

MSc Project Report

“Labelyzt” - Labelling GitHub Issues Automatically using Text Classification based on Neural Networks

Hemika Yasinda Kodikara

2018

A report submitted as part of the requirements for the degree of
MSc in Computing: Software Technology at Robert Gordon University, Aberdeen, Scotland

Abstract

Repository Management Services such as GitHub has become a gateway for Open-Source Softwares allowing developers around the world to collaborate. Service providers such as GitHub provides capability of tracking issues of a projects or a repository allowing to manage tasks and perform analysis. Such issue tracking system also provides the capability of categorizing issues by means of labels. Categorizing issues is significantly helpful in resolving them. These labels are attached to an issue by a user of the repository. This is a daunting and time wasting task as large repositories tend to have lots of issues.

“Lablyzt” is a GitHub Integration Application which automates the task of attaching a suitable label to an issue. The application embraces text classification using Convolutional Neural Networks to categorise the issue to either of the followings:

1. Bug.
2. Question.
3. Documentation.
4. Improvement/Feature.

The results show that a correct label is predicted 60% of the times.

Project classification according to 2012 ACM Computing Classification Systems:

- Computing methodologies~Supervised learning by classification.
- Computing methodologies~Neural networks.
- Software and its engineering~Software libraries and repositories.

Keywords:

Neural Network, GitHub, Text Classification.