

1 - Introdução

1.1 - Motivação

Software livre é um tipo de programa desenvolvido por um grupo de pessoas que além de disponibilizar o software gratuitamente, ainda coloca a disposição seu código fonte para que este seja distribuído e alterado livremente. Este tipo de software ganhou muita exposição com projetos como o Linux e Apache, mas a comunidade de software livre não se restringe de maneira alguma a apenas esses dois nomes. Existem diversos outros projetos famosos como o Mozilla , Jboss ou mesmo o CVS (Control Version System), mas existem também milhares de outros que não tem a mesma divulgação, mas nem por isso perdem em qualidade para seus concorrentes comerciais. Apenas na página SourceForge.net [SourceForge03] (uma página dedicada a hospedar projetos de software livre), existem mais de 60.000 projetos ativos.

O modo como o desenvolvimento de um software, dito livre, ocorre é considerado por muitos como pouco convencional, e não condiz com os métodos de desenvolvimento de software propostos pela engenharia de software. O fato é que existem hoje em dia milhares de projetos de software livre em andamento, e muitos com um nível de sucesso igual ou mesmo superior aos seus equivalentes comerciais [Godfrey00] [Mockus02].

Este tipo de software é desenvolvido em comunidades muitas vezes com membros separados por grandes distâncias e se comunicando unicamente por ferramentas de groupware de uso comum na Internet (como listas de discussão, e-mail e bate papo); estes membros usam seu tempo livre para o desenvolvimento e são em sua maioria usuários do software que desenvolvem.

Em projetos grandes (como é o caso do Linux e Apache), o número de membros desta comunidade pode chegar à casa dos milhares e mesmo que apenas uma pequena parcela desta contribua efetivamente com o código, uma outra parte bem maior contribui enviando relatórios de erros encontrados (estudos com uma versão do Apache revelaram que mais de 3.000 pessoas enviaram mensagens

relatando erros encontrados [Mockus02]). O número de pessoas envolvidas em projetos deste tipo por si só já o torna um processo que merece ser analisado à parte.

Outro ponto deste processo que merece atenção, é que durante a vida de um projeto de software livre, este pode vir a passar pelos cuidados de diversos desenvolvedores diferentes, e mesmo ter diversos caminhos paralelos de desenvolvimento. Por essas razões, dificilmente um processo deste tipo chega a um final definitivo, já que mesmo que a equipe original perca o interesse ou ache que o software já atingiu o estágio de qualidade desejado, nada impede que outra equipe venha a continuar o desenvolvimento deste (esta equipe pode ser formada por exemplo, por usuários que necessitam de novos módulos adicionados ao projeto). Podemos dizer, portanto, que um projeto de software livre se encontra sempre em um contínuo processo evolutivo. Este processo de evolução em si também merece um estudo à parte, já que estudos iniciais contam que ele nem sempre obedece às leis de Lehman [Godfrey00].

Um dos problemas do método de desenvolvimento de software livres é a pouca atenção dada à documentação. Isso ocorre devido a própria natureza informal de um projeto deste tipo, e o resultado é que as únicas fontes de documentação disponíveis são normalmente o código fonte e os arquivos contendo mensagens trocadas entre os participantes do projeto. Esta falta de documentação acaba por gerar em alguns projetos, uma indesejada barreira de entrada para novos participantes na evolução do software. Como a entrada de novos membros é algo sempre encorajado e necessário para estes tipos de projeto, deve-se fazer um esforço para diminuir ao máximo essa barreira.

Uma proposta apresentada neste trabalho é a união de cenários com o código fonte de um software livre. Serão apresentados com o decorrer do mesmo, os benefícios obtidos com a proposta, e como é possível diminuir a barreira de entrada de um projeto de software livre usando esta técnica. Para validar a proposta, será apresentado um projeto de software livre desenvolvido na PUC-Rio chamado C&L. O autor desta dissertação foi o gerente responsável pela evolução

desta ferramenta de software livre durante sua primeira fase de evolução e contribuiu com boa parte do código que hoje é usado nela. A seguir será apresentada uma ferramenta de extração de cenários diretamente do código, baseada na proposta do trabalho, que tem como objetivo transformar os cenários do código em uma documentação externa. Esta ferramenta foi desenvolvida inteiramente pelo autor desta dissertação e seu código fonte se encontra disponível livremente para qualquer usuário.

O trabalho está dividido em 6 capítulos, o primeiro mostra uma visão geral da área de software livre e analisa sua situação atual, o segundo capítulo se foca mais no software livre em si, falando de sua definição, comunidade, características gerais e como este está presente no Brasil, o terceiro se concentra na evolução de software e como ela está relacionada com software livre, o quarto fala sobre a ferramenta C&L, será apresentada sua história, sua utilidade e o que foi aprendido com ela. O quinto capítulo fala sobre a proposta apresentada nesta dissertação, seus benefícios e a experiência de sua utilização no C&L, bem como também apresenta a ferramenta de extração de cenários diretamente do código baseada na proposta do trabalho. O capítulo seguinte conclui a dissertação, apresentando as contribuições desta, além de apresentar as dificuldades encontradas e trabalhos futuros.

1.2 - Visão geral da área de software livre

Não existe uma definição exata do termo “Software Livre”, mas geralmente a definição usada é a de que “Software Livre é um tipo de software que é distribuído de acordo com os termos estabelecidos pelo Open Source Definition” [Joseph02]. Open Source Definition é um documento feito e mantido pela OSI (Open Source Initiative), e que apesar das aparências, não é uma licença, mas uma especificação dos termos de distribuição de um software. Este é composto por dez critérios (todos serão mostrados no capítulo 2), e para que um determinado software seja considerado um software livre, é necessário que este siga os dez critérios, caso apenas um destes não seja respeitado, o software é considerado não elegível para receber a certificação da OSI.

A história do software livre começa junto com alguns dos primeiros sistemas desenvolvidos na década de 40. Durante essa década, os computadores eram usados basicamente como grandes calculadoras, sendo valorizados pela rapidez com que realizavam cálculos complexos, e os programadores eram em sua maioria engenheiros, matemáticos ou cientistas com conhecimento profundo em matemática. A partir da década de 50, começaram a surgir aplicações de processamento de dados empresarial, estas aplicações necessitavam de um grande volume de entrada e saída de dados, o que na época era bem problemático devido as limitações do hardware. Fazer um software nesta época era uma tarefa árdua devido a falta de memória e a lentidão dos periféricos, sendo necessário as vezes esperar um dia inteiro para uma aplicação ser compilada. Era necessário, portanto, uma grande habilidade por parte dos programadores na questão da economia de recursos gastos, o que transformava a criação de software em uma tarefa difícil e demorada.

Todas esses problemas terminaram por contribuir para que os sistemas desenvolvidos com sucesso fossem amplamente compartilhados. Não existia simplesmente uma ideologia de camaradagem, mas sim uma busca por eficiência; um exemplo deste comportamento foi à colaboração entre as indústrias de aviação militares e não militares (concorrentes vorazes) que ocorreu nesta época nos EUA, e que resultou na formação do PACT (Project for Advanced of Coding Techniques) [Leonard00].

Em 1956, o governo dos EUA proibiu a gigante das telecomunicações AT&T de entrar em mercados que não tivessem ligação com a telefonia, como consequência desta proibição, ela ficou impossibilitada de atuar comercialmente no mercado de computadores. Mais tarde na década de 60, essa proibição acabou por fazer com que o sistema operacional Unix que fora desenvolvido em 1969 no AT&T Bell Labs por Dennis Ritchie e Ken Thompson, não pudesse ser vendido comercialmente. Com isso o Unix passou a ser distribuído gratuitamente em conjunto com seu código fonte, para universidades e entidades de pesquisa.

A partir deste momento o Unix ganhou popularidade e diversos canais de distribuição do sistema operacional e de suas aplicações apareceram pelo mundo. O mais famoso destes foi o Berkeley Software Distribution (BSD) que foi estabelecido em 1977. O grupo BSD fez diversas modificações no Unix, uma parte destas foram feitas pelos seus próprios membros e outras foram feitas através de contribuições de código fonte vindas de pessoas de fora. O resultado era distribuído gratuitamente, e o sucesso foi tanto que esta versão foi escolhida pelo DARPA (Defense Advanced Research Projects Agency) para ligar os nós de pesquisa da Arpanet, que mais tarde se tornaria a Internet. Algumas das maiores contribuições do BSD foram os famosos Sendmail e BIND (Berkeley Internet Name Domain).

Mais tarde, na década de 80, com o fim do monopólio da AT&T, esta rumou para a comercialização do Unix, o que acabou por gerar uma batalha judicial com o grupo BSD que durou vários anos. A incerteza do resultado desta batalha acabou por gerar uma migração dos programadores que contribuía com o Unix para o Linux, contribuindo assim para o crescimento deste. A batalha judicial terminou apenas no começo da década de 90, e mais tarde o projeto BSD se dividiu em vários projetos concorrentes, como o FreeBSD e o OpenBSD.

Outro grande acontecimento da década de 80 na área de software livre foi sem dúvida o aparecimento da Free Software Foundation (FSF) em 1985, fundada por Richard Stallman. A FSF é famosa por ter criado a linha de produtos de software livre conhecida como GNU (GNU's Not Unix!), e também por seu manifesto GNU escrito por Stallman no mesmo ano da fundação da FSF. O projeto GNU gerou e ainda gera diversos sistemas, alguns destes são largamente conhecidos e usados, como por exemplo, o compilador C GCC(GNU C Compiler) e o editor Emacs. A FSF é conhecida pelo seu grande idealismo, e todos os seus sistemas podem ser copiados ou modificados de acordo com a vontade do usuário, contanto que estes não adicionem restrições próprias ao resultado, não sendo possível, por exemplo, cobrar pelo todo ou parte de um software gerado a partir de um projeto GNU. A maioria da verba usada pelo

projeto GNU vem hoje em dia de doações individuais, sendo que essas doações em 2002 representaram cerca de 67% do fundo operacional do projeto [GNU03].

Durante todos este tempo, vários casos de sucesso de sistemas desenvolvidos com a filosofia do software livre apareceram e ainda aparecem, sendo que podemos citar alguns como grandes exemplos de sucesso, como o Linux, a família de produtos da Apache Software Foundation (como o Apache HTTP Server e o Jakarta Tomcat), as linguagens Perl e PHP, além de diversos outros sistemas.



1.3 - Análise da área

Nos últimos anos, a comunidade de software livre e seus produtos ganharam uma grande atenção por parte da mídia mundial. No princípio, o movimento de software livre foi tratado como um simples movimento contra um inimigo comum, neste caso a Microsoft e outras grandes produtoras de software. Mas a verdadeira batalha é para aumentar a produção de sistemas de alta qualidade e confiabilidade que o mercado tanto precisa, apesar da crescente complexidade destes [Jai01].

Hoje em dia, projetos de software livre estão sendo desenvolvidos em todas as grandes categorias de sistemas, como sistemas operacionais, linguagens de programação, servidores de aplicação e aplicações empresariais. Grandes

empresas tecnológicas como a IBM e a Sun Microsystems estão devotando consideráveis somas de tempo e recursos em projetos de software livre, pois elas vêm com preocupação o fato de empresas menores como a Red Hat e a VA Linux estarem tomando a dianteira na produção de sistemas de qualidade e menor custo [Gunnison01].

O crescimento da área de software livre se deve também em grande parte a popularização da Internet, já que uma forte característica que está presente na maioria dos projetos deste tipo é a grande distribuição geográfica entre os participantes. Devido a este fato, a Internet com suas ferramentas de groupware tem um papel fundamental na integração dos membros dos projetos; as ferramentas mais comuns usadas em projetos deste tipo são simples e de fácil obtenção, como sistemas para correio eletrônico, gerenciadores de listas de discussões, bate-papos e controle de versões. Existem inclusive alguns portais especializados em hospedar projetos de software livre que já possuem integradas algumas destas ferramentas.

Alguns fatos que demonstram a importância de projetos deste tipo para o mercado de sistemas atual:

- até Julho de 2003, o servidor HTTP mais usado pelo mercado era o Apache com mais do que o dobro de usuários que o segundo colocado (Microsoft IIS). O Apache domina o mercado desde 1996, e na última pesquisa realizada pelo Netcraft [Netcraft03], o servidor Apache possuía 63,72% do mercado, enquanto o concorrente da Microsoft possuía apenas 25,95%.

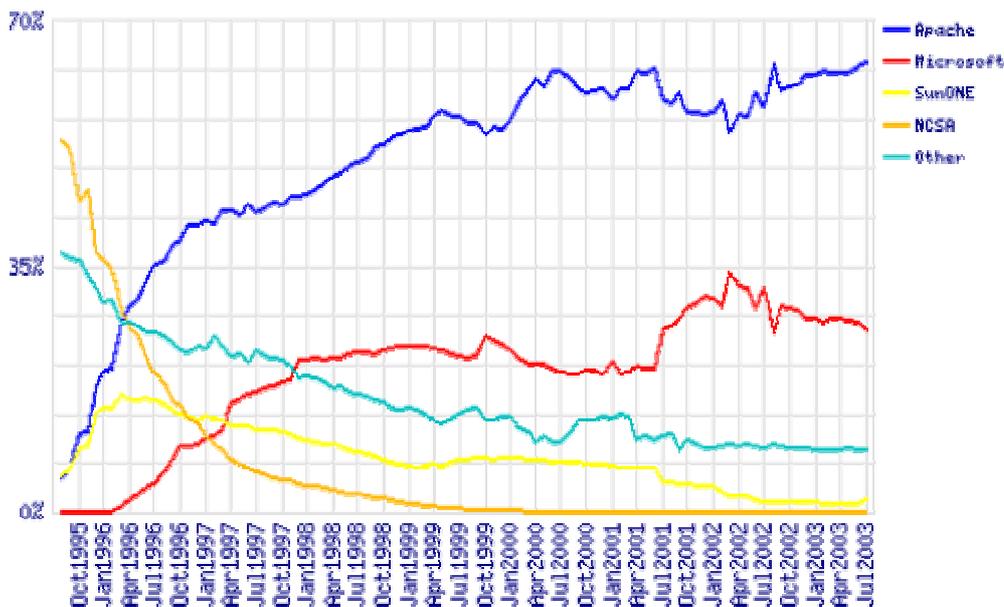


Figura 1.1 - Distribuição do mercado de servidores HTTP de 1995 até 2003

- em 2001 o GNU/Linux era o segundo sistema operacional mais popular para realizar a hospedagem de páginas na Internet, com quase 30% do mercado, e os sistemas BSD (FreeBSD, NetBSD e OpenBSD) tinham cerca de 6% [Netcraft03].
- em novembro de 2001, uma pesquisa publicada pela Evans Data Survey com 400 desenvolvedores representando 70 países, mostrou que 39,6% dos desenvolvedores norte americanos e 48,1% dos desenvolvedores dos demais países, planejam focar a maioria de suas aplicações para o GNU/Linux no ano seguinte. Uma pesquisa posterior realizada em outubro de 2002 mostrou que o interesse continuava a crescer e que cerca de 59% destes dos pesquisados pretendiam escrever código para GNU/Linux.
- em fevereiro de 2002, uma pesquisa realizada pelo OpenForum Europe com o título de “Análise de oportunidade de mercado para Software Livre” realizada com vários CIOs e diretores financeiros, mostrou que 37% das empresas já faziam uso de algum tipo de software livre, e 49% delas esperavam fazer uso destes no futuro.

- pesquisa realizada em setembro de 2001 mostrou que o servidor de correio eletrônico mais usado no mundo é o Sendmail com 42% do mercado mundial, mais de duas vezes o valor obtido pelo servidor concorrente da Microsoft [David03].
- a linguagem mais usada para a realização de scripts em páginas na Internet é o PHP, que em uma pesquisa realizada em Junho de 2002, estava presente em 24% das páginas na Internet, mais do que o seu principal concorrente (ASP da Microsoft) [David03].

Estes são apenas alguns dos muitos exemplos existentes que servem para ilustrar a importância que sistemas deste tipo tem no mercado atual. De fato, o uso de software livres se tornou tão comum nos dias de hoje, que muitas vezes fazemos uso deles sem ao menos notar que estes são de fato software livres.

Não existe um número preciso relativo à quantidade de projetos de software livre existentes hoje em dia, mas sabe-se que este número chega a algumas dezenas de milhares como dito anteriormente, existem mais de 60.000 projetos hospedados apenas no portal sourceforge.net [SourceForge03]. Este número expressivo por si só já é um indicativo da popularidade deste tipo de projeto, mas dentre estes todos, existem aqueles que se destacam dos demais devido a seu tamanho e popularidade. Como exemplos de projetos desta importância, podemos citar os famosos Linux e Apache HTTP Server. Segue uma pequena descrição da história de cada um destes projetos.

1.3.1 Linux

O projeto começou em 1991 com o na época estudante da universidade de Helsinki, Linus Torvalds. Este usou como modelo para a sua criação, o também sistema operacional Minix, desenvolvido por Andrew Tanenbaum como um clone simplificado do Unix, que tinha seu código aberto, apesar de que qualquer modificação no código fonte necessitasse da permissão do autor.

O intuito de Torvalds era o de criar um sistema operacional parecido com o Unix para máquinas IBM PC. Para esse fim, o autor procurou abertamente ajuda da comunidade e a resposta foi e ainda é enorme. Estima-se que mais de 1000 programadores contribuíram com código para o desenvolvimento do Kernel do Linux, e seu autor chega a dizer que este é o maior projeto colaborativo da história da humanidade [Joseph02].

Hoje em dia, o Linux se tornou mais popular até mesmo que o Unix (sistema em que inicialmente se espelhou), e é o sistema operacional da plataforma PC mais portado do mundo. Grandes empresas como a IBM, Sun Microsystems e Oracle portam seus produtos com sucesso para este sistema operacional, que se tornou um grande caso de sucesso de um projeto de software livre.

1.3.2 Apache HTTP Server

O projeto começou inicialmente em 1995, e seus desenvolvedores (todos voluntários) usaram a Internet como ferramenta para trabalhar em conjunto e criar remendos para o então popular e gratuito web server da NCSA (National Center for Supercomputing Applications). O resultado foi chamado de A PatCHy Server, e mais tarde de Apache.

Este projeto não tem líder nem endereço físico, as decisões são tomadas via votação por correio eletrônico pelo “Apache Core Group” (ACG), que reúne os membros que mais contribuíram enviando código para o projeto. O único modo de se fazer parte do ACG, é sendo convidado por algum dos membros, o que não impede de maneira alguma que um indivíduo não membro contribua para o projeto, ele apenas não fará parte do grupo que vota sobre os rumos do projeto (planos de desenvolvimento, códigos que serão adicionados na próxima versão e demais decisões).

O resultado do projeto foi e ainda é um estrondoso sucesso. Grandes empresas como a IBM participam do projeto, e este é desde 1996 o servidor HTTP mais usado no mercado mundial.

Em 1999, o Apache Group se tornou a Apache Software Foundation, e possui hoje um grande número de outros importantes projetos de software livre ativos, como o Jakarta Tomcat, Ant, e outros.