

Anteprojeto de Tese de Doutorado

# Sobre a Adoção de Boas Práticas de Desenvolvimento de Software em Projetos de Software Livre

Rodrigo Rocha Gomes e Souza  
rodrigo@dcc.ufba.br

Dezembro de 2009

## 1 Motivação

No contexto de desenvolvimento de software, boas práticas são estratégias ou atividades que comprovadamente contribuem para o sucesso de um projeto de software [2]. A aplicação de boas práticas dentro de uma equipe torna-a mais competitiva ao equilibrar as diferenças de desempenho entre seus membros. Não obstante, a transferência de práticas entre desenvolvedores pode ser prejudicada por diversos fatores, como falta de motivação ou de experiência [7].

Projetos de software livre são tipicamente caracterizados pela participação de voluntários com diferentes níveis de envolvimento, o que é um fator que potencialmente dificulta a transferência de práticas na equipe. Outro fator é a alta rotatividade de desenvolvedores [5]. Além disso, comparado a projetos desenvolvidos por empresas, projetos de software livre geralmente possuem uma estrutura de controle menos centralizada [4], o que reforça a importância da coordenação e da comunicação entre desenvolvedores na propagação de boas práticas.

Projetos de software livre têm adquirido importância em diversos segmentos. Um estudo mostrou que, em 2004, mais de 1 milhão de desenvolvedores no Estados Unidos estavam trabalhando em projetos de software livre. Além disso, projetos de software livre dominam o mercado mundial de servidores web, servidores de e-mail e servidores DNS [8].

Infelizmente, a maior parte do conhecimento disponível sobre os fatores que influenciam a adoção de práticas em projetos de software livre é proveniente de observações isoladas e estudos de caso [1]. Essas observações, embora de grande valia, são essencialmente tendenciosas. Uma compreensão mais abrangente dos fatores é necessária para que se tomem medidas mais efetivas para a adoção de boas práticas.

## 2 Objetivo

O principal objetivo desta pesquisa é entender como a adoção de boas práticas em um projeto de software é afetada pela evolução de aspectos ligados ao processo, ao produto e à organização de desenvolvedores do projeto. Pretende-se assim identificar em quais contextos determinadas práticas são efetivamente aplicadas e entender os mecanismos que reforçam a adoção de práticas de desenvolvimento de software.

Os seguintes aspectos do processo, do produto e da organização social de cada projeto serão investigados:

- idade do projeto;
- tamanho do software;
- número de desenvolvedores;
- número de downloads do software;
- atividade em listas de discussão;
- concentração do esforço de codificação em um grupo pequeno de desenvolvedores.

A análise será focada nas atividades de implementação e controle de qualidade. Serão investigadas, entre outras, as seguintes práticas:

- implementação de testes de unidade automatizados antes da implementação de funcionalidades;
- implementação dos testes por pessoas diferentes daquelas que implementam funcionalidades;
- correção de *bugs* antes da implementação de novas funcionalidades;
- integração contínua das mudanças com o repositório de código-fonte;
- posse coletiva de código.

Espera-se, com este trabalho, chegar a respostas para questões de pesquisa como as que se seguem:

- A adoção de boas práticas é afetada pelo tamanho da equipe? Quais práticas são mais afetadas?
- A adoção de práticas é mais uniforme em projetos mais centralizados?
- Existe relação entre a adoção de práticas e o perfil de um desenvolvedor? Por exemplo, será que desenvolvedores que participam ativamente de listas de discussão do projeto tendem a seguir mais boas práticas?

- A saída de desenvolvedores de um projeto prejudica a adoção de boas práticas? De que forma?

O conhecimento produzido por esta pesquisa poderá ser usado para se adquirir maior controle sobre as práticas adotadas em um projeto, tornando o seu desenvolvimento mais previsível.

### 3 Metodologia

Pretende-se estudar projetos de software livre hospedados no repositório SourceForge.net, que abriga cerca de 230 mil projetos<sup>1</sup>. O SourceForge.net disponibiliza publicamente, para cada projeto, informações sobre o número de *downloads* e idade do projeto, além de sistema de controle de versão, sistema de rastreamento de *bugs* e listas de discussão.

A grande disponibilidade de dados relacionados ao desenvolvimento favorece a realização de estudos em larga escala. Nesta pesquisa serão utilizados predominantemente métodos quantitativos de análise de dados para viabilizar o estudo de uma amostra significativa de projetos de software livre. Espera-se assim obter conclusões generalizáveis sobre o uso de boas práticas em projetos de software livre.

A princípio serão realizados estudos de caso com um pequeno conjunto de sistemas, em caráter exploratório, a fim de se adquirir familiaridade com as informações disponíveis e com ferramentas de extração e análise de dados. As questões de pesquisa serão refinadas de acordo com observações realizadas nessa etapa.

A seguir será realizado um estudo em larga escala, considerando uma amostra de projetos estratificada de acordo com dados como número de desenvolvedores e tamanho do software. Serão extraídos dados de repositórios de controle de versão, sistemas de acompanhamento de *bugs* e listas de discussão.

A identificação das práticas adotadas por um projeto será feita a partir do uso de técnicas de mineração de processos [6] sobre o histórico dos dados extraídos do projeto. Tais técnicas têm como finalidade extrair modelos de processos a partir de dados temporais.

A organização social dos desenvolvedores será estudada através de técnicas de análise de redes sociais [3]. Essas técnicas permitirão investigar aspectos como a transferência e a concentração de trabalho entre desenvolvedores.

Para responder às questões de pesquisa, serão usados métodos de análise exploratória de dados e inferência estatística, relacionando características de projetos e de desenvolvedores, o histórico de desenvolvimento e a aplicação de boas práticas. A partir dos resultados da análise poderão ser realizadas investigações mais aprofundadas sobre projetos selecionados e comparações com resultados obtidos em estudos de caso.

---

<sup>1</sup><http://sourceforge.net/about>

## 4 Proposta de Cronograma

As seguintes atividades serão realizadas durante este projeto, de acordo com os prazos indicados na Tabela 1:

1. **revisão bibliográfica;**
2. **realização de estudos de caso;**
3. **realização de um estudos em larga escala;**
4. **divulgação científica;**
5. **redação e defesa da tese.**

Atividade	2010.1	2010.2	2011.1	2011.2	2012.1	2012.2	2013.1	2013.2
1	•	•	•					
2			•	•				
3					•	•		
4				•		•		
5						•	•	•

Tabela 1: Proposta de Cronograma de Atividades

## Referências

- [1] Kevin Crowston and James Howison. The social structure of free and open source software development. *First Monday*, 10(2), 2005.
- [2] Ivar Jacobson. Enough of Processes- Let's Do Practices. *Requirements Engineering, IEEE International Conference on*, 0, 2007.
- [3] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.
- [4] Eric S. Raymond. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly and Associates, Inc., Sebastopol, CA, USA, 2001. Foreword By-Young, Bob.
- [5] Gregorio Robles and Jesus Gonzalez-Barahona. Contributor Turnover in Libre Software Projects. *Open Source Systems*, pages 273–286, 2006.
- [6] Vladimir Rubin, Christian Günther, Wil van der Aalst, Ekkart Kindler, Boudewijn van Dongen, and Wilhelm Schäfer. Process Mining Framework for Software Processes. pages 169–181. 2007.

- [7] Gabriel Szulanski. Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm. *Strategic Management Journal*, 17:27–43, 1996.
- [8] David A. Wheeler. Why Open Source Software / Free Software (OSS/FS, FOSS, or FLOSS)?: Look at the Numbers! 2007.