CloneDetector

Program pentru depistarea proiectelor Java copiate

Adrian-Vlad LEP, Coordonator: conf.dr.ing. Ioana SORA

Universitatea Politehnica Timisoara

Facultatea de Automatica si Calculatoare

Calculatoare si Tehnologia Informatiei

Iunie 2011

Există diverse platforme ce detectează lucrări literare, proiecte şcolare sau progrme software copiate. Programul CloneDetector analizează static codul sursă şi depistează proiectele Java copiate pe baza unui set de reguli Prolog. CloneDetector foloseşte o abordare originală, combinând metodele de depistare bazate pe metrici şi cele bazate pe grafuri de apeluri (call graph) create din arbori sintactici abstracţi(AST). Prin acest instrument se doreşte depistarea clonelor de tipul 3, adică a proiectelor copiate în care pe lângă redenumire s-au operat schimbări de natură sintactică, s-au adăugat sau şters porţiuni de cod.

Sistemul a fost dezvoltat pentru depistarea proiectelor scolare Java copiate. Punctul de plecare va fi codul sursă predat. Figura 1 descrie paşii parcurşi.1. Arhitectura sistemului
 Pentru compararea a două proiecte, din codul Java se va genera cu ajutorul plugin-ului Eclipse JTransformer arborele sintactic abstract (AST) al fiecărui proiect. AST-ul va fi format din clauze Prolog şi memorat sub formă unor fişiere qlf (quick load file), fişiere compilate Prolog.

Analiză a două proiecte începe la nivel de clase. Încercăm să găsim pentru fiecare clasa din primul proiect o clasa echivalentă în cel de al doilea. Două clase vor fi considerate identice dacă vor avea similarităţi la nivel de clasa, dacă metodele lor vor fi similare şi dacă clasele care vor fi apelate din metodele celor două clase vor fi asemănătoare.

Din testele realizate am observat ca analiza grafului de apeluri aduce o imbunatatire majora, crescand numarul de imperecheri corecte la nivel de metode, implicit la nivel de clase si proiecte.

Pentru a satisface nevoile profesorilor cat mai bine si a usura munca utilizatorilor, au fost dezvoltate diverse facilitati, fara a afecta scalabilitatea algoritmului. CloneDetector permite excluderea unor clase, pachete sau foldere din analiza. Aceasta se dovedeste foarte utila in cazul in care toate proiectele au portiuni de cod date de la inceput.

Pentru a putea analiza diferit in functie de tema data, am creat doua profile de analiza unul larg si al doilea mai exact. Aceste profile pot fi selectate la rulare sau noi profile pot fi usor definite in fisierul profiles.pl.

De asemenea, datorita metodei de implementare se pot adauga sau sterge relativ usor filtre noi in functie de novoile ulterior aparute.

In concluzie CloneDetector se doreste a fi un tool usor de folosit si eficient in depistarea proiectelor software copiate. In urma testelor realizate asupra unor proiecte concrete, suntem increzatori in utilitatea acestei platforme. S-a obsevat ca programele ce folosesc algoritmi de comparare a sirurilor de caractere, desi foarte rapide, pot depista doar clonele de tipul 1, eventual 2 si 3 daca redenumirile nu sunt complete. Instrumentele ce se bazeaza pe grafuri de apeluri pot semnala toate tipurile de clone dar datorita algoritmilor complexi, au o durata de procesare mai mare. CloneDetector poate depista toate clonele de tipul 1 si 2 indiferent de numarul de redenumiri efectuate, semnaland si clone de tipul 3 cu probabilitatea lor de a fi copiate.