

Podobnosti v zdrojovom kóde

Marek Roštár
Vedúci: Michal Kompan

Cieľ práce

- Navrhnuť metódu odhaľovania podobnosti v zdrojových kódach
- Dôraz na podporu viacerých programovacích jazykov
- Motivácia: odhaľovanie potenciálnych plagiátorov, zlepšenie kvality kódu

Rôzne plagiátoské útoky

- Pridanie bielych znakov
- Pridanie komentárov
- Zmena poradia funkcií
- Premenovanie premenných
- Úprava algebraických výrazov
- Vytiahnutie funkcií do iného súboru

Rôzne metódy detekcie

- Text/token
- Abstraktný syntaktický strom
- Control flow graph

Návrh riešenia

- Preprocessing + Abstraktný syntaktický strom
- Porovnanie rôznych úrovní preprocessingu
- V ruby
- Testovacie dáta pravdepodobne zo školy

Problémy

- Mapovanie rôznych jazykov
- Časova nefektivita na krátkých zdrojových kódoch voči token/text zameranému prístupu