



ESTUDOS EMPÍRICOS SOBRE O USO DE ESTRUTURAS DE DADOS EM PROJETOS OPEN-SOURCE JAVA

Theócrita Moura Maciel Malheiro Júnior¹, Tiago Lima Massoni²

RESUMO

Por apresentarem um alto índice de utilização em linguagens de programação orientadas a objetos, as bibliotecas de coleções muitas vezes são utilizadas de forma inadequada, fazendo com que usuários precisem adaptar seus programas para substituir as estruturas de dados em uso. Este tipo de modificação consiste em um refatoramento, onde o código precisa passar por alterações que preservam a sua semântica.

O objetivo do presente trabalho é identificar indícios de quais estruturas de dados oferecidas pelo Java Collections Framework (JCF) estão sendo utilizadas de forma inadequada. Para alcançar este objetivo, fizemos uma análise baseada em heurísticas, onde foram definidas métricas para as estruturas de dados List, Set, Map e Queue. Para a definição destas métricas utilizamos as árvores de decisão que foram desenvolvidas com base em estudo feito na literatura de JCF. Posterior a definição destas métricas, fizemos uma análise estática do código-fonte de 20 projetos de código aberto disponíveis em repositórios públicos.

Os quesitos de avaliação são inserção de dados, busca e remoção, os quais serão explorados de forma empírica, investigando o desempenho de cada tipo de coleção. Com a implementação dos testes e os resultados obtidos, será possível ter conhecimento do perfil de cada um dos clientes de Collections, onde poderemos sugerir possíveis refatoramentos.

Palavras-chave: Estrutura de Dados, Java Collection Framework, Refatoramento.

EMPIRICAL STUDIES ON THE USE OF DATA STRUCTURES IN OPEN-SOURCE JAVA PROJECTS

ABSTRACT

By having a high rate of use in object-oriented programming languages, the collection libraries are often used improperly, causing users to adapt their programs to replace the data structures in use. This type of modification consists of a refactoring, where the code needs to pass by changes preserving the semantics.

The objective of this work is to identify what data structures offered by the Java Collections Framework (JCF) are being used inappropriately. To accomplish this, we did an analysis based on heuristics, defining metrics for the data structures List, Set, Map and Queue. For the definition of these metrics, we used decision trees that have been developed based on JCF literature. After the definition of these metrics, we did a static analysis of the source code of 20 open source projects available in public repositories.

The criteria of evaluation are data entry, search and removal, which will be explored empirically, investigating the performance of each type of collection. With the implementation of tests and the obtained results, it is possible to have a profile of each customer that use collections, and by that, suggest possible refactorings.

Keywords: Data Structure, Java Collections, Framework, Refactoring.

¹Aluno do Curso de Ciência da Computação, Departamento de Sistemas e Computação, UFPG, Campina Grande, PB, e-mail: theocrito.junior@ccc.ufcg.edu.br

²Ciência da Computação, Professor Doutor, Departamento de Sistemas e Computação, UFPG, Campina Grande, PB, e-mail: massoni@dsc.ufcg.edu.br