

nodigt U graag uit op de openbare verdediging van het proefschrift van

## Camilo Velázquez-Rodríguez

ter behaling van de graad van Doctor in de Wetenschappen

Titel van het proefschrift:

Extracting Library Features from Incomplete  
Code on Stack Overflow

Promotor:  
Prof. dr. Coen De Roover

De verdediging heeft plaats op

Vrijdag 29 maart 2024 om 17u in aula I.2.02

De verdediging kan ook online gevolgd worden  
via: <https://msteams.link/J84H>

Samenstelling van de jury

Prof. dr. Viviane Jonckers (VUB, voorzitter)  
Prof. dr. Wolfgang De Meuter (VUB, secretaris)  
Prof. dr. Ann Nowé (VUB)  
Prof. dr. Sam Verboven (VUB)  
Prof. dr. Davide Di Ruscio, Università degli Studi  
dell'Aquila, Italië)  
Prof. dr. Bin Lin, Radboud Universiteit, Nederland)

### Curriculum vitae

Camilo Velázquez-Rodríguez behaalde zijn M.Sc. in Toegepaste Wiskunde en Informatica voor Administratie aan de Universidad de Holguín in 2016. Hij begon zijn PhD aan het Software Languages Lab (SOFT) in 2018, ondersteund door het Excellence of Science Research Project SECO-ASSIST.

Zijn onderzoek richtte zich op het ontwerpen van technieken en hulpmiddelen om ontwikkelaars te helpen bij het selecteren van een software bibliotheek in een groot software-ecosysteem door het extraheren van functies. Camilo's onderzoek resulteerde in vijf publicaties in internationale peer-reviewed tijdschriften en conferenties, waarvan er één werd geëerd met een distinguished paper award. Camilo presenteerde zijn onderzoek op internationale conferenties en workshops. Daarnaast heeft hij diverse bachelor- en masterscripties mede begeleid.

### Abstract van het doctoraatsonderzoek

In de hedendaagse softwareontwikkeling is het gebruikelijk om features van externe bibliotheken te hergebruiken. Features hergebruiken, in plaats van ze helemaal opnieuw te implementeren, kan de ontwikkeltijd verkorten en de algemene systeemkwaliteit verbeteren. Het selecteren van een geschikte bibliotheek om features uit te hergebruiken kan echter moeilijk zijn voor ontwikkelaars vanwege het gebrek aan geautomatiseerde toolondersteuning. Sommige bibliotheekindices stellen ranglijsten van bibliotheken voor, maar deze zijn bevooroordeeld op het aantal bibliotheekdownloads, toegekende sterren, enz. Dit kan ertoe leiden dat ontwikkelaars de populairste bibliotheek selecteren in plaats van de bibliotheek met de feature die ze hoopten te hergebruiken.

Deze thesis levert drie bijdragen. De eerste, genaamd LiFUSO, is een geautomatiseerde benadering voor het opsommen en beschrijven van de features die door een bibliotheek worden geboden, op basis van openbaar beschikbare informatie op sociale coderingsplatformen zoals Stack Overflow (SO) en GitHub. LiFUSO analyseert bibliotheekgebruik binnen SO-berichten en extraheert gebruikspatronen die indicatief zijn voor bibliotheekfeatures. Daartoe wordt zowel rekening gehouden met de codefragmenten als met de omringende natuurlijke taal binnen elk SO-bericht waarin de bibliotheek wordt besproken. RESICO, de tweede bijdrage van de thesis, is een geautomatiseerde aanpak voor het omzetten van API-typerefenties binnen een codefragment naar hun overeenkomstige, volledig gekwalificeerde naam. Als een op leren gebaseerde benadering van tekstclassificatie moet RESICO worden getraind op een corpus van programma's waarvoor een compiler de juiste type-informatie heeft bepaald. Eenmaal getraind kan het syntactisch onjuiste codefragmenten als invoer gebruiken. De laatste bijdrage combineert LiFUSO en RESICO zodat de eerste niet langer beperkt is tot SO-berichten die zijn getagd met de naam van een bibliotheek, maar ook informatie kan extraheren uit berichten waarin veel bibliotheektypen zijn herkend. We evalueren de impact van het verbreden van de reikwijdte van de analyse op de kwantiteit en kwaliteit van de ontdekte bibliotheekfeatures.

De bijdragen zijn relevant voor de software-engineeringgemeenschap als geheel. De typeresolutie van RESICO kan worden overgenomen door tools die potentieel onvolledige codefragmenten moeten analyseren, of door tools die moeten bepalen welke bibliotheken binnen een codefragment worden gebruikt ---net als in onze derde bijdrage. LiFUSO maakt de weg vrij voor toolondersteuning bij het selecteren van een bibliotheek uit vele alternatieven. Ten slotte helpen deze bijdragen bij het begrijpen van de voor- en nadelen van datagerichte dan wel algoritmische oplossingen voor software-engineeringproblemen.