
**University of L'Aquila**

**DEPARTMENT OF ENGINEERING COMPUTER SCIENCE AND MATHEMATICS**

**Master degree in Software Engineering for Adaptive Systems**

# AUTOMATED APPROACHES TO ASSESS THE SIMILARITY OF OPEN SOURCE PROJECTS

Thesis Advisor:
**Davide Di Ruscio**

Thesis Co-Advisor:
**Phuong T. Nguyen**

Candidate:
**Riccardo Rubei**

# Table of Contents

# Introduction

## Scenario

# Introduction

## Challenges

- Searching for canditate components.
- Evaluating a set of retrieved canditate components to find the most suitable one.
- Adapting the selected components to fit the spicific requirements.

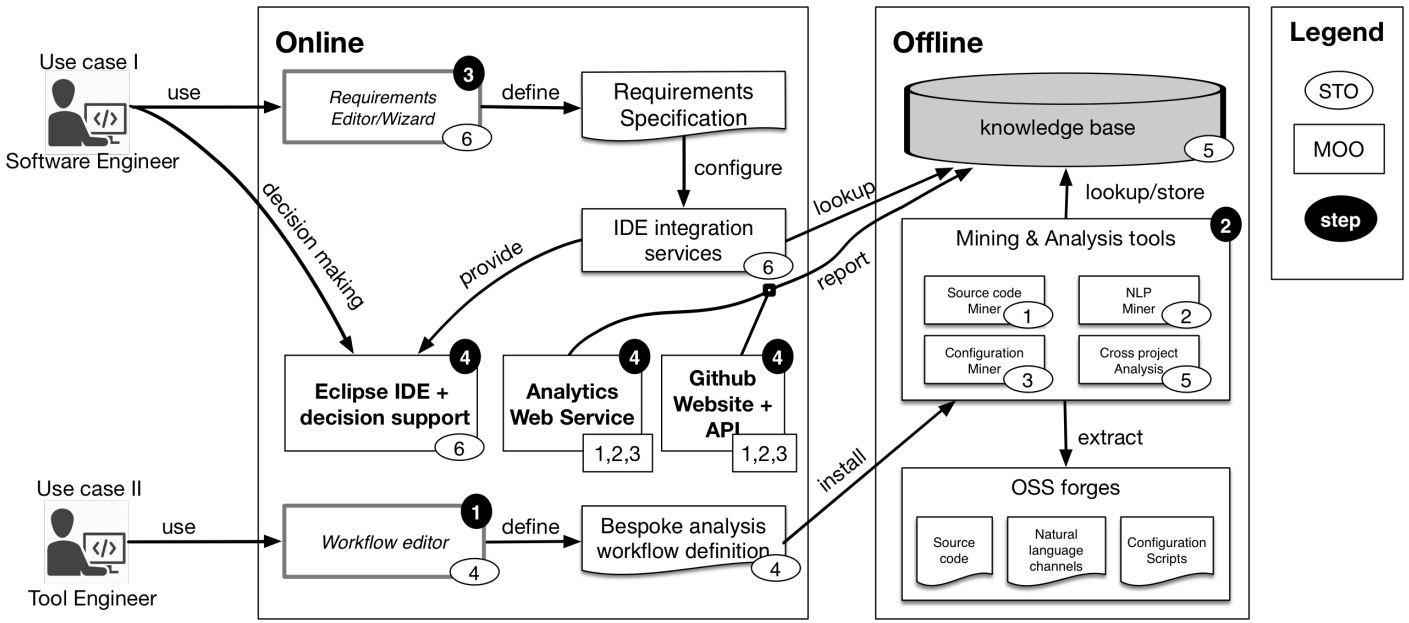
# Introduction

## CROSSMINER

- CROSSMINER aims at addressing such challenges by providing advanced techniques and tools supporting the identification and adoption of existing high-quality open source software components instead of implementing in-house propietary solutions with similar functionalities.

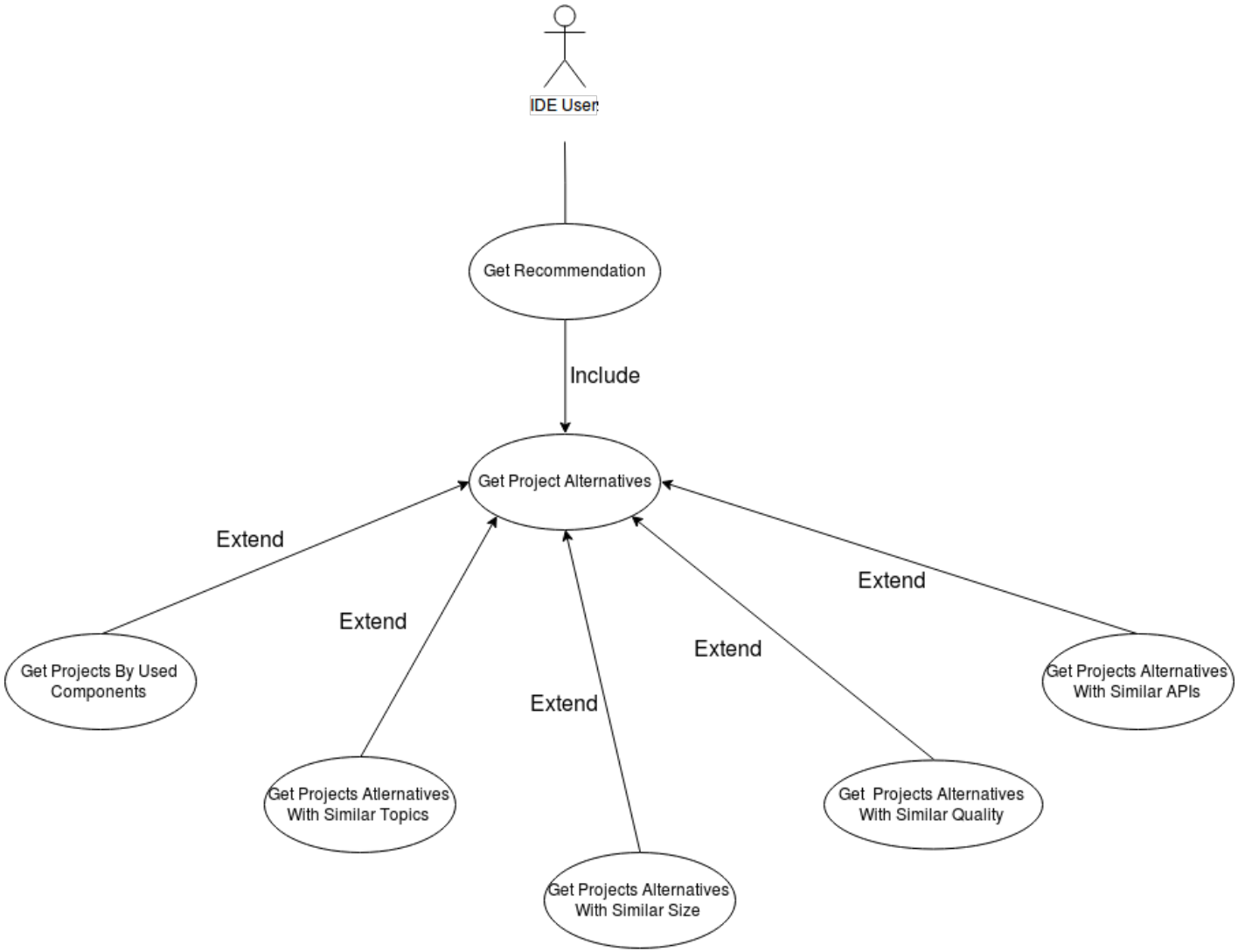
# Introduction

## CROSSMINER



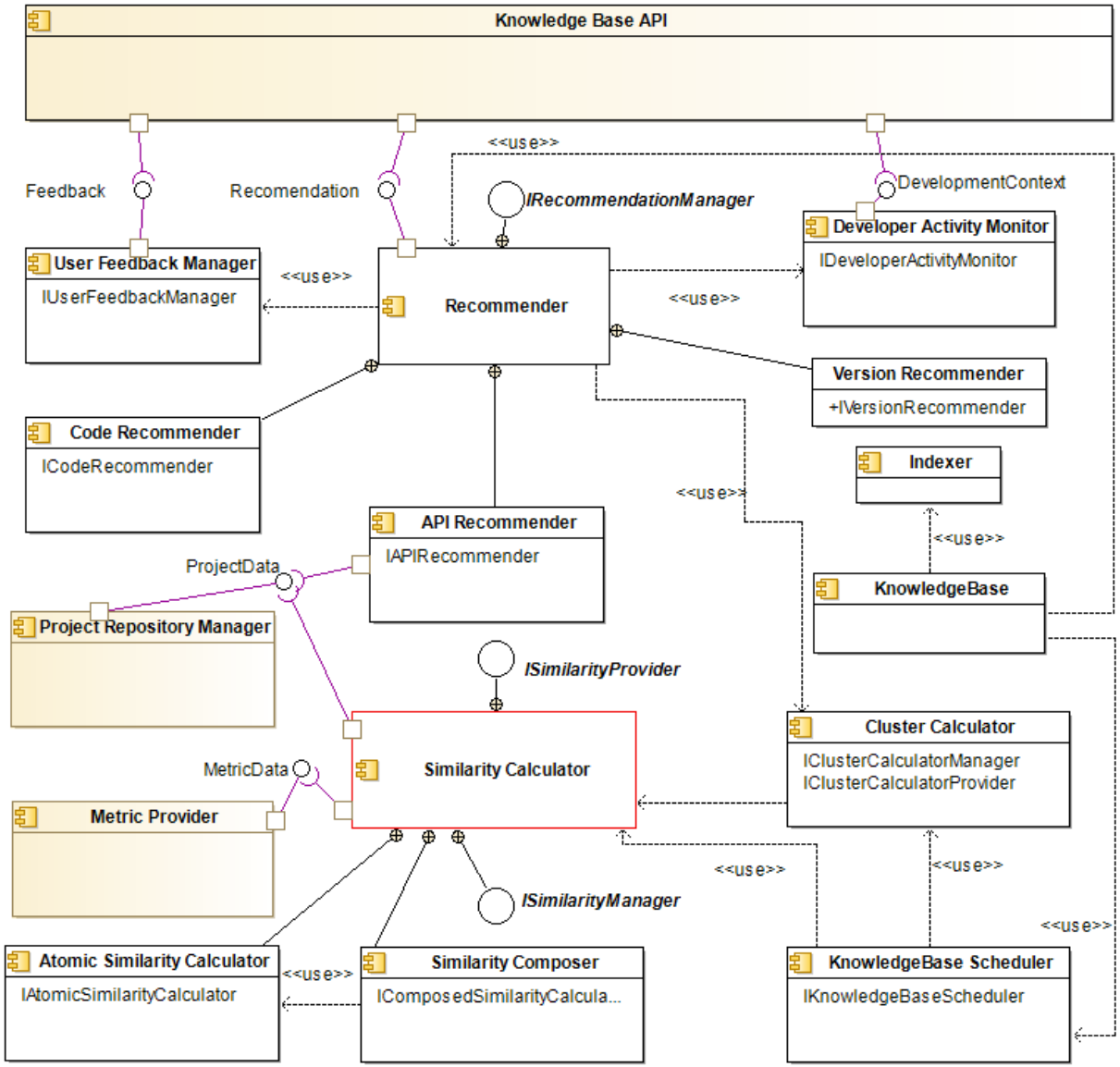
# Introduction

## CROSSMINER



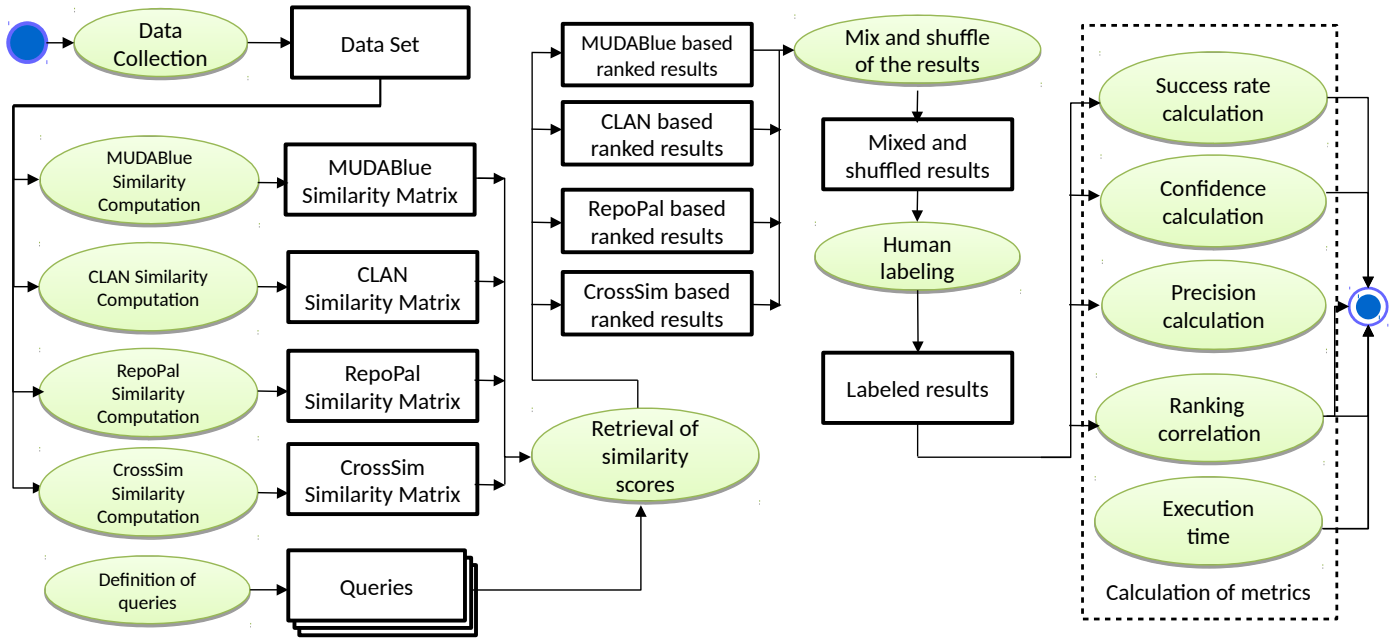
# Introduction

## CROSSMINER



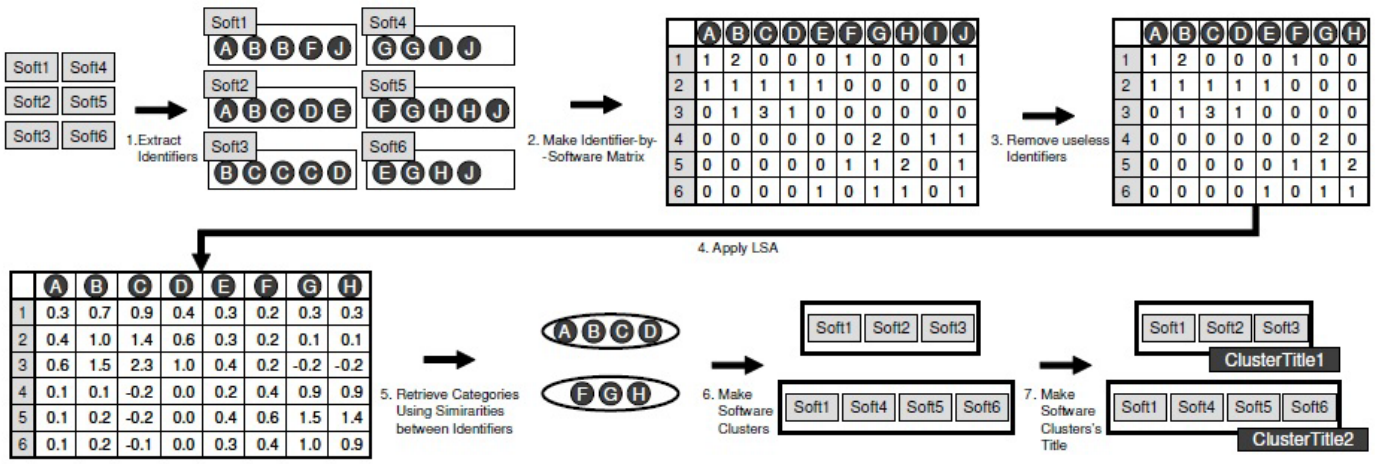
# Contribution

## Evaluation Process



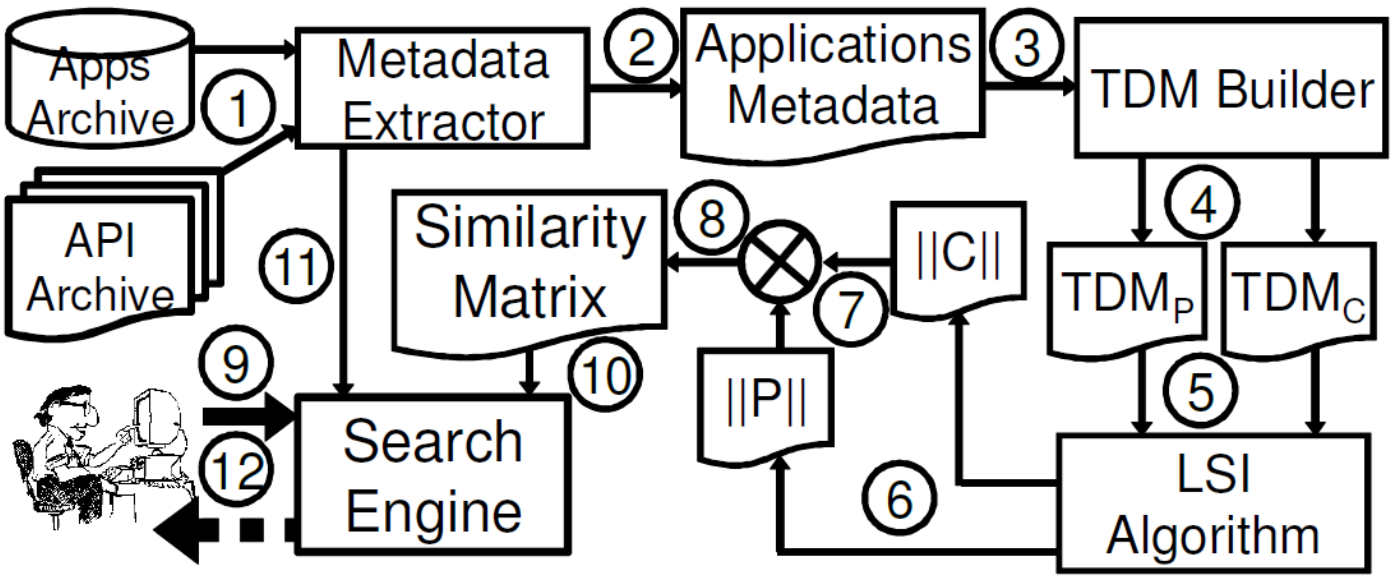
# Contribution

## MudaBlue



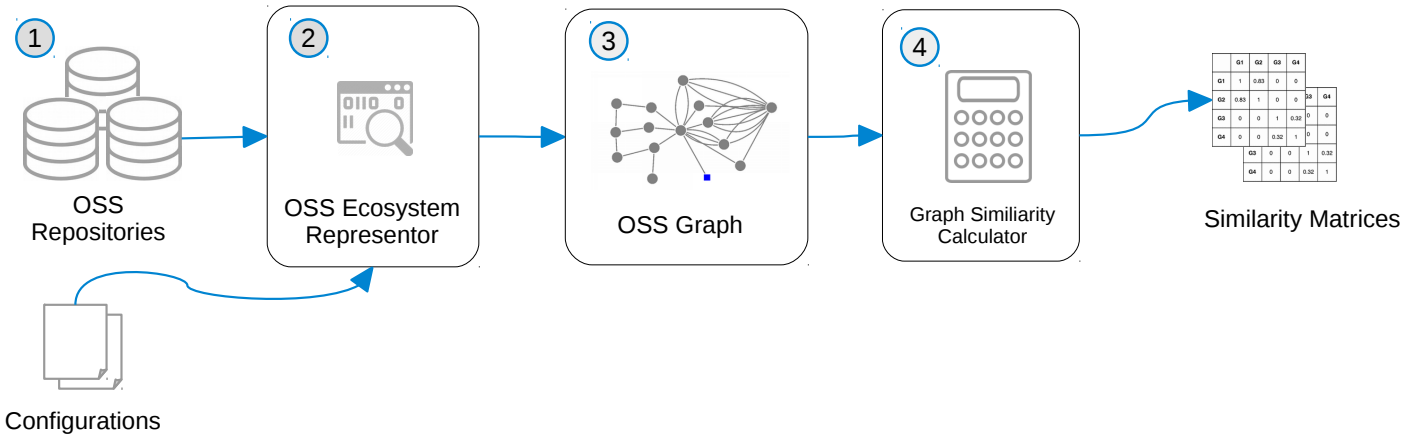
# Contribution

## CLAN



# Contribution

## CROSSSIM



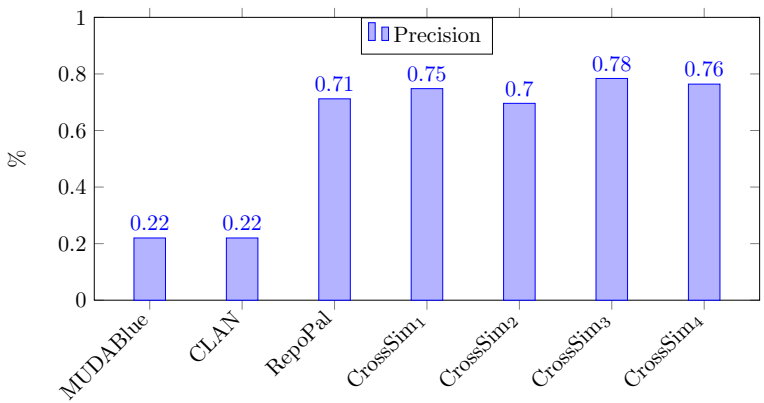
# Evaluation

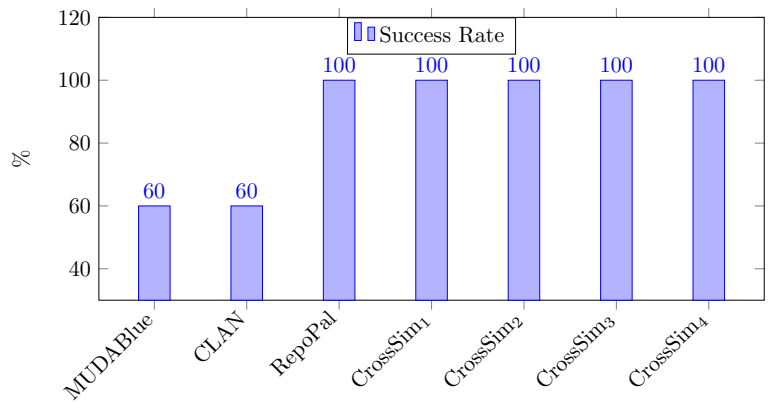
## User Study

- User study: Human evaluators label the similarity between query and retrieved projects
- User study: 10 people involved with experience plus a double check
- Similarity scales: *Dissimilar*, *Neutral*, *Similar*, and *Highly Similar*
- Evaluation metrics: Success Rate, Confidence, Precision

# Evaluation

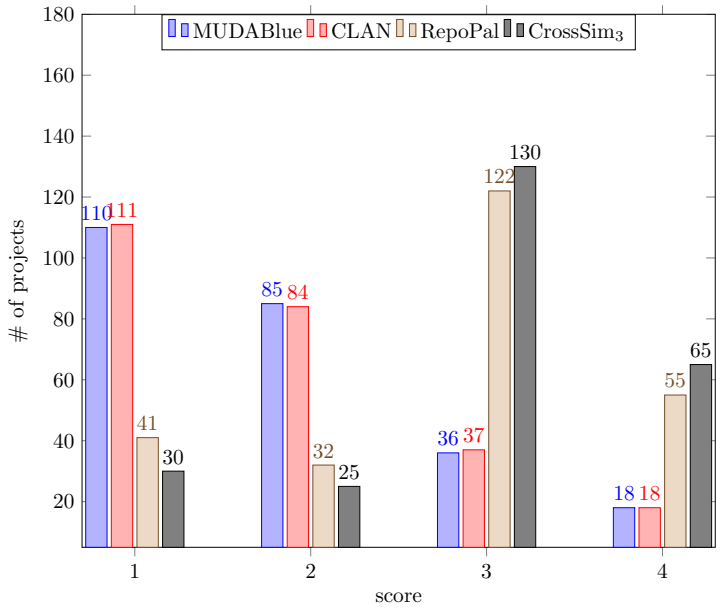
## Results

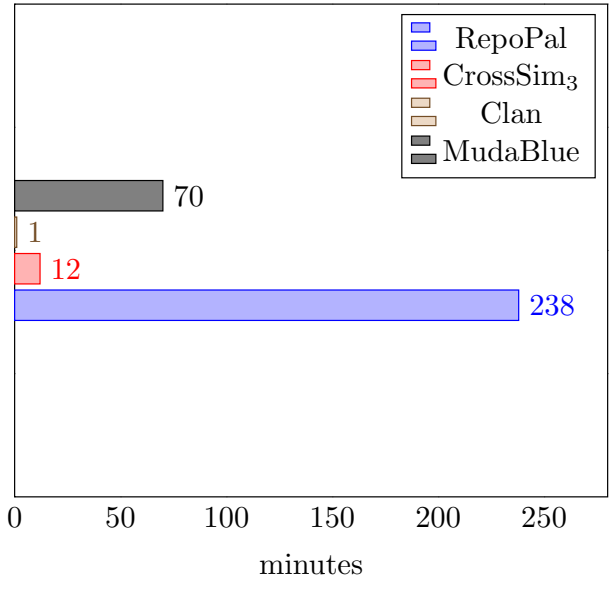


# Evaluation

## Results



# Conclusion

## What Has Been Done

- Implementation of two approaches
- Evaluating the results
- Confirmation of the goodness of CrossSim

# Conclusion

## What Else to be Done

- Integrate CrossSim inside the system
- Provide API recommandation
- Provide snippets of code