Pontifícia Universidade Católica
do Rio de Janeiro

Fernando de Freitas Silva

Uma nova abordagem de mineração de repositórios de software utilizando ferramentas da Web Semântica

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação de Informática da PUC-Rio.

Orientador: Prof. Daniel Schwabe

Rio de Janeiro

Agosto de 2013

Fernando de Freitas Silva

Uma nova abordagem de mineração de repositórios de software utilizando ferramentas da Web Semântica

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Daniel Schwabe
Orientador
Departamento de Informática - PUC-Rio

Prof. Arndt Von Staa
Departamento de Informática - PUC-Rio

Prof. Alessandro Garcia
Departamento de Informática - PUC-Rio

Prof. José Eugenio Leal
Coordenador(a) Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 15 de agosto de 2013

Fernando de Freitas Silva

Graduou-se em Bacharel em Sistemas de Informação pela Pontifícia Universidade Católica do Rio de Janeiro em 2009.

Ficha Catalográfica

Silva, Fernando de Freitas

Uma nova abordagem de mineração de repositórios de software utilizando ferramentas da Web semântica / Fernando de Freitas Silva ; orientador: Daniel Schwabe. – 2013.

178f : il. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2013.

Inclui bibliografia

Informática – Teses. 2. Web semântica. 3. Manutenção de software. 4. Engenharia de software. 5. Semântica. 6. Repositórios de software. I. Schwabe, Daniel. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

Agradeço aos meus pais, Walquíria de Freitas Silva e Fernando Jorge Lopes Silva, por sempre me apoiarem, incentivarem e por terem me dado uma criação excepcional. Sem vocês eu não estaria onde estou hoje. Muito obrigado!

Agradeço a minha esposa e melhor amiga, Geyzyely, pelo apoio durante esta empreitada e por ter compreendido as madrugadas que tive que passar longe de você para estudar. Obrigado por estar ao meu lado e espero que eu possa te apoiar do mesmo modo que você me apoiou. Te amo.

Agradeço ao meu "amigão" e filho, Joaquim, que apesar de sua pouca idade entendia os momentos que eu não podia brincar ou estar com ele. Obrigado por ser um filho maravilhoso e que você alcance coisas maiores do que eu já imaginei.

Tenho muito a agradecer ao meu orientador Professor Daniel Schwabe pelo constante incentivo e apoio durante Mestrado. Sem suas observações e direcionamentos eu não teria chegado até aqui. Muito obrigado!

Agradeço ao Gustavo Robichez de Carvalho e ao Professor Carlos José Pereira de Lucena pelo apoio durante minha graduação e pela oportunidade de ingressar no Mestrado. Muito obrigado!

Agradeço aos meus colegas da PrimeUp, ao sócio Leandro Daflon e ao meu gerente João Manoel Silvestre pelo compreensão nos momentos que tive que me ausentar do trabalho durante meu Mestrado. Muito obrigado!

Agradeço à PUC-Rio que me formou Bacharel em Sistemas de Informação e me deu a oportunidade de me tornar Mestre em Informática sem custo algum.

Finalmente, gostaria de agradecer à CAPES pela bolsa concedida durante os dois anos do Mestrado. Este apoio foi imprescindível para que eu pudesse concluir meus estudos. Obrigado!

Resumo

Silva, Fernando de Freitas; Schwabe, Daniel. **Uma nova abordagem de mineração de repositórios de software utilizando ferramentas da Web Semântica.** Rio de Janeiro, 2013. 178p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A Mineração de Repositórios de Software é um campo de pesquisa que extrai e analisa informações disponíveis em repositórios de software, como sistemas de controle de versão e gerenciadores de issues. Atualmente, diversos trabalhos nesta área de pesquisa têm utilizado as ferramentas da Web Semântica durante o processo de extração a fim de superar algumas limitações que as abordagens tradicionais enfrentam. O objetivo deste trabalho é estender estas abordagens que utilizam a Web Semântica para minerar informações não consideradas atualmente. Uma destas informações é o relacionamento existente entre as revisões do controle de versão e as mudanças que ocorrem no Abstract Syntax Trees dos arquivos modificados por essas revisões. Adicionalmente, esta nova abordagem também permite modelar a interdependência entre os projetos de software, suas licenças e extrair informações dos builds gerados por ferramentas de integração contínua. A validação desta nova abordagem é demonstrada através de um conjunto de questões que são feitas por desenvolvedores e gerentes durante a execução de um projeto e que foram identificadas em vários trabalhos da literatura. Demonstramos como estas questões foram convertidas para consultas SPARQL e como este trabalho consegue responder às questões que não são respondidas ou são respondidas parcialmente em outras ferramentas.

Palavras-chave

Web Semântica; Manutenção de Software; Engenharia de Software Semântica; Repositórios de Software.

Abstract

Silva, Fernando de Freitas; Schwabe, Daniel (Advisor). **A New Approach for Mining Software Repositories using Semantic Web Tools.** Rio de Janeiro, 2013. 178p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The Mining of Software Repositories is a field of research that extracts and analyzes information available in software repositories, such as version control systems and issue trackers. Currently, several research works in this area have used Semantic Web tools during the extraction process to overcome some limitations that traditional approaches face. The objective of this work is to extend the existing approaches that use Semantic Web tools to mine information not considered in these works. The objective of this work is to extend these approaches using the Semantic Web to mine information not currently considered. One of these information is the relationship between revisions of version control and the changes that occur in the Abstract Syntax Trees of files modified by these revisions. Additionally, this new approach also allows modeling the interdependence of software projects, their licenses and extracting information from builds generated by continuous integration tools. The validation of this approach is demonstrated through a set of questions that are asked by developers and managers during the execution of a project and have been identified in various works in the literature. We show how these questions were translated into SPARQL queries and how this work can answer the questions that are not answered or are partially answered in other tools.

Keywords

Semantic Web; Software Maintenance; Semantic Software Engineering; Software Repositories.

Sumário

Apêndice E – Consultas SPARQL “Perguntas Frequentes dos Desenvolvedores” 171

Lista de figuras

Lista de tabelas

Lista de quadros