# 1. INTRODUÇÃO

# 1.1 Contextualização

O acompanhamento de alunos em turmas numerosas de disciplinas voltadas ao aprendizado de técnicas de programação é uma dificuldade encontrada pelos professores e/ou monitores, que buscam usar a tecnologia para facilitar o gerenciamento de turmas grandes.

Um problema encontrado em laboratórios de programação é a cópia total ou parcial de soluções entre colegas. Em cenários de turmas numerosas, a detecção deste tipo de conduta se torna bastante complicada. A análise de similaridade em códigos-fonte é objetivo de estudo de diversos trabalhos e algumas ferramentas foram desenvolvidas para auxiliar na avaliação.

# 1.2 Motivação

Apesar da grande quantidade de ferramentas disponíveis, poucas delas são capazes de identificar de maneira eficaz todas as semelhanças léxicas e semânticas entre pares de códigos.

# 1.3 Objetivos

Tendo em vista as dificuldades apresentadas por essas ferramentas, esse trabalho propõe uma ferramenta que extrai as características-chave únicas da linguagem e o processamento dessas para criação de uma métrica que define o nível de similaridade entre dois códigos-fonte e a comparação dos resultados encontrados com as outras ferramentas de análise de similaridade em códigos-fonte.

Os objetivos específicos são:

1. Efetuar um estudo teórico sobre as principais métricas de software e técnicas de detecção de plágio em código-fonte e estudar os resultados dos principais algoritmos relacionados.

ii. Analisar técnicas de normalização de códigos-fonte que melhorem o desempenho dos algoritmos de análise sintática no contexto de analise de similaridade.

1. Introduzir métricas de complexidade a ferramenta proposta.

# 1.4 Estrutura do documento

O trabalho está organizado em cinco capítulos. O presente capítulo descreve uma breve introdução ao tema, contextualizando o assunto abordado, a motivação, os objetivos e as contribuições deste trabalho.

No Capítulo 2 são abordados os principais conceitos teóricos necessários para compreensão deste trabalho. Entre eles estão os principais conceitos relativos a técnicas de plágio, ao funcionamento de algoritmos e de ferramentas para análise de similaridade e detecção de plágio, as técnicas de normalização e as métricas de complexidade.

No Capítulo 3 apresenta a ferramenta desenvolvida, bem como o método de conformidade que foi desenvolvido para a análise dos resultados de múltiplas ferramentas.

No Capítulo 4, apresentam-se os resultados obtidos com códigos gerados manualmente. Por fim, no Capítulo 5 são realizadas as considerações finais, bem como as possíveis melhorias a serem consideradas em trabalhos futuros.