Supporting Code Comprehension via Annotations: Right Information at the Right Time and Place

Marjan Adeli, Nicholas Nelson, Souti Chattopadhyay, Hayden Coffey, Austin Henley, and Anita Sarma

{adelima, nelsonni, chattops, anita.sarma}@oregonstate.edu
khcoffey1@vols.utk.edu, azh@utk.edu
Code Comprehension

- Code Comprehension:
  - Navigating through a codebase
  - Building a mental model of that code

- Large portion of developers' activities
  - 58% of the effort [1]

Difficulties in code comprehension

Understanding code requires developers to:

• Manage different types of artifacts

• Locate relevant information in different places

• Understand the relationships between artifacts
Facilitating code comprehension in IDEs

- Code Bubbles
- Code Canvas
- Debugger Canvas
- Synectic
Synectic: A canvas-based IDE
Annotation Overlay

- **Annotation notes** for capturing design rationale, expected API usage patterns, corner-cases, etc.

- **Annotation links** for connecting notes to cards or groups

- Multiway connections between annotations and cards to describe relationships.
Study Design - RQs

- RQ1: How do annotations affect code comprehension among newcomers?
  - Do annotations increase the **accuracy** of responses?
  - Do annotations reduce the **time to task completion**?
  - Do annotations reduce **cognitive load**?
User Study

Controlled lab study
- Between-subject design
- 22 participants (graduate students)
- 4 code comprehension tasks

Synectic treatment
- 11 participants
- 4 code comprehension tasks
- Onboarding information added as annotations

Eclipse treatment
- 11 participants
- 4 code comprehension tasks
- Onboarding information added as text document
Study Design – Tasks

• Code Comprehension Task
  • Navigation portion
  • Comprehension portion

• Designed as onboarding tasks
  • Locating code related to a feature
  • Learning how to make changes to those features
Results

🎯 Accuracy

Rank Based Non-Parametric (RBNP) ANOVA test
\( p\)-value < 0.001, statistic = 19.46488

⏰ Time

RBNP ANOVA test
\( p\)-value = 0.22, statistic = 1.607723

🧠 Cognitive Load

RBNP ANOVA test
\( p\)-value = 0.003, statistic = 11.52591
Discussion

• Quantitative results
  • Accuracy & Cognitive Load differences were significant
  • Time differences were not statistically significant

• Qualitative results
  • Sillito et al.’s four stages of comprehension model\(^1\) to explain comprehension
  • Information Foraging Theory (IFT) to explain navigation

Discussion – Stages of comprehension

Sillito et al. identified 4 categories of comprehension:

1. Finding the initial focus point
2. Building on those focus points
3. Understanding the concepts between related entities
4. Understanding concepts across multiple groups of related entities
Discussion – Stages of comprehension

1. Finding the initial focus point
Discussion – Stages of comprehension

2. Building on those focus points
3. Understanding the concepts between related entities
Discussion – Stages of comprehension

4. Understanding concepts across multiple groups of related entities
Summary

• Annotations in a canvas-based IDE resulted in:
  • Lower cognitive load among newcomers
  • More accurate comprehension responses
  • Required no additional time compared traditional IDEs

• Design challenges for annotations within IDEs:
  • Manage different types of artifacts
  • Locate relevant information in different places
  • Understand the relationships between artifacts

“Right information, at the right place, and the right time”