exploratory Study
• Conclusion and Future Works
Exploratory Study

• We performed an exploratory study to analyze the viability of the proposed model to support comments analysis for TD identification
Exploratory Study.

Selecting the projects and participants

Extracting comments

Filtering comments

Analyzing comments

- jEdit and Lucene;
- Four masters students analyzed the projects;

Set of comments

Set of candidate comments

Set of comments that can indicate a TD item
Agenda.

• Introduction
  – Motivation

• A Contextualized Vocabulary Model for Identifying Technical Debt (CVM-TD)

• Exploratory Study

• Conclusion and Future Works
Contributions of work

• Our findings show that:

  – (i) the dimensions considered by the model are used by developers when writing comments,
  – (ii) CVM-TD provides a vocabulary which may be used to detect TD, and
  – (iii) In spite of our outcomes being promising, the proposed model needs to be improved and evaluated with other projects.
Future works.

• Identify points that may improve CVM-TD;
• Implement CVM-TD model and integrate it with a comment analysis tool;
• Evaluate the use of CVM-TD in other software projects.
References


Questions
A Contextualized Vocabulary Model for Identifying Technical Debt on Code Comments

Mário André, André Batista, Manoel Mendonça, Rodrigo Spínola