

O Código que os Desenvolvedores Escreveram e o Código que Poderiam Escrever: Uma Perspectiva sobre Identificadores Sugeridos por LLMs

Julyanara Rodrigues Silva¹; marcelo de almeida maia; Carlos Eduardo de Carvalho Dantas²

¹Estudante, IFTM *Campus* Uberlândia Centro, MG, bolsista IFTM, julyanara.silva@estudante.iftm.edu.br

²Professor(a), IFTM *Campus* Uberlândia Centro, MG, Dr. Ciência da Computação, carlooseduardodantas@iftm.edu.br

A nomeação apropriada de identificadores é crucial no código-fonte, já que os nomes podem representar até 70% de todos os caracteres e desempenham um papel fundamental ao ajudar desenvolvedores a entenderem a intenção por trás do código. A renomeação também está entre as operações de refatoração mais comuns realizadas por desenvolvedores. Embora ferramentas de análise estática possam detectar violações de convenções de nomenclatura, ainda há uma lacuna em ferramentas capazes de sugerir nomes semanticamente alinhados com o propósito do código. Neste estudo, foram coletadas 152 instâncias em Java e 168 em Python de pull requests (PRs) no GitHub, onde desenvolvedores melhoraram nomes de identificadores. Foi investigado se modelos de linguagem de grande escala (LLMs) são capazes de sugerir nomes semanticamente consistentes com aqueles escolhidos por desenvolvedores. Também foram comparadas três configurações de temperatura, variando de determinística a criativa, para avaliar como esse parâmetro influencia as sugestões de nomes. Os resultados mostram que o ChatGPT teve um desempenho ligeiramente superior aos outros modelos para identificadores em Python, correspondendo exatamente aos nomes escolhidos pelos desenvolvedores em 16,2% dos casos e alcançando alta similaridade semântica em 65,1%. De modo geral, os três LLMs tiveram melhor desempenho em Python do que em Java. Em relação à configuração de temperatura, a temperatura moderada apresentou resultados ligeiramente melhores em Python, enquanto a configuração criativa teve desempenho um pouco superior em Java. Esses achados sugerem que tarefas como nomeação de identificadores podem se beneficiar de configurações de parâmetros diferentes daquelas geralmente utilizadas em tarefas mais determinísticas, como refinamento de código.

Palavras-chave: chatgpt. deepseek. gemini. github. identificadores. LLM.

Apoio: IFTM.