Pontifícia Universidade Católica
do Rio de Janeiro

Camila Patrícia Bazílio Nunes

History-Sensitive Recovery of Features in Code of Evolving Program Families

TESE DE DOUTORADO

Thesis presented to the Programa de Pós-Graduação em Informática of the Departamento de Informática, PUC-Rio as partial fulfillment of the requirements for the degree of Doutor em Informática

Advisor : Prof. Carlos José Pereira de Lucena
Co-advisor: Prof. Alessandro Fabricio Garcia

Rio de Janeiro
October, 2012

Pontifícia Universidade Católica
do Rio de Janeiro

Camila Patrícia Bazílio Nunes

History-Sensitive Recovery of Features in Code of Evolving Program Families

Thesis presented to the Programa de Pós-Graduação em Informática, of the Departamento de Informática do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Doutor.

Prof. Carlos José Pereira de Lucena
Advisor
Departamento de Informática — PUC–Rio

Prof. Alessandro Fabricio Garcia
Co-advisor
Departamento de Informática — PUC–Rio

Prof. Vander Alves
UnB

Prof. Uirá Kulesza
UFRN

Prof. Renato Fontoura de Gusmão Cerqueira
Departamento de Informática - PUC-Rio

Prof. Karin Breitman
Departamento de Informática - PUC-Rio

Prof. José Eugenio Leal
Coordinator of the Centro Técnico Científico da PUC-Rio

Rio de Janeiro, October 19, 2012

Camila Patrícia Bazílio Nunes

She completed her undergraduate studies in Computer Science at the Federal University of Alagoas (UFAL) in 2006. She received her Master degree in Informatics from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 2009.

Bibliographic data

Nunes, Camila

History-Sensitive Recovery of Features in Code of Evolving Program Families / Camila Patrícia Bazílio Nunes; advisor: Carlos José Pereira de Lucena; co–advisor: Alessandro Fabricio Garcia. — Rio de Janeiro : PUC–Rio, Departamento de Informática, 2012.

155 f: il ; 30 cm

1. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2012.

Inclui Bibliografia.

1. Informática – Tese. 2. Famílias de Programas Evolutivas. 3. Mapeamento de Características. 4. Heurísticas. 5. História Multi-Dimensional. I. Lucena, Carlos. II. Garcia, Alessandro. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Acknowledgments

First and foremost, I would like to deeply thank my supervisors Carlos Lucena and Alessandro Garcia for guiding me during the whole execution of the work. I am very grateful to Professor Carlos Lucena, who helped in the very beginning of my research career. In particular, my deepest thanks to Alessandro Garcia who always encouraged me and kept me on the right way.

I am also thankful to the members of my examining committee who have generously contributed their time and expertise: Vander Alves, Uirá Kulesza, Renato Cerqueira, Karin Breitman. Thanks also to all my colleagues and professors from the Computer Science Department at PUC-Rio. Many thanks to the Opus research group and the Software Engineering Laboratory (LES) at PUC-Rio. During my PhD, I also had the pleasure of working with other great researchers from different universities in the last four years. They have contributed a lot to this work by providing valuable suggestions that greatly improved the quality of my thesis.

I would like also to thank my family, my parents, Julio Nunes and Girleide Nunes, and my brothers, Sady and Leonardo. They always believed in me and encouraged me to go on this journey.

My doctoral studies were financially supported by CNPq and CAPES.

Abstract

Nunes, Camila; Lucena, Carlos; Garcia, Alessandro. **History-Sensitive Recovery of Features in Code of Evolving Program Families**. Rio de Janeiro, 2012. 155p. DSc Thesis — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A program family might degenerate due to unplanned changes in its implementation, thus hindering the maintenance of family members. This degeneration is often induced by feature code of the program family that is changed individually in each member without considering other family members. In extreme cases, the program family code is fully or partially replicated and individually changed across several evolving members. Hence, as a family evolves over time, it might no longer be possible to identify and classify the implementation elements realizing common and variable features. One of the imminent activities to address these problems is the history-sensitive recovery of program family's features. This recovery process encompasses the historical analysis of each family member in order to identify and classify the implementation elements (i.e. methods, attributes) according to their variability nature. Existing work fails to analyse the evolution of the family members with the goal of recovering features' implementation elements. Additionally, existing techniques for feature analysis are not effective as they only take into consideration the history of a single member product. In summary, the contributions of this thesis are threefold: (i) a catalogue of mapping mismatches to guide software engineers in promoting the correctness and completeness of their feature mappings. This catalogue is useful to ensure a better effectiveness of the recovery process during the mapping analysis; (ii) a suite of five heuristics for the automatic expansion of feature mappings throughout the program family history. Those heuristics rely on both the multi-dimensional historical analysis of program families and the catalogue of mapping mismatches; and (iii) a suite of history-sensitive heuristics for classifying the implementation elements realizing each family feature according to their variability degree.

Keywords

Evolving Program Families. Feature Mapping. Heuristics. Multi-dimensional history.

Resumo

Nunes, Camila; Lucena, Carlos; Garcia, Alessandro. **Recuperação Sensível a História de Características no Código de Famílias de Programas Evolutivas**. Rio de Janeiro, 2012. 155p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Uma família de programas pode degenerar devido a mudanças não planejadas e, consequentemente, tendem a prejudicar a manutenção dos membros da família. Esta degeneração é frequentemente causada pelo código de uma característica ("feature") da família que é modificada individualmente em cada membro sem considerar outros membros da família. Em casos extremos, o código da família é completamente ou parcialmente replicado e individualmente modificado por diversos membros em evolução. Assim, à medida que uma família evolui, pode não ser mais possível identificar e classificar os elementos de código implementando as características comuns e variáveis. Uma das atividades iminentes para resolver esses problemas é a recuperação sensível à história de características da família. Este processo de recuperação inclui a análise histórica de cada membro da família a fim de identificar e classificar os elementos de implementação (e.g. métodos, atributos) de acordo com a sua natureza de variabilidade. Os trabalhos existentes falham em analisar a evolução dos membros de uma família com o objetivo de recuperar os elementos de implementação das características. Além disso, as técnicas existentes para a análise de características não são efetivas, pois elas apenas levam em consideração a história de um único membro por vez. Em resumo, as contribuições desta tese são divididas em três partes: (i) um catálogo de incompatibilidades de mapeamento para guiar engenheiros de software na corretude e completude de seus mapeamentos de características. Este catálogo é útil para garantir uma melhor eficácia do processo de recuperação durante a análise dos mapeamentos; (ii) um conjunto de cinco heurísticas para a expansão automática de mapeamentos de características ao longo do histórico da família de programas. Essas heurísticas são baseadas na análise histórica multi-dimensional da família e no catálogo de incompatibilidades de mapeamentos; e (iii) um conjunto de heurísticas sensíveis a história para classificar os elementos de implementação de cada característica da família de acordo com seu grau de variabilidade.

Palavras–chave

Famílias de Programas Evolutivas. Mapeamento de Características. Heurísticas. História Multi-Dimensional.

Table of Contents

List of Figures

List of Tables