Scalable motif search in graphs using distributed computing

Andrew Esler

A thesis presented in partial fulfilment of the requirements for the degree of a Masters in Computer Science

Massey University

Turitea

New Zealand

2012

Acknowledgments

I would like to thank my supervisor Jens Dietrich for the funding and motivation for this project. Patrick Rynhart provided extensive help with Massey's computing infrastructure. Thanks also to those who were sounding boards for ideas and pointed out my follies.

Abstract

Motif detection allows software engineers to detect antipatterns in software. By decreasing the number of antipattern instances in a piece of software, the overall quality of the software is improved. Current methods to find these antipatterns are slow and return results only when all antipatterns have been found. The GUERY framework is able to perform motif detection using multiple cores and deliver results as they are generated. By scaling GUERY to run on multiple machines, it was hoped that research requiring many queries on a graph could be performed significantly faster than is currently possible.

The objective of this thesis was to research and prototype mechanisms whereby GUERY could be run using a cluster of computers and results delivered as a stream to interested systems. A system capable of running on a cluster of machines and delivering a stream of results as they are computed was developed.

Contents

List of Figures

List of Tables

Listings