

Duplicitný kód sa vyskytuje v zdrojových súboroch z rôznych dôvodov. V mnohých prípadoch je motiváciou pre kopírovanie kódu lenivosť programátora, alebo pokus o použitie cudzieho zdrojového kódu. V priebehu rokov bolo vyvinutých niekoľko metód na detekciu duplicitného zdrojového kódu. Prístupy k problému sa líšia v spôsobe akým porovnávajú zdrojové súbory. Metódy založené na analýze syntaktických vlastností kódu často používajú abstraktné syntaktické stromy. Skúmaním stromovej reprezentácie sú tieto metódy schopné odhaliť duplicitný kód, ktorý prešiel zmenami formátovania ako aj premenovaním identifikátorov. Fragmenty duplicitného kódu sú odhalené identifikovaním podstromov s rovnakým tvarom. Po nájdení podozrivých častí stromu nasleduje hlbšia analýza vlastností stromu za účelom odhalenia duplikátov. Ako súčasť tejto práce sme vyvinuli nástroj na detekciu duplicitného zdrojového kódu na základe porovnávania abstraktných syntaktických stromov.