

**EXAMENSARBETE** Measuring Semantic Distances between Software Artifacts to Consolidate

Issues from the Development and the Field

**STUDENT** Mahmoud Nasser**HANDLEDARE** Elizabeth Bjarnason (LTH), Markus Borg (SICS)**EXAMINATOR** Per Runeson (LTH)

# Mätning av semantiska avstånd mellan mjukvaruartefakter med användandet av ett sökmotorbaserat verktyg

## POPULÄRVETENSKAPLIG SAMMANFATTNING **Mahmoud Nasser**

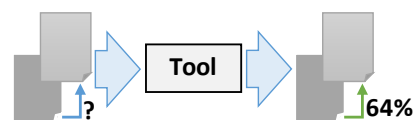
Regelbundna manuella granskningar för att identifiera länkar mellan ständigt tillkommande mjukvaruartefakter är viktigt för att hålla ordning och reducera kostnaden för ett mjukvaruprojekt. Tidskrävande är det också; därför utreder detta arbete hur ett verktyg kan åtgärda detta med en snabbare och effektivare granskningsprocess.

Olika mjukvaruartefakter, dokument som beskriver mjukvarufunktioner, ser ibland väldigt olika ut även om de behandlar samma del i mjukvaran. En funktion måste nämligen beskrivas på olika sätt för olika sammanhang; vanlig naturlig text för att kund och utvecklare ska vara överens om funktionsutformning, programkod eller testfall i tekniska respektive testningssammanhang.

Olika artefakter som beskriver en funktion har en länk mellan sig eftersom de är lika i funktion och hör ihop. Ju mindre lika två artefakter är i detta avseende, desto större *semantiskt avstånd* de sägs ha. Dessa länkar kan förloras och förändras allteftersom mjukvaran ändrar skepnad. Om man inte underhåller denna data så blir det problem senare när man t.ex. upptäcker en bugg, då man i sammanhang av programkod eller testfall sällan förstår vad som går fel utan att hitta relevanta artefakter.

Detta brukar hanteras genom regelbundna granskningsmöten där artefakter som förändrats gås igenom och man lägger till, tar bort och länkar om mellan olika artefaktpar utifrån frågan om de beskriver samma funktion. Granskningarna, som utförs manuellt, är tidskrävande och kostsamma.

Vi har med vårt arbete utvecklat ett verktyg som använder en sökmotor för att indexera två olika men besläktade defektartefakttyper. Vi definierar ett sorts mått för semantiska avstånd och matar in en defektartefakt i verktyget för att få förslag på alla andra artefakter som är semantiskt nära. Verktyget analyserar olika sorters länkar och defektdata för att skapa ett helhetsmått. Träffsäkerheten evalueras genom att testa verktyget med förberedda sökningar som facit.



Verktyget sätter ett mått på semantiska avstånd.

Våra resultat visar att för vissa syften blir verktyget väldigt träffsäkert, medan det i bästa fall blir måttlig träffsäkerhet för flera andra syften. Vi har kommit fram till att detta osäkra och blandade resultat beror på den begränsade definitionen i vad för sorts defektdata man tar hänsyn till. Oavsett detta, snabbar verktyget upp granskningen, och i bästa fall, särskilt ifall det utvecklas vidare, kan det hitta fler komplexa och användbara länkar.