

**Informatics Institute of Technology**

In collaboration with

**University of Westminster, UK**

**“InteRec”**

**INTEGRATOR RECOMMENDATION FOR PULL-BASED DISTRIBUTED  
SOFTWARE DEVELOPMENT**

A Dissertation By

**Mr Raveen Savinda Rathnayake**

(IIT ID: 2015358 | UOW ID: W1610086)

Supervised By

**Mr Guhanathan Poravi**

Submitted in partial fulfilment of the requirements for the

BEng. (Hons) in Software Engineering Degree

Department of Computing

**April 2019**

© The copyright for this project and all its associated products resides with

Informatics Institute of Technology.

## Abstract

Over the last decade, open source software development, dramatically changed with the introduction of the pull-based development mechanism. The pull-based method became popular because of the convenience for collaborative contributions. With the introduction of pull-based development in the distributed software development, the number of contributions received for projects has risen as this method lowered the entry barrier for contributions. Nowadays social coding platforms such as GitHub, GitLab and Bitbucket, integrated this pull-based model in their platforms. Due to the rising number of contributions received for pull-based development platforms via pull-requests, the people (integrators) who decide to accept or reject the pull-requests, faced difficulty in managing the pull-requests. As a result of that, it is observed that a significant delay occurred in accepting or rejecting a pull-request. In order to address this issue, integrator recommendation systems were introduced to recommend appropriate integrators to review pull-requests.

This research focuses on providing a solution for the difficulty mentioned above, faced by the integrators. The existing solutions have used three main traditional approaches to recommend integrators. File path similarity calculation, similarity calculation for the textual content of the PR and integrator activeness calculation are the three main approaches. In this research, an algorithm was designed to combine these three traditional approaches and create a novel hybrid solution. Furthermore, research was conducted to find the optimal date limit when using the past PR data to recommend integrators for pull-requests. The solution, InteRec achieves an average of 83% accuracy in terms of Top-3 accuracy and outperforms the traditional approaches.

### Subject Descriptors:

- Information systems ~ Collaborative and social computing systems and tools
- Information systems ~ Expert search
- Information systems ~ Rank aggregation
- Information systems ~ Similarity measures

**Keywords:** Pull-Request, Reviewer recommendation, Integrator recommendation, Pull-based software development, Distributed software development