Analysing and Measuring Open Source Projects

Miguel Regedor, miguelregedor@gmail.com

MI-STAR 2011, UCE15, Universidade do Minho

31 de Janeiro de 2011

Tabel of Contents

Open Source Software

Nowadays, Open Source Software is disseminated

Open Source Projects

- Android
- Apache
- Mozilla Firefox
- Open Office
- Ubuntu

Open Source Software

- Thousands of OSS packages can be found online and free to download

Open Source Software

- Thousands of OSS packages can be found online and free to download
- OSS is not only used by computer specialists

Open Source Software

- Thousands of OSS packages can be found online and free to download
- OSS is not only used by computer specialists
- Open Source can be now considered the largest software industry in the world

Open Source Software

- Thousands of OSS packages can be found online and free to download
- OSS is not only used by computer specialists
- Open Source can be now considered the largest software industry in the world

Measuring the savings that people are making in licence fees, the open source industry is worth 60 billion dollars.

Development Process

Open Source communities work in a *bazaar style*

Development Process

Open Source communities work in a *bazaar style*

Traditional Software Development Process

- Similar to building cathedrals
- Few specialized individuals working in isolation

Open Source Development Process

- Chaotic way
- Resemble a great babbling *bazaar*.

Quality

Can software that is developed in such chaotic way be trusted as a high quality product?

Quality

Can software that is developed in such chaotic way be trusted as a high quality product?

The shock is that in fact the *bazaar style* seemed to work

Quality

Can software that is developed in such chaotic way be trusted as a high quality product?

The shock is that in fact the *bazaar style* seemed to work

However, how can the quality of an open source software project be measured?

Meusuring Open Source Sofware quality

- A OSSP is built up from hundreds, sometimes thousands, of files

Meusuring Open Source Sofware quality

- A OSSP is built up from hundreds, sometimes thousands, of files
- To analyse manually a software project is a very hard and time consuming task

Meusuring Open Source Sofware quality

- A OSSP is built up from hundreds, sometimes thousands, of files
- To analyse manually a software project is a very hard and time consuming task

With that in mind

A system capable of automatically analysing and producing reports would enable users to make better choices, and developers to further improve their software.

Open Source Project Hosting Web Sites

Name	Established	Available VCS	Users	Projects	Alexa rank
SourceForge	1999	CVS SVN Bazar GIT Mercurial	2,000,000	236,319	136
GitHub	2008	GIT	505,000	1,516,000	742
Google Code	2006	SVN Mercurial	?	250,000	900
Code Plex	2006	SVN Microsoft TFS Mercurial	151,782	15.955	2,343
Assembla	2006	SVN GIT	180,000	60,000	6,628
Launchpad	2005	Bazar	1,140,345	19,016	12,466
BerliOS	2000	CVS SVN GIT Mercurial	47,285	5,448	17,299
Bitbucket	2008	Mercurial	51,600	27,769	12,047
Gitorious	2008	GIT	?	8,336	28,531

Assessing Open Source Software

IISO/IEC 912 software quality attributes

- **Functionality** (meeting of the functional requirements)
- **Reliability** (fault tolerance and recoverability)
- **Usability** (effort to learn, and operate the software)
- **Efficiency** (performance and resource consumption)
- **Maintainability** (effort to modify the software)
- **Portability** (effort to transfer to another environment)

Software Metrics

- Lines of Code

Software Metrics

- Lines of Code
- Cyclomatic Complexity

Software Metrics

- Lines of Code
- Cyclomatic Complexity
- Fan-In and Fan-Out

Software Metrics

- Lines of Code
- Cyclomatic Complexity
- Fan-In and Fan-Out
- Object-Oriented Metrics

Software Metrics

- Lines of Code
- Cyclomatic Complexity
- Fan-In and Fan-Out
- Object-Oriented Metrics
- Coding Standards?

Available Tools

- **FindBugs** find "bugs"in Java Programs
- **FxCop** reports information about the assemblies
- **PMD** scans Java source code for potential problems
- **PreFast** identifies defects in C/C++ programs
- **RATS** finds C/C++, Perl and Python security problems
- **SWAAT** scans Java, JSP, ASP .Net and PHP
- **Flawfinder** scans C and C++

Conclusion: Summary and Future Work

Karatsuba Algorithm (1960)

Divide and Conquer

Future Work

1. Chose a language: Ruby

Future Work

1. Chose a language: Ruby
2. Select consensual good and bad projects

Future Work

1. Chose a language: Ruby
2. Select consensual good and bad projects
3. Find patterns

Future Work

1. Chose a language: Ruby
2. Select consensual good and bad projects
3. Find patterns
4. Create a systematic way to measure it

Future Work

1. Chose a language: Ruby
2. Select consensual good and bad projects
3. Find patterns
4. Create a systematic way to measure it
5. Continue the work

Analysing and Measuring Open Source Projects

Questions?