

Code Understanding With LLM VS Code Extension

Ekistoclecio heleno Duarte de Lima
Ian Karlo Torres dos Santos
Flávio José Canuto Vasconcelos Júnior
Heitor Brayner Prado



Centro de
Informática
UFPE

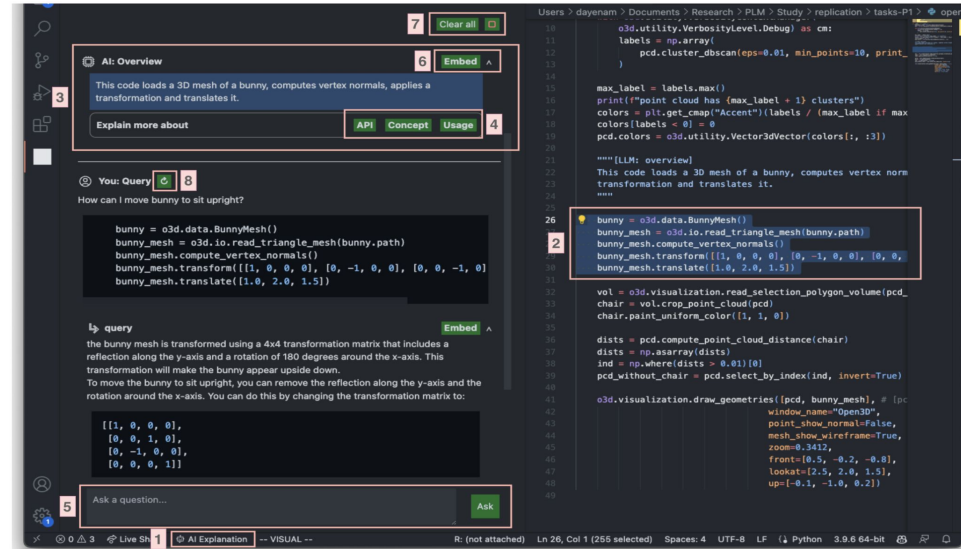


UNIVERSIDADE
FEDERAL
DE PERNAMBUCO

Relembrando...

Artigo: [Using an LLM to help with Code Understanding](#)

Objetivo: Simular um processo tradicional de engenharia de software



Relembrando... (Críticas)

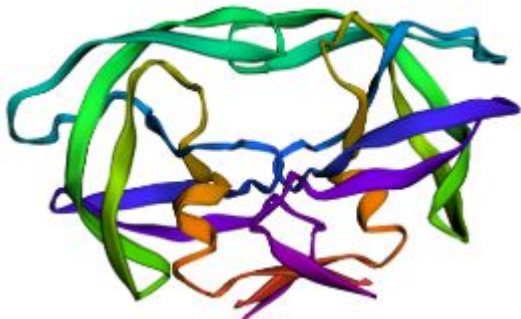
- Confusão entre Eficiência e Compreensão
 - Entendimento X Tempo de conclusão
 - Variáveis Confundidoras
 - Fatores psicológicos
- **RQ1:** Até que ponto o GILT afeta a compreensão dos desenvolvedores, o tempo de conclusão das tarefas e as taxas de conclusão das tarefas ao lidar com código desconhecido?

Nosso Objetivo

- Desenvolver e avaliar uma extensão para o VSCode que utiliza LLMs para ajudar os desenvolvedores a entenderem melhor o código
- Foco especial na explicação de métodos e funcionalidades dentro de bibliotecas de código.
- A ferramenta deve fornecer explicações claras sobre o que cada método faz e como ele pode ser utilizado, com base no contexto do código.

Bibliotecas Escolhidas

- **Py3Dmol** - é um biblioteca python para visualização molecular livre de dependências em notebooks iPython.
- **Gensim** - biblioteca Python para modelagem de tópicos, indexação de documentos e recuperação de similaridade com grandes corpora. O público-alvo é a comunidade de processamento de linguagem natural (NLP) e recuperação de informação (IR).



Metodologia

- Foram escolhidos dois repositórios com bibliotecas de código diferentes para testar a eficácia da extensão na compreensão de código.
 - Critérios de Seleção:
 - Popularidade das Bibliotecas
 - Diversidade dos Métodos
 - Documentação Limitada
- Foram realizadas entrevistas com dois desenvolvedores sem conhecimento prévio das bibliotecas aplicadas
- Foram aplicados dois questionários, cada um com sete questões, para analisar as respostas enviadas pelos desenvolvedores utilizando a ferramenta.

Metodologia

Avaliação das Respostas

Cada resposta foi classificada com base no nível de conhecimento demonstrado:

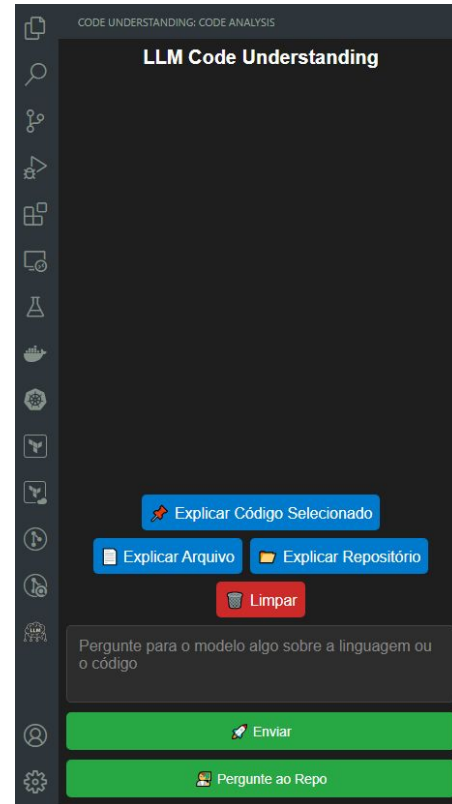
- Nenhum conhecimento
- Pouco conhecimento
- Médio conhecimento
- Bom conhecimento
- Alto conhecimento

Cálculo de Desempenho

- Para cada questionário:
 - Foi calculada a pontuação média de conhecimento
- Análise comparativa entre:
 - Uso da extensão VSCode
 - Uso de pesquisa web

Extensão VSCode

- Conexão com a API do Gemini 2.0 Flash
- Usa o contexto do repositório aberto no momento.
- Funcionalidades: Explicar código selecionado, explicar arquivo, explicar repositório, perguntar usando o repositório como contexto e fazer uma pergunta livre



Extensão VSCode

- **Explicar código selecionado:** Usuário pode selecionar uma parte do código e solicitar uma explicação da LLM.

```
'Por favor, forneça uma análise abrangente deste trecho de código. Sua explicação deve incluir:
```

- ```
1. Propósito geral e funcionalidade do trecho
2. Funções importantes, variáveis e seus papéis
```

```
Não repita nenhum código em sua explicação.
```

```
Caso você não encontre alguma das regras acima, não adicione nada.
```

```
Escreva o texto de forma mais direta, sem dividir muito em tópicos.
```

```
Aqui está o trecho de código:
```

```
${code}
```

```
A resposta da explicação deve estar em formato markdown e em PORTUGUÊS.
```

```
,
```

# Extensão VSCode

- **Explicar arquivo:** Permite ao usuário enviar o arquivo aberto no VSCode para obter uma explicação detalhada do conteúdo.

```
'Por favor, forneça uma análise abrangente deste arquivo. Sua explicação deve incluir:
1. Propósito geral e funcionalidade do trecho
2. Funções importantes, variáveis e seus papéis
```

```

Não repita nenhum código em sua explicação.
Caso você não encontre alguma das regras acima, não adicione nada.
Escreva o texto de forma mais direta, sem dividir muito em tópicos.
```

```
Aqui está o conteúdo do arquivo:
```

```
${fileContent}
```

```
A resposta da explicação deve estar em formato markdown e em PORTUGUÊS.
,
```

# Extensão VSCode

- **Explicar repositório:** Envia todos os arquivos .ts, .js, .json, .py, .java e .md do repositório aberto no VSCode para uma explicação abrangente.

```
'Por favor, faça uma análise abrangente desta base de código com múltiplos arquivos. Cada arquivo está marcado com '## File X' no início e '#endfile' no final.
```

```
Forneça uma explicação detalhada que inclua:
```

- ```
1. Propósito geral e funcionalidade do trecho  
2. Funções importantes, variáveis e seus papéis
```

```
Não inclua nenhum trecho de código em sua explicação.
```

```
Caso você não encontre alguma das regras acima, não adicione nada.
```

```
Escreva o texto de forma mais direta, sem dividir muito em tópicos.
```

```
Aqui estão os arquivos:
```

```
${filesContent}
```

```
A resposta da explicação deve estar em formato markdown e em PORTUGUÊS.
```

```
`
```

Extensão VSCode

- **Perguntar usando o repositório como contexto:** Permite ao usuário fazer perguntas utilizando o repositório aberto no VSCode como base de contexto.

```
'Por favor, analise de forma abrangente a seguinte base de código e responda à pergunta com base em seu conteúdo. Cada arquivo na base de código está marcado com '## File X' no início e '#endfile' no final.  
  
Primeiro, entenda a base de código:  
1. Arquitetura geral e propósito  
2. Componentes principais e seus relacionamentos  
3. Funcionalidades e características importantes  
  
Em seguida, forneça uma resposta detalhada à pergunta, referenciando especificamente partes relevantes da base de código.  
A resposta deve estar em formato markdown e em português.  
Caso você não encontre alguma das regras acima, não adicione nada.  
  
Base de código:  
${filesContent}  
  
Pergunta: ${question}  
  
Por favor, forneça uma resposta abrangente que aborde diretamente a pergunta enquanto aproveita seu entendimento da base de código fornecida.'
```

Extensão VSCode

- **Perguntar livre:** O usuário pode fazer qualquer pergunta à LLM, sem utilizar o repositório ou arquivos como contexto.

```
'Por favor, forneça uma resposta clara e detalhada para a seguinte pergunta. Se a pergunta for sobre código ou conceitos técnicos, inclua exemplos relevantes e explicações.  
Pergunta: ${question}'
```

Resultados da pesquisa - Gensim

Questão	Com Extensão	Sem Extensão
Q1	Bom conhecimento	Pouco conhecimento
Q2	Bom conhecimento	Pouco conhecimento
Q3	Bom conhecimento	Nenhum conhecimento
Q4	Bom conhecimento	Pouco conhecimento
Q5	Bom conhecimento	Nenhum conhecimento
Q6	Bom conhecimento	Nenhum conhecimento
Q7	Bom conhecimento	Pouco conhecimento
Média Geral	Bom conhecimento	Pouco conhecimento

Resultados da pesquisa - py3Dmol

Questão	Com Extensão	Sem Extensão
Q1	Alto conhecimento	Bom conhecimento
Q2	Médio conhecimento	Nenhum conhecimento
Q3	Bom conhecimento	Médio conhecimento
Q4	Bom conhecimento	Médio conhecimento
Q5	Alto conhecimento	Médio conhecimento
Q6	Bom conhecimento	Nenhum conhecimento
Q7	Bom conhecimento	Pouco conhecimento
Média Geral	Bom conhecimento	Médio conhecimento

Melhorias

- Utilização de um banco vetorial para salvar informações do repositório.
- Adição do histórico das conversas (contexto) nas chamadas do Llm (funções de perguntar ao modelo, pergunta livre).
- Refinar os filtros na busca de arquivos para evitar sobrecarga de contexto.
- Melhoria da formatação do texto de resposta.
- Adição de opções para deixar as respostas mais curtas e diretas ou mais longas e explicadas.
- Utilização de outros modelos (seleção de modelos).