Mining Software Repositories

Wouter Poncin
Outline

• Introduction
• Problem description
• Existing approaches
• New approach
• Summary
Introduction

- Large projects
- Many developers
- Communication

Chapter 11: Ask for what you need

Introduction

- Many information sources
Introduction

- Process mining
Introduction

- Large projects
- Many developers
- Communication
- Many information sources
- Process mining

- “Who is working on this artifact?”
- “Were you allowed to work on this artifact?”
Problem description

What is the problem to be solved?

- How to find related developers (of some artifact) using information from software repositories?
- To enable process mining / analysis over multiple sources of information
- Literature study
Existing approaches
What has already been done?

• Work on ‘isolated systems’:
  • “How Developers Drive Software Evolution” [GKSD05]
  • “Mining Email Social Networks” [Bir06]
  • “Who are Source Code Contributors and How do they Change?” [Ger09]
  • “Mining CVS Repositories to Understand Open-Source Project Developer Roles” [Yu07]

• Work on `linking systems’:
  • “Developer identification methods for integrated data from various sources” [Rob05]
New approach
What is new in this project?

• Various information sources
• Extendable framework
  • Easily incorporate new information sources
• Enable process mining
  • Conformance checking
  • Performance analysis
  • Organization analysis
Issues
What are the issues to be solved?

- Various data sources
  - Data extraction
  - Data transformation
- Extendable framework
  - Multiple data ‘inputs’
  - Multiple analysis ‘outputs’
Summary

- Mining software repositories
- Related developers
- Process mining
- Literature study
Questions
References