Pontifícia Universidade Católica
do Rio de Janeiro

Thiago Pinheiro de Araújo

SDiff: Uma ferramenta para comparação de documentos com base nas suas estruturas sintáticas

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio.

Orientador: Prof. Arndt von Staa

Rio de Janeiro

Março de 2010

Thiago Pinheiro de Araújo

SDiff: Uma ferramenta para comparação de documentos com base nas suas estruturas sintáticas

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Arndt von Staa
Orientador

Departamento de Informática - PUC-Rio

Prof. Renato Fontoura de Gusmão Cerqueira

Departamento de Informática - PUC-Rio

Prof. Marcus Vinicius Soledade Poggi de Aragão

Departamento de Informática - PUC-Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro
Técnico Científico - PUC-Rio

Rio de Janeiro, 8 de março de 2010

Thiago Pinheiro de Araújo

Graduou-se em Engenharia de Computação na Pontifícia Universidade Católica do Rio de Janeiro (Brasil, Rio de Janeiro).

Ficha Catalográfica

Araújo, Thiago

SDiff: Uma ferramenta para comparação de documentos com base nas suas estruturas sintáticas / Thiago Pinheiro de Araújo; orientador: Arndt von Staa. — Rio de Janeiro : PUC–Rio, Departamento de Informática, 2010.

v., 95 f.: il. ; 29,7 cm

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

Inclui referências bibliográficas.

1. Gerência de configuração. 2. Controle de versão. 3. Ferramenta de desenvolvimento de software. 4. Engenharia de Software. I. Staa, Arndt von. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Para meus pais, Ítalo e Antônia.

Agradecimentos

Ao meu pai, Ítalo José de Araújo, meu melhor amigo, que se despediu há pouco tempo. Sem sua presença, experiência de vida, participação, incentivo, motivação, alegria e interesse na minha formação, não seria possível conquistar tudo que conquistei. Muito obrigado por tudo, e olhe por nós.

A minha mãe, Antônia dos Santos Cunha Pinheiro, e a minha tia, Maria do Carmo Cunha Pinheiro, pelo apoio e carinho nos momentos mais difíceis. Obrigado pela incansável dedicação.

A Ana Carolina Andrade, pelo seu amor, pela sua paciência, pelos seus conselhos e pelo seu apoio e compreensão em cada momento de nossas vidas.

Ao meu orientador, prof. Arndt von Staa, por toda sua dedicação e interesse na minha formação acadêmica. Obrigado pelos seus conselhos, pela sua orientação, pelo seu entusiasmo com a pesquisa, e por todas as conversas que tivemos, sempre construtivas para minha vida acadêmica e pessoal.

A Maria Isabel Cristovão Prado, pela sua amizade em todos esse anos, pelo carinho e atenção que têm com a nossa família, e também pelo papel fundamental que teve na minha carreira. Muito obrigado.

A Carlos Eduardo Crestana e Ricardo Gomes, pela grande amizade e confiança no meu potencial. Agradeço também a disponibilidade e paciência nas discussões que contribuíram para a realização deste trabalho.

A João Magalhães pela sua orientação, apoio e compreensão.

Ao CNPq e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos professores da Comissão examinadora.

A todos os professores e funcionários do Departamento pelos ensinamentos e pela ajuda.

A todos os meus familiares, amigos de infância, amigos da graduação e do mestrado, que de alguma forma contribuíram para a realização deste trabalho.

Resumo

Araújo, Thiago; Staa, Arndt von. **SDiff: Uma ferramenta para comparação de documentos com base nas suas estruturas sintáticas.** Rio de Janeiro, 2010. 95p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Associado a cada sistema de controle de versão existe uma ferramenta de comparação responsável pela extração das diferenças entre duas versões de um documento. Estas ferramentas costumam realizar a comparação baseando-se na informação textual dos documentos, em que o elemento indivisível na comparação é a linha ou a palavra. Porém, o conteúdo versionado normalmente é fortemente estruturado (como exemplo, linguagens de programação) e a utilização deste mecanismo pode desrespeitar limites sintáticos e outras propriedades do documento, dificultando a interpretação das alterações. Nesse trabalho foi construída uma ferramenta para identificar as diferenças entre duas versões de um documento utilizando um mecanismo de comparação baseado na sua estrutura sintática. Desta forma, é possível identificar com maior precisão as diferenças relevantes ao leitor, reduzindo o esforço para compreender a semântica das alterações. A ferramenta construída é capaz de suportar diferentes tipos de documentos a partir da implementação de componentes que tratem das sintaxes desejadas. O componente implementado como exemplo neste trabalho trata a sintaxe da linguagem de programação C++.

Palavras-chave

Gerência de configuração; Controle de versão; Ferramenta de desenvolvimento de software; Engenharia de software.

Abstract

Araújo, Thiago; Staa, Arndt von (Advisor). **SDiff: a comparison tool based in syntactical document structure.** Rio de Janeiro, 2010. 95p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Associated with each version control system there's a comparison tool for extracting the differences between two versions of a document. These tools tend to make a comparison based on textual information from documents, in which the indivisible element is the line or word. But the content versioned is usually highly structured (for example, programming languages) and the use of this mechanism can disrespect syntactical limits and other properties of the document, becoming difficult to interpret what really changed. In this work we created a tool to identify differences between two versions of a document using a comparison mechanism based on the syntactic structure. Thus, it is possible to identify more precisely the relevant differences to the reader, reducing the effort to understand the semantics of the changes. The tool can support different types of documents by implementing components that interprets the desired syntax. The example syntax component implemented in this work deals with the syntax of the programming language C++.

Keywords

Software Configuration Management; Version control system; Software development tool; Software Engineering.

Sumário

Lista de figuras

Lista de tabelas