

SIXTH FRAMEWORK PROGRAMME
PRIORITY 2
INFORMATION SOCIETY TECHNOLOGIES



**SIXTH FRAMEWORK
PROGRAMME**

FLOSSWORLD

**Free/Libre and Open Source Software: Worldwide
Impact Study**



Track 2 Study Report - Brazil

(Referred to as D23b in the work packages description in the proposal)

Project Reference: 015722

Kind of Project: Specific Support Action

Start Date: May 2005

End Date: June 2007



UNITED NATIONS
UNIVERSITY

UNU-MERIT



Universidad
Rey Juan Carlos

Authors and contributors

Coordination, Drafting and Edition

Jesús M. González-Barahona, GSyC/Libresoft, Universidad Rey Juan Carlos
Juan Luis Prieto, GSyC/Libresoft, Universidad Rey Juan Carlos
Daniel Izquierdo-Cortázar, GSyC/Libresoft, Universidad Rey Juan Carlos
Álvaro del Castillo, GSyC/Libresoft, Universidad Rey Juan Carlos

Reviewing

Gregorio Robles, GSyC/Libresoft, Universidad Rey Juan Carlos
Teófilo Romera, GSyC/Libresoft, Universidad Rey Juan Carlos

Retrieval process and Drafting

Roberto Andradás-Izquierdo, GSyC/Libresoft, Universidad Rey Juan Carlos
Jorge Gascón-Pérez, GSyC/Libresoft, Universidad Rey Juan Carlos
Daniel Izquierdo-Cortázar, GSyC/Libresoft, Universidad Rey Juan Carlos
Juan Luis Prieto, GSyC/Libresoft, Universidad Rey Juan Carlos

Graphics

Juan Martínez-Romo, GSyC/Libresoft, Universidad Rey Juan Carlos

Local support and Reviewing

Julio Pereira, National Information Technology Institute (ITI) of the Office of the President
Marcelo Thompson, Oxford Internet Institute (University of Oxford)
Maurício Augusto Coelho, National Information Technology Institute (ITI) of the Office of the President
Sergio Amadeu, Professor, Faculdade Cásper Líbero.
Tadao Takahashi, Presidente do Programa Sociedade da Informação do Brasil
Diana Hay, National Information Technology Institute (ITI) of the Office of the President
Renato Martini, National Information Technology Institute (ITI) of the Office of the President
Rubens Queiroz de Almeida, University of Campanias

Copyright

©2007 GSyC/LibreSoft

Some rights reserved. This report is distributed under the Creative Commons Attribution-ShareAlike 3.0 licence, available in <http://creativecommons.org/licenses/by-sa/3.0>

This report is available in <http://flossworld.org>

This report has been funded by the European Commission under contract number FP6-IST-015722.

Disclaimer

The opinions expressed in this Study are those of the authors and do not necessarily reflect the views of the European Commission. Contract FP6-IST-015722

Contents

1	Executive Summary	5
2	Introduction	6
2.1	Details of the considered region	6
2.2	Summary of results	8
2.3	Problems and constraints found	10
2.4	Details of the methodology	10
2.5	Contributions	10
3	Results	11
3.1	General information	11
3.2	Forges information	12
3.3	SCM information	12
3.4	Source code information	15
3.5	Mailing list	17
3.6	Authorship information	18
3.7	SourceForge Analysis	19
4	Appendixes	21
4.1	General information	21
4.2	Data from repositories	22
4.3	Sourceforge analysis	25
4.4	Source Code	29

List of Tables

2.1	Brazilian forges	8
2.2	Registered users and projects in Brazilian forges (30th of May 2007). SourceForge has been included for completeness (data June, 2006).	8
2.3	Information sources that could be extracted from Brazilian forges (April 2007).	8
3.1	Data collected by Brazilian partners	11
3.2	Committers with the most commits	13
3.3	Committers and commits for some selected projects in Brazil	13
3.4	Programming Languages used in <i>CodigoLivre</i> forge	15
3.5	Top 10 of mailing lists (data 13 March 2007)	17
3.6	Top 10 of mailing list posters (data 13 March 2007)	17
3.7	Categories of authorship of source code files	18
3.8	The most active Brazilian projects detected on SourceForge	19
4.1	Brazilian Communities	21
4.2	SCM results in Brazilian forges	22
4.2	SCM results in Brazilian forges	23
4.2	SCM results in Brazilian forges	24
4.2	SCM results in Brazilian forges	25
4.3	SCM results in SourceForge forge	25
4.3	SCM results in SourceForge forge	26
4.3	SCM results in SourceForge forge	27
4.3	SCM results in SourceForge forge	28
4.3	SCM results in SourceForge forge	29
4.4	Basic Cocomo model applied to software releases detected in Codigo Livre forge	29
4.4	Basic Cocomo model applied to software releases detected in Codigo Livre forge	30
4.4	Basic Cocomo model applied to software releases detected in Codigo Livre forge	31

List of Figures

2.1	Brazil World Map	6
2.2	Projects and users per forge	9
3.1	General information obtained from surveys (comparison among countries)	12
3.2	Committers per project	14
3.3	Commits per project	14
3.4	Progammig languages used in <i>CodigoLivre</i> forge	16
3.5	Distribution of lines of code per project in <i>CodigoLivre</i> forge	16
3.6	Posters and Messages per Mailing List	18
3.7	Authorship	19
3.8	Distribution of commits per project in SourceForge for Brazilian developers	20

Chapter 1

Executive Summary

This report shows the main results of the analysis performed, in the context of the FLOSSWorld project, of some quantitative aspects of libre (free, open source) software in Brazil. The sources of the analysis have been an exhaustive data retrieval of several facts related to libre software in that country (such as a list of Linux user groups, magazines focused on libre software, etc.) and a detailed, quantitative data mining of sites hosting libre software development (forges).

In Brazil there are 5 local forges which are quite popular among the libre software community. From these 5 forges, the most important one is *CodigoLivre* which has around 14200 registered users and 1866 projects. Then comes *LuaForge* with around 800 users and 250 projects and *AgroLivre* with 615 users and just 16 projects. On the other hand, we do not have a lot of information about the last two, *Incubadora Virtual* and *QuerenciaLivre*. Also Brazilian developers use SourceForge to collaborate in libre software projects and to develop Brazilian projects as well. We identified 851 Brazilian projects in SourceForge and it is remarkable that *CodigoLivre* has even more.

In *CodigoLivre* there are 96 mailing lists including general mailing lists and specific projects mailing list having them around 1667 different posters. We can say that this mailing lists are very active because they had a total of more than 5700 messages posted when the data was retrieved, in January 2007.

It is worth to mention that there are at least 9 enterprises currently submitting code to the forges in Brazil. This shows a certain interest from Brazilian industries in SourceForge.

Like in some of the other countries in the Flossworld study, *C* is the most popular programming language among the Brazilian developers with the 82% (around 4,390,322 lines) of the total lines of code of the software.

Comparing the Source Code Management systems of the local forges and SourceForge we see that projects hosted in local forges have a valid SCM tool but only around 640 of the 1840 use it to develop. Only 643 developers use their SCM account regularly. Also, there are 300 projects in SourceForge with a CVS repository in use. About 414 developers are working on these repositories.

But there is a big difference in projects hosted in SourceForge. Most of the commits in SourceForge are being made in *tdrsoft*, a project which has more than 25800 commits, more than three times the next project.

Chapter 2

Introduction

Within the context of the FLOSSWorld project, this report is devoted to the quantitative study of libre software development in Brazil. The information presented here is based mainly in public data found in the repositories of libre software projects hosted in the analyzed area, in SourceForge (the largest hosting site for libre software projects) and in a survey completed with the help of the partners in FLOSSWorld.

The data found in repositories (usually identified by the local partners in the project) has been downloaded, stored in a database, and later carefully mined and analysed (using a semi-automatic process that has been complemented by human validation). Most of the data used was obtained from source code, source code management systems and mailing list archives. In addition to repositories in the region, SourceForge has also been analyzed, as the largest hosting site, worldwide, for libre software projects. The details of the methodology used are specified in a separate document, also produced by the FLOSSWorld project, the “Metodology report”.

In the following chapters, the main results produced by this methodology are shown. Before that, in this chapter, some details about the methodology itself are discussed.

2.1 Details of the considered region

Geographic area.

Brazil, officially the Federative Republic of Brazil, is the largest and most populous country in South America and the fifth largest in the world in both area and population. Brazil has an extension of 8,514,877 km^2 . The 0.65% of its total extension is occupied by water.

Spanning a vast area between central South America and the Atlantic Ocean, it is the easternmost country of America and borders every other South American country other than Ecuador and Chile (viz. Uruguay, Argentina, Paraguay, Bolivia, Peru, Colombia, Venezuela, Guyana, Suriname and the French department of French Guiana).



Figure 2.1: Brazil World Map

Time zone.

Brazil's time zone is between UTC -2 and UTC -5, but UTC -3 is used officially. Calculating time zones is very useful when studying developers. This is because some developers do not say where they are from when they register in a forge. But usually, they put their time zone instead. With the time zone, and using some heuristics it is possible to guess the nationality of developers.

Other countries that share the time zone with Brazil are:

1. The East part of The United States.
2. Argentina.
3. Groenland.
4. Venezuela.

[Both the description of the geographic area and the map where obtained in the Wikipedia¹]

Demographic information.

The estimated population of Brazil (2006) is 188,078,261 people, being the fifth most populated country of the world.

Population density is about $22/km^2$. Brazil figures in the 182nd place in the list of countries by population density.

Economic information.

Brazil's Gross Domestic Product is \$1,577 trillion (2005 estimation). This puts it in the 9th place in the list of countries by GDP². These values are derived using PPP³.

Using values above, calculated GDP (PPP) per capita is \$8,584, putting Brazil in 68th place into List of countries by GDP (PPP) per capita.

Using nominal values Brazil has a nominal GDP of \$795,666 Billion, putting Brazil in 10th place into List of countries by GDP (nominal) and in 74th place into List of countries by GDP (nominal) per capita, with a renta per capita of \$4,320. This list has total of 182 countries.

Internet top level domain.

The Brazilian TLD⁴ (Domain Name System Top Level Domain, which is .br) is an important indicator to recover information from Brazilian email addresses. If someone has a Brazilian TLD, she is supposed to be a Brazilian user or developer.

¹<http://en.wikipedia.org/wiki/Brazil>

²Gross Domestic Product: the value of all final goods and services produced within a nation in a given year.

³Purchasing Power Parity: is a mean that represents the relative cost of living and the inflation rates of the countries.

⁴methodology report reference

2.2 Summary of results

The Brazilian FLOSSWorld partners have identified five forges that could be considered as Brazilian local forges. These forges have been spidered in order to identify the projects they contain. Table 2.2 lists a relation of the number of projects and users registered in each of the forges. The world-wide most popular forge, SourceForge, has been added to the table as many Brazilian developers and Brazilian-driven projects have been found there. The number of registered users at SourceForge gives only the estimation of Brazilian developers⁵. The 2560 projects in SourceForge that are Brazilian-driven have a majority (i.e. more than 50%) of Brazilian developers in their teams.

Forge	Forge name
http://codigolivre.org.br/	CodigoLivre
http://luaforge.net/	LuaForge
http://repositorio.agrolivre.gov.br/	AgroLivre
http://iv.incubadora.fapesp.br/	IncubadoraVirtual
http://querencialivre.rs.gov.br/	QuerenciaLivre

Table 2.1: Brazilian forges

Forge	Registered users	Registered Projects
CodigoLivre	14200	1866
LuaForge	802	250
AgroLivre	615	16
IncubadoraVirtual	ND	ND
QuerenciaLivre	ND	ND
Sourceforge	21291	851

Table 2.2: Registered users and projects in Brazilian forges (30th of May 2007). SourceForge has been included for completeness (data June, 2006).

It is important to point out that not all registered users are active developers in the forges. Many of them could register and never join a development project, for instance. Projects, as well, may not make use of all development-related tools offered by the forges (and therefore, will not offer data from several data sources). This fact is shown in detail in the summary table 2.3, where the number of SCM (CVS/SVN) repositories, mailing lists, software releases and size of the software is given.

This fact is specially dramatic for Brazil. Being one of the countries of the study with the most local forges (5), it is specially disappointing that not much data could be extracted from them.

Forge	SCM repos	Committers	Commits	MailingLists	Releases	SLOC
CodigoLivre	1950	319	145,932	96	89	5,330,730
LuaForge	ND	ND	ND	ND	ND	ND
AgroLivre	ND	ND	ND	ND	ND	ND
IVirtual	ND	ND	ND	ND	ND	ND
QLivre	ND	ND	ND	ND	ND	ND
SourceForge	300	414	166,815	ND	ND	ND

Table 2.3: Information sources that could be extracted from Brazilian forges (April 2007).

⁵Methodology report: How to obtain nationality from Sourceforge

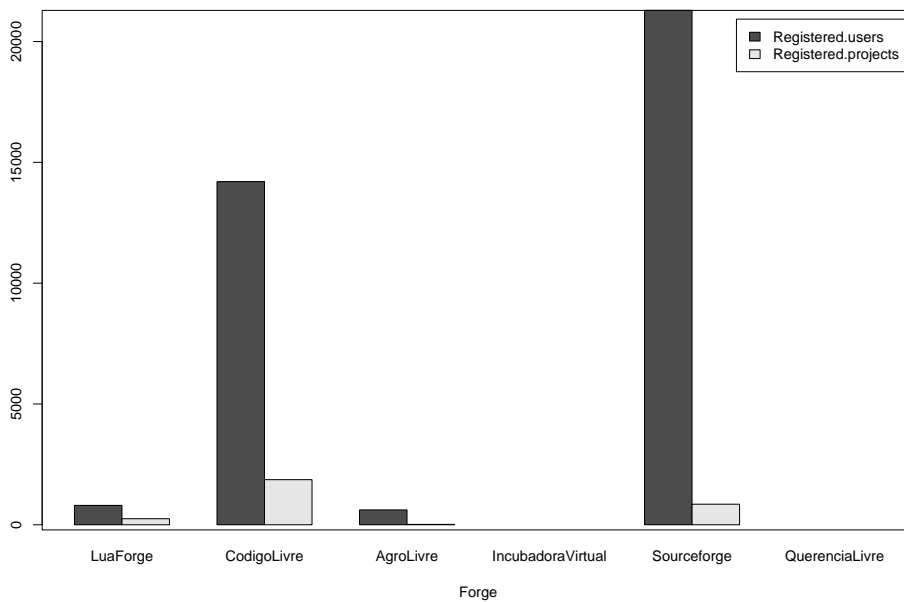


Figure 2.2: Projects and users per forge

The next section describes the set of difficulties arisen during the data retrieval process. These difficulties explain the *ND* (No-Data) value that appears in so many cells of table 2.3.

Results presented in table 2.3 are: the field *SCM repos* which is the number of Source Code Management systems, the number of mailing lists, the number of code releases and the number of lines of code in these releases.

2.3 Problems and constraints found

In Brazil there is a big free software development movement. We have found five different local Brazilian forges. Unfortunately only one of them was suitable for our tools to analyse it, *CodigoLibre*. But an added problem for data gathering appeared in *CodigoLibre*. The only way to gain access to the Source Code Management systems in *CodigoLibre* is by asking the site administrators for a valid account, since software repositories are not open to anonymous users. The rest of the data did not represent major problems, since most of it is publicly available.

Nobody in the data gathering or data mining teams was able to read Portuguese. Since Portuguese is the language used in all of the local Brazilian forges, sometimes, apparently simple operations took longer than expected.

2.4 Details of the methodology

This section presents a brief description of the different tools used in the data retrieval process.

Two forges have been analysed, Sourceforge and *CodigoLivre*. This is why data comes from different data sources. The analysis was made on January of 2007 and different tools have been used in order to make the gathering of data as automatic as possible.

1. CVSanaly⁶: CVSanaly was used to retrieve data from Sourceforge's and CodigoLivre's software repositories. Some repositories were protected by passwords. So we needed the help of the site administrators in these cases and they were always pleased to help.
2. PyTernity⁷: PyTernity can be run on source code form software releases downloaded by hand. It obtains information about the authorship of the source code.
3. SLOCCCount⁸: SLOCCCount can be run on the same files as PyTernity. SLOCCCount offers data about the size of the software, such as lines of source code.
4. Spider⁹: The Spider tool was able to get information about *LuaForge* and *CodigoLivre*. While data from *AgroLivre* was retrieved by hand because there are just 14 registered there.

2.5 Contributions

This report about the situation in Brazil has been drafted by the GSyC/LibreSoft team of the Universidad Rey Juan Carlos (Madrid, Spain, coordinated by Jesus M. Gonzalez-Barahona), who has also performed the data retrieval, mining and analysis, and produced its final version. In addition to the general collaboration by all partners, two Brazilian FLOSSWorld partners have contributed specifically (ITI and CAMP, University of Campanias) by providing information about libre software developing sites, feedback about the drafts of the report, and many other useful information, and generic help with issues specific to their region (including assistance with language issues). The coordinator of FLOSSWorld (UNU-MERIT) has also provided specific assistance and feedback during all the stages of preparation of this report.

⁶Methodology Report - Chapter Tools - Section CVSanaly

⁷Methodology report - Chapter Tools - Section PyTernity

⁸Methodology report - Chapter Tools - Section SLOCCCount

⁹Methodology report - Chapter Tools - Section Spider

Chapter 3

Results

3.1 General information

Brazilian partners collaborated providing different sources of useful data for the report¹. For more information, please refer to section 4.1 where this data is detached.

There is a summary in table 3.4.

Region	Communities	Media	Platforms
Brazil	20	1	1

Table 3.1: Data collected by Brazilian partners

Where *Communities* means group of users interested in libre software, *Media* are any kind of journal magazine related to FLOSS and *Platforms* are web sites which provide any kind of support to the libre software world, such as forges. Also, in figure 3.1 there is a comparison of the information obtained from all the countries studied in FLOSSWorld.

¹Methodology report - Chapter Data Sources - Section Primary Data Sources

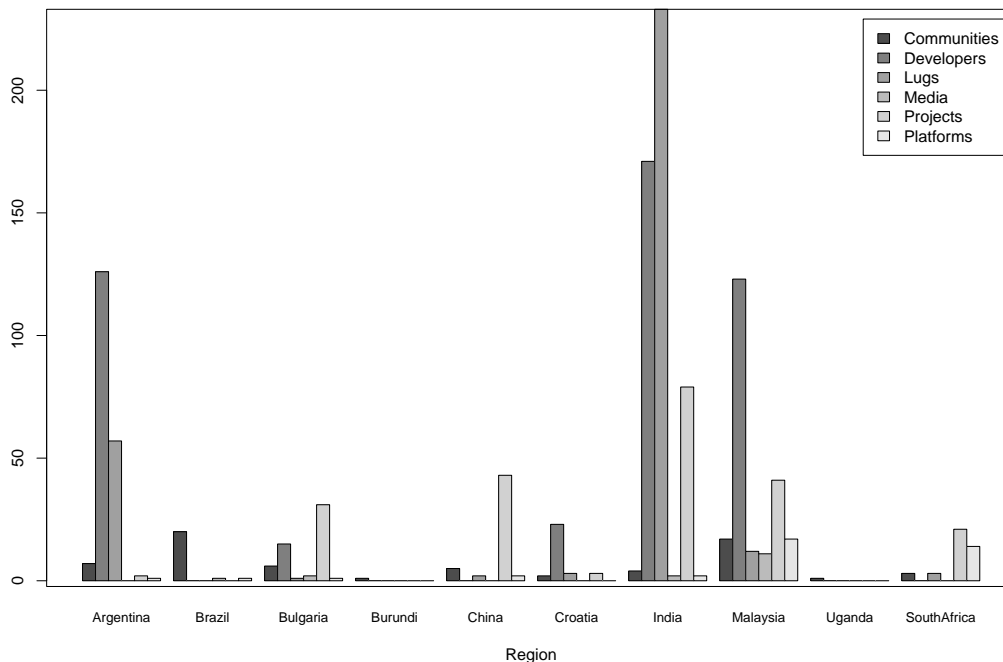


Figure 3.1: General information obtained from surveys (comparison among countries)

There are twenty free and open source software communities in Brazil which have a average of more than 100 participants each. All Brazilian communities' web sites are in Portuguese. This is quite a strange fact because in other countries the general rule is that these web sites are in English.

Linux Magazine is the name of a media found in Brazil. This magazine is oriented to different publics, from programmers and managers to general public. *Linux Magazine* has a monthly frequency and it is written in Portuguese. On its web² we can see the index of the magazine and some articles in pdf format. Other services are offered in the web page, such as obtaining a subscription for the magazine through the web page.

We also have data related to platforms and this will be presented in the following sections. As a summary we can tell that there are 5 local forges in Brazil, *CodigoLivre* being the biggest one with around 1950 repositories and 785 mailing lists.

3.2 Forges information

3.3 SCM information

In Brazil around 1842 projects are hosted in local forges which are quite popular compared to forges in other countries. This projects have Source Code Management systems (cvs or svn). But 1199 of them do not show any valuable information about committers. They appear to have just the *root* user, who makes all the commits. This may be because they are small projects with a very limited group of developers, maybe just one. The rest of them (643) seem to follow the classical approach.

We extracted some data from projects that used cvs (643) in Brazil. The commiter with the most commits is called *davidferreira* who did 10392 commits till the moment of data gathering. In the relation between project and committers it is good to mention that the projects with committers and the projects

²<http://www.linux-magazine.com.br/>

with more commits are more or less the same in this case. The project *Sagu* has 12 developers using the cvs and all of them sum around 11486 commits.

Table 3.2 and table 3.3 illustrate these facts.

Committer name	Num of commits
davidferreira	10392
beto	8897
jeff_otoni	8218
vgartner	6451

Table 3.2: Committers with the most commits

Project name	Number of committers	number of commits
sagu	12	11486
sgi	1	8218
pjmoo	1	7804
miolo2	2	5557
lfs_br	1	4248
tulip	2	4125
koizo	2	3899
sagu2	3	3873
vpe	1	3664
care2xbr	2	3149

Table 3.3: Committers and commits for some selected projects in Brazil

Figures 3.2 and 3.3 show committers per project and commits per project respectively.

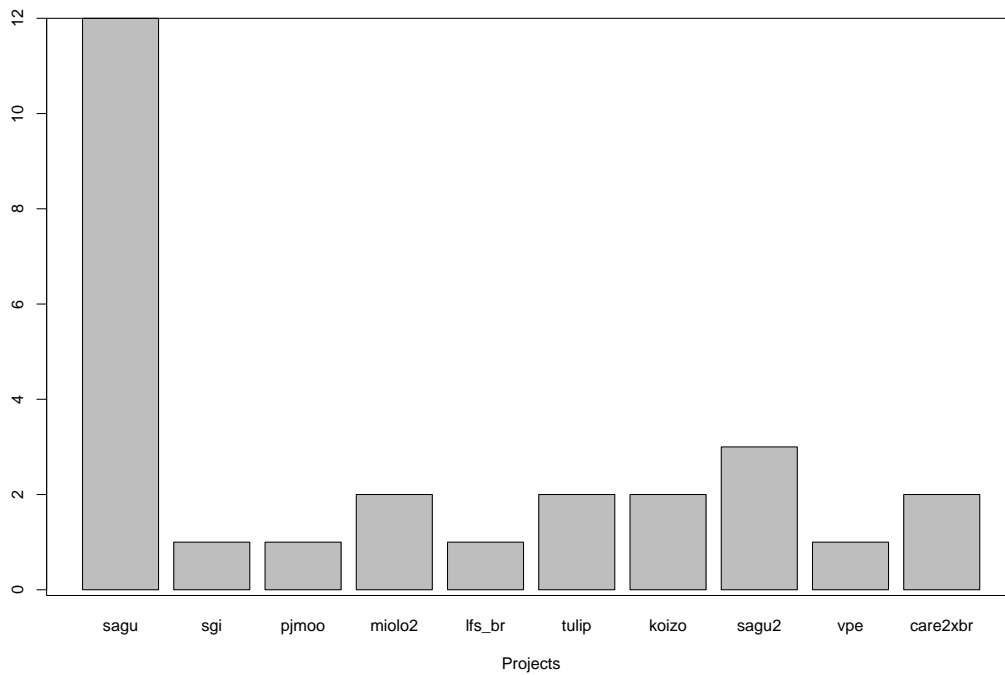


Figure 3.2: Committers per project

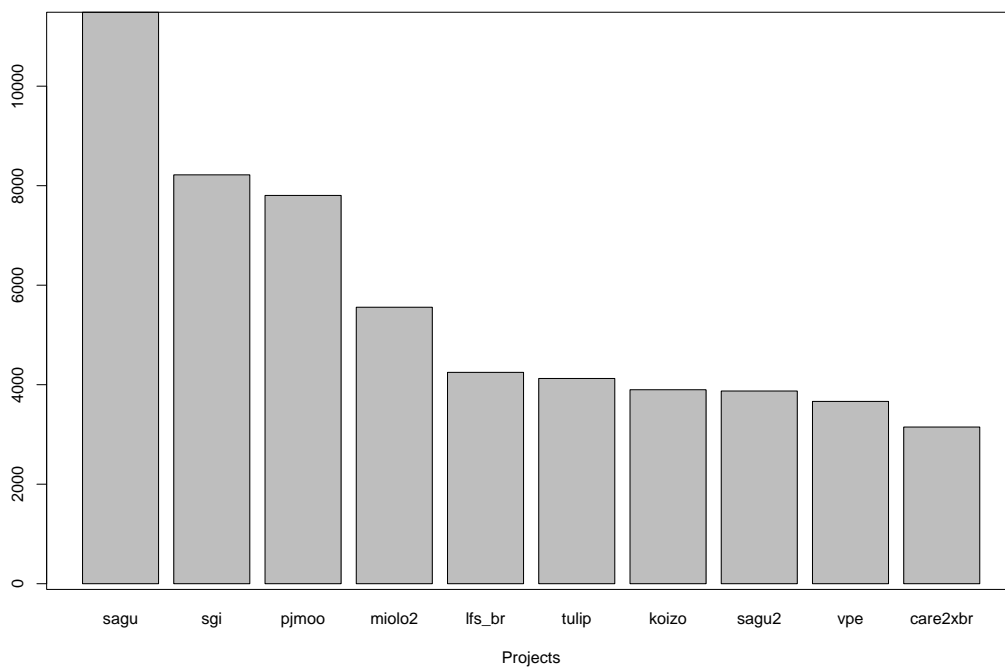


Figure 3.3: Commits per project

3.4 Source code information

We have found 89 Brazilian projects with relevant data. Data related to source code information like number of lines, estimated development cost, effort estimations or average developers per year.

For example, the average number of developers per project is 1.385 developer/year. The average number of code lines is 71.035 for each one, that suggests that there are a lot of projects with a very small number of lines. Programming languages usage data, clearly shows that *Ansi C* is the most popular programming language in the development of free software in Brazil.

Programming Language	Lines of Code	Percentage
C	4390322	82.3587
PHP	365592	6.8581
ASM	248461	4.6609
SH	63580	1.1927
C++	61519	1.1540
F90	58361	1.0948
Python	57034	1.0699
Pascal	44525	0.8352
Perl	22825	0.4281
Java	9851	0.1847
Yacc	6219	0.1166
Lex	1513	0.0283
Ruby	275	0.0051
Lisp	218	0.0040
C#	211	0.0039
Awk	99	0.0018
Sed	93	0.0017
TCL	32	0.0006
Total	5330730	100.0

Table 3.4: Programming Languages used in *CodigoLivre* forge

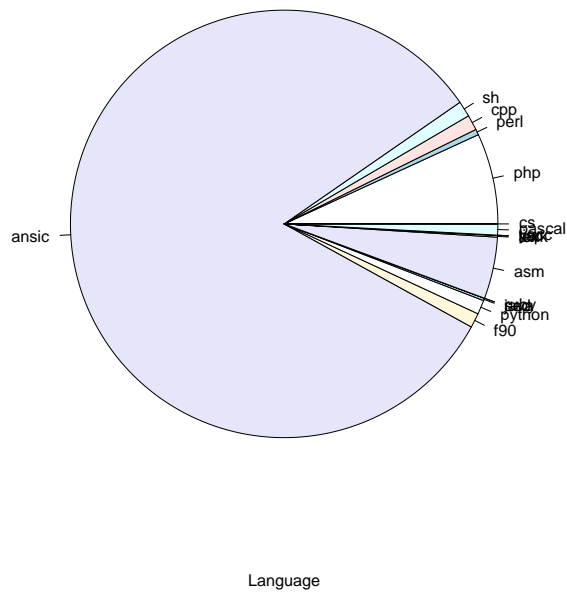


Figure 3.4: Programming languages used in *CodigoLivre* forge

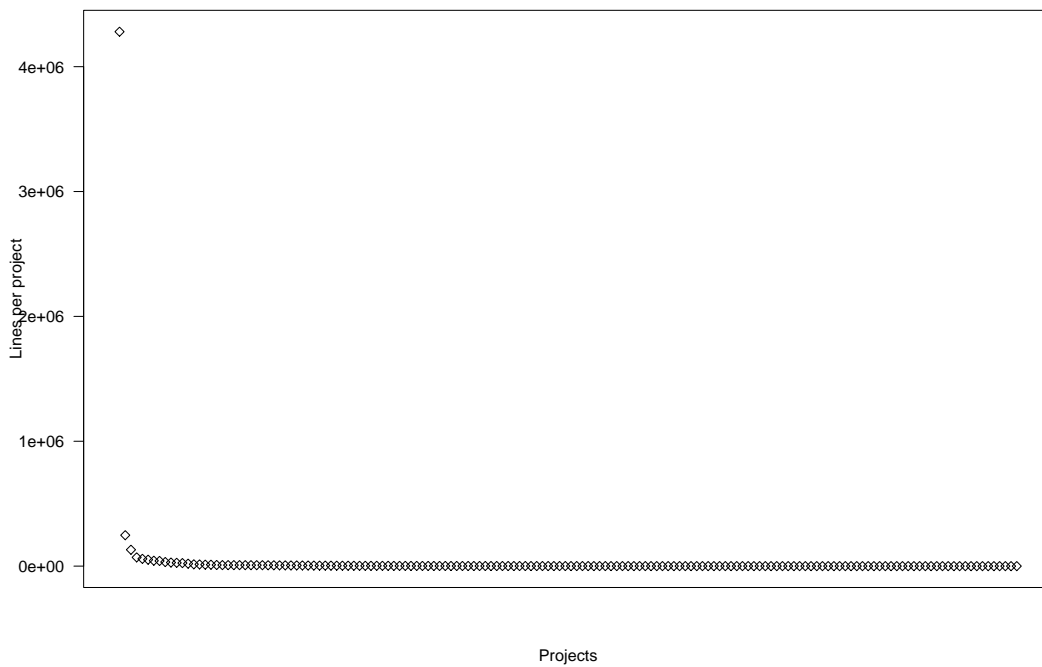


Figure 3.5: Distribution of lines of code per project in *CodigoLivre* forge

3.5 Mailing list

We have found 96 mailing lists in total in Brazil local forges, including local and global. These mailing lists belong to a total of 73 projects. In these 96 mailing lists 1667 unique posters have been found who had already sent 5709 unique messages by the time of the analysis (data 13 March 2007).

Project	Mailing List	Posters	Messages
gnuteca	gnuteca-devel	415	537
agata	agata-brasil	136	433
gnuteca	gnuteca-unal	119	119
gnuteca	gnuteca-users	91	268
agata	agata-english	90	253
pila	pila-devel	89	94
lfs	lfs-br-i18n	88	477
miolo	miolo-users	80	616
botbox	botbox-devel	75	79
agata	agata-espanol	65	107

Table 3.5: Top 10 of mailing lists (data 13 March 2007)

Poster alias	Messages posted
abntex-pub@server.codigolivre.org.br	678
noreply@users.codigolivre.org.br	267
miolo-users@codigolivre.org.br	232
lizardo@linuxfromscratch.org	101
miolo-i18n@server.codigolivre.org.br	99
pablo@solis.coop.br	94
pablo@daloglio.net	83
rodzadra@gmail.com	61
siages-devel@server.codigolivre.org.br	61
valcir@unicamp.br	58

Table 3.6: Top 10 of mailing list posters (data 13 March 2007)

Figure 3.6 shows number of posters and number of messages per mailing list.

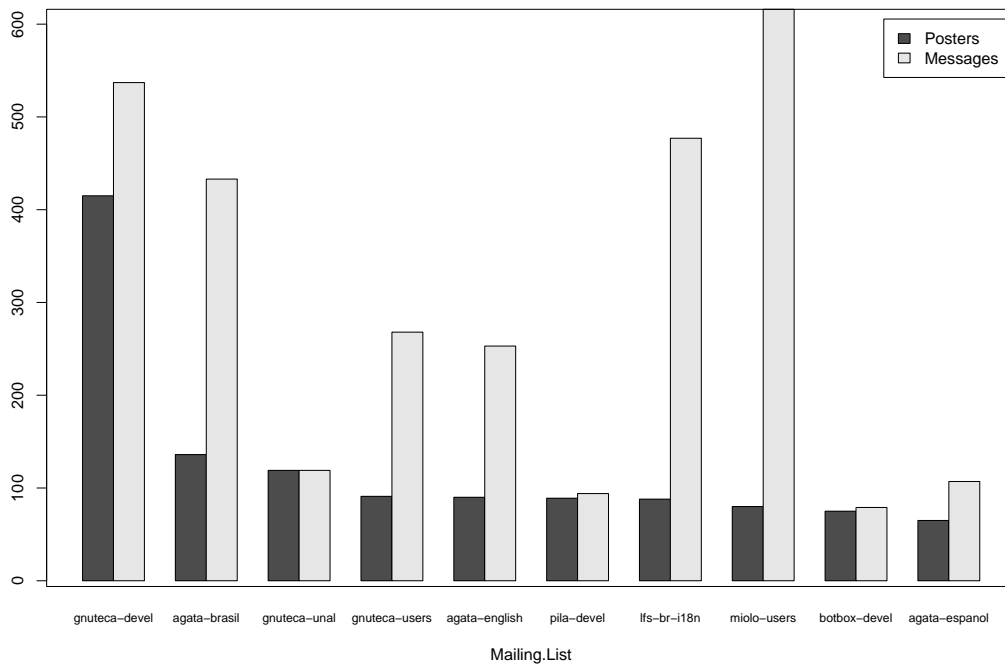


Figure 3.6: Posters and Messages per Mailing List

3.6 Authorship information

Looking to authorship results obtained by parsing source code with PyTernity³ is possible to see that, like in all countries, individual authors are responsible for the most files. Followed by enterprises, which suggests that industry is taking a relevant role in the Brazilian development scenario. Table 3.7 and figure *fig:authorship* show this information.

Type of author	Detected number
Individual authors	26
Team-Groups	5
Universities	1
Enterprises	9
Foundation or public entities	1
Unknown	4

Table 3.7: Categories of authorship of source code files

³Methodology Report - Chapter Tools - PyTernity

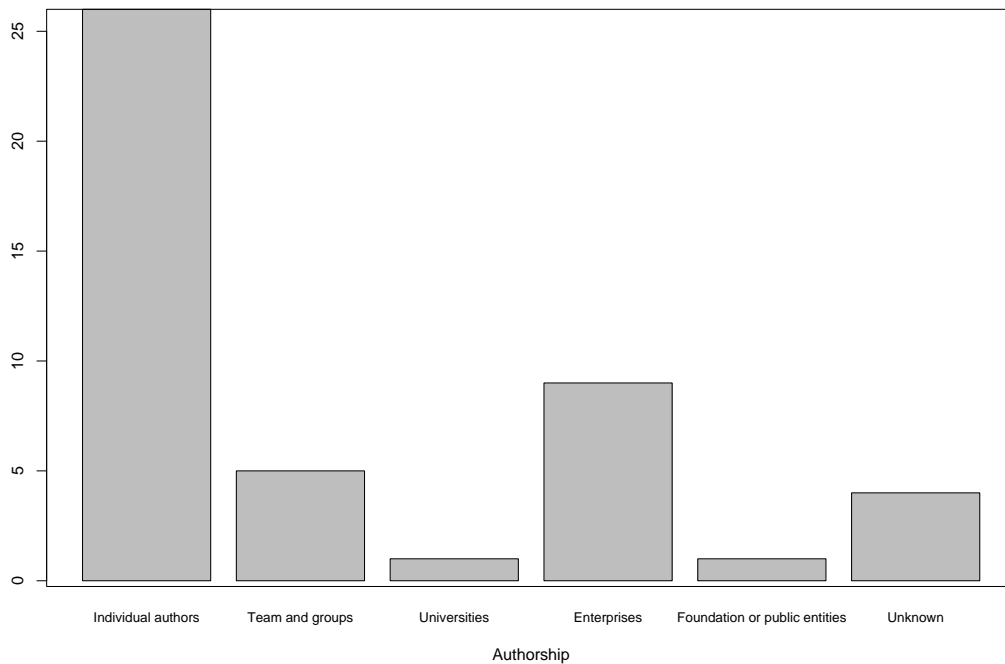


Figure 3.7: Authorship

3.7 SourceForge Analysis

Brazil has a big number of projects in SourceForge, more than other country included in FLOSSWorld. There are around 256 projects in SourceForge that are identified as Brazilian-driven. Unfortunately only around 300 of them have an active CVS repository which they use regularly. These projects with CVS information show an average of 322 commits. This is because only a very few of them have less than 100 commits and only 20 have less than 10.

The project with the most commits is *tdrsoft* that has 25825 commits. That is more than three times the commits in the second one, *chessd* with 7849.

There are 414 developers working in Brazilian projects which use CVS. Moreover, 111 of them collaborate in more than one project at the same time, but there are only 9 developers in the project with the most developers.

In Table 3.8 we can see the top 5 of the projects which use the CVS..

Project name	Num of commits
tdrsoft	25825
chessd	7849
sntool	3506
bioma	3188
beholder	3036

Table 3.8: The most active Brazilian projects detected on SourceForge

Figure 3.8 shows a distribution of commits per project in SourceForge for Brazilian developers.

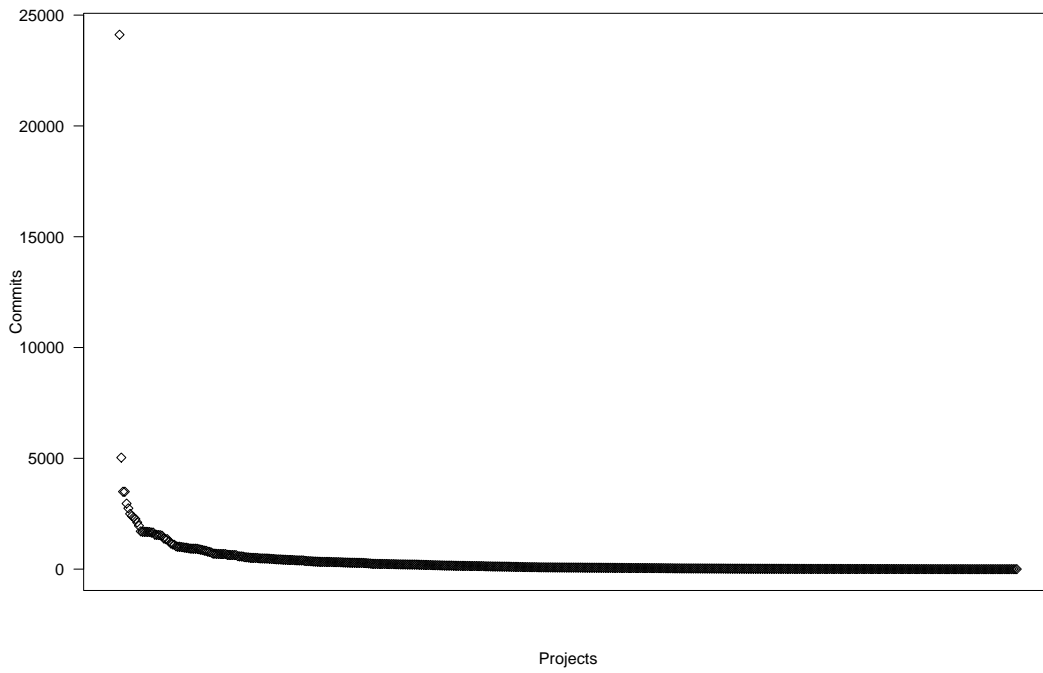


Figure 3.8: Distribution of commits per project in SourceForge for Brazilian developers

Chapter 4

Appendixes

4.1 General information

Table 4.1: Brazilian Communities

Name	Size	URL
PSL-Brasil	large < 100	http://www.softwarelivre.org/
PSL-Rio de Janeiro	large ; 100	
PSL-Bahia	n/a	http://twiki.im.ufba.br/bin/view/PSL
PSL-MG – Projeto Software Livre Minas Gerais	large < 100	http://psl-mg.softwarelivre.org/tiki/tiki-index.php
Projeto Software Livre Rio Grande do Sul	very large < 1000	http://psl-rs.softwarelivre.org/
Projeto Software Livre Mulheres	large < 100	http://mulheres.softwarelivre.org/
Projeto LinuxChix Brasil	large < 100	http://www.linuxchix.org.br/
PSL-AM - Projeto Software Livre Amazonas	n/a	
PSL-DF - Projeto Software Livre do Distrito Federal	n/a	
Usuarios Debian Sao Paulo	n/a	
Grupo de Usuarios Debian de Roraima	n/a	
Projeto Software Livre Pernambuco	n/a	http://psl-pe.softwarelivre.org/
Comunidade Sol	n/a	http://www.comunidadesol.org/
Grupo de Usuarios GNUrias	small < 15	http://www.gnurias.org.br/
Grupo de Usuarios Debian-PR (GUD-PR)	n/a	http://www.debian-pr.org/pmwiki.php
Grupo de Usuarios Debian GNU Linux Rio de Janeiro	n/a	http://www.debian-rj.org/pmwiki.php
Grupo de Usuarios Debian do Rio Grande do Sul	n/a	http://www.debian-rs.org/
Comunidade PHPNuke Brasil	n/a	http://phpnuke.org.br/
Projeto Software Livre Goias	n/a	http://twiki.softwarelivre.org/bin/view/PSLGO/WebHome
Projeto Mono-Brasil	n/a	http://monobrasil.sl.org.br/

4.2 Data from repositories

The following table shows Brazilian projects which are in local forges that have more than 100 commits in the SCM. There are around 490 projects with less than 100 commits and most of them, around 450, even have less than 50 commits. So they are small projects with a very small number of developers, usually 1 or 2.

Table 4.2: SCM results in Brazilian forges

Project name	Number of commits
sagu	11486
sgi	8218
pjmoo	7804
miolo2	5557
lfs_br	4248
tulip	4125
koizo	3899
sagu2	3873
vpe	3664
care2xbr	3149
xoopsbrasileiro	2858
fiascodobrasil	2734
geoagents	2720
investimento	2410
eflow	2345
tdrsoft	2344
orbe	2165
commando	1938
boaideia	1766
naweb	1724
codejun	1631
pslib	1628
factprocess	1444
fileboard	1274
centru	1265
kdedevelbr	1247
naprosys	1240
frameworkgsw	1225
lagartixacms	1192
miolo	1187
gnuteca	1169
mensagemweb	1159
wel	1144
diretolivre	1128
afabrica	1088

Continued on next page

Table 4.2: SCM results in Brazilian forges

Project name	Number of commits
mercury	1072
abntex	1011
cdlivre	994
instalador_ll	992
cbov	943
nossoxoops	921
pythondoc	898
jsms	882
gstee	831
projetoenix	805
pila	796
serverpublisher	766
othelloweb	733
dbexplorer	705
gnudata	693
projetolinuxbsd	686
pensandoemjava	651
jurisbac	632
arebr	622
gstgc	591
kpreg	584
emailer	574
eventor	570
modularpsi	558
oeme	537
jcvscodigo	526
bislge	521
theblog	510
openbound	480
bugio	435
pit	399
anthurium	395
saite	381
ibquota	357
classgenerator	347
mmc	340
manualphp	340
sqlitebrasil	338
sappetrinets	337
sigae	335
jmessenger	332
simumems	314
sgcpd	295
labjor	289

Continued on next page

Table 4.2: SCM results in Brazilian forges

Project name	Number of commits
jservermonitor	288
sagu2ufr	287
mioloide	286
sgqi	283
framecalixto	282
vitra	278
gerar	271
xadresco	268
omega	260
formalina	257
iarajs	256
dt2j	254
notes	253
pgdocs	253
gnucom	244
jetlearning	237
scotty	236
opaweb	235
simulife	232
jabba	228
gecon	218
fred	206
ltsp_doc_pt_br	204
sectools	203
jexecsql	190
portal_lsn	190
losango	187
modus	187
rau_tu	181
guaranix	179
declarafacil	175
myportfolio	171
perto	170
gedesolcpg	168
antropologia	167
elitecache	167
paintestlab	166
dukenuke	160
viaipadmin	158
powered	155
autublave	154
ticket_it	153
compuclass	150
lesp	150
Continued on next page	

Table 4.2: SCM results in Brazilian forges

Project name	Number of commits
sape	150
sultan	143
cgti	139
dform	138
quila	134
mobisim	132
schambeck	129
sintegra	128
grana	126
gruj_truco	125
jet	124
batalhao	123
sapes1	122
msxk7	120
gawnuts	120
contsoft	119
g-g	112
pgmanager	112
gtk_akount	108
isk	108
guara	108
papillon	106
fr0gsms	106
biblios	104
watcher	104
kompass	103
bizu	102
jerp	101
queuesimulator	101

4.3 Sourceforge analysis

The next table shows Brazilian sourceforge projects that have more than 100 commits in the SCM.

Table 4.3: SCM results in SourceForge forge

Project name	Num of commits
tdrsoft	25825
chessd	7849
sntool	3506
bioma	3188
Continued on next page	

Table 4.3: SCM results in SourceForge forge

Project name	Num of commits
beholder	3036
zige	2784
lmdsq	2523
kildclient	2495
protux	2423
planeta	2395
xmlnuke	2376
nntk	2326
ractiontools	2220
vmailadmin	2150
phporb	1956
calpersystem	1938
pantheon	1784
s2ilib	1776
guj	1650
xhgtk	1595
zcomix	1587
arara	1539
brainbox	1536
modulos	1511
freemmg	1419
ocomonphp	1414
spiritdk	1388
openreality	1122
amphibian	1100
adaptweb	1091
freepki	1070
g6zip	1033
neucis	1031
sami	1015
cmssp	1015
maca	989
linuxstok	972
kofflinux	964
iflux	947
mesheditor	932
bigfish	929
kalc	923
deathworks3d	919
eventmanager	907
dcnigtmare	903
cmph	839
eclipsework	831
opentel	792
Continued on next page	

Table 4.3: SCM results in SourceForge forge

Project name	Num of commits
freelms	782
libglass	781
flick3d	769
pucerp	763
linfca	754
dracobuilder	734
mudufsc	713
athosframework	703
manualphp	684
notes2	683
phpvortex	682
kprojectfb	681
xhbdoept	657
pymorph	646
krico	638
epigrass	636
wct	635
cirts	610
treepropsheetex	577
mpide	565
revelationp	554
artofwar	551
hiberclipse	537
ultralife	509
ool	490
publicit	474
atom3d	472
gals	464
arteduca	456
arandu	449
highdb	445
cgiframeork	442
osgmtool	433
charba	431
lupa	430
shob	423
nikut	411
hotwork	403
lojanaweb	397
holo	397
arenax	376
urisc	351
mpbot	345
pcassembly	337

Continued on next page

Table 4.3: SCM results in SourceForge forge

Project name	Num of commits
libbx	332
breath	331
vassalo	330
openps	330
uts	324
tplusplus	323
nrp	322
luapalms	319
tupines	319
arca	319
uebimiau	317
livecd	313
asxa	313
filedonkmanager	297
minihelpdesk	293
gmdb	288
pwgp	287
beamblocks	285
ucpapyrus	284
manaproject	247
aircbot	243
flogware	239
cmdevel	238
abstractplugin	237
blogware	233
frebsdcc	232
kaffeinesuite	226
fourwaychess	218
blogamp	217
geniusvp2	216
replicadorbr	216
pybber	212
mindelo	211
threedverse	207
mogno	199
lula	189
linxtris	182
bestoffice	178
dolly	176
dsateclipse	168
flowcontrol	160
goldrec	154
incoop	151
scrd	151
Continued on next page	

Table 4.3: SCM results in SourceForge forge

Project name	Num of commits
avrprog	150
apdata	150
aspecies	148
xinca	147
phpdc	147
myfon	142
moraisaf	142
gnuadventure	140
dmml	139
liza	139
karuna	137
htmlcxx	132
turnchess	125
nerdservices	123
nfswap	122
slackpkg	121
open2support	121
dbfw4j	121
gbible	119
pluggathon	118
sentencedetect	118
kantata	116
sendymail	113
oky	106
holeofbugs	102
uml4j	101

4.4 Source Code

Table 4.4: Basic Cocomo model applied to software releases detected inCodigo Livre forge

Project	Num Lines	Effort D/Y (D/M)	Schedule	Avg num of developers	Cost
kernelbr	4,540,063	1,383.39 (16,600.69)	8.36 (100.37)	165.40	\$ 186,877,320
sgi	132,528	33.84 (406.10)	2.04 (24.50)	16.57	\$ 4,571,562
soc	73,691	18.27 (219.28)	1.62 (19.39)	11.31	\$ 2,468,464
brams	59,708	14.65 (175.81)	1.49 (17.83)	9.86	\$ 1,979,137
sagu-pro	52,253	12.74 (152.84)	1.41 (16.90)	9.04	\$ 1,720,515

Continued on next page

Table 4.4: Basic Cocomo model applied to software releases detected inCodigo Livre forge

Project	Num Lines	Effort D/Y (D/M)	Schedule	Avg num of developers	Cost
siages	43,387	10.48 (125.73)	1.31 (15.69)	8.01	\$ 1,415,368
lotonyckx	42,089	10.15 (121.78)	1.29 (15.50)	7.85	\$ 1,370,941
sagu2	37,584	9.01 (108.13)	1.23 (14.82)	7.30	\$ 1,217,292
sagu	28,855	6.83 (81.93)	1.11 (13.34)	6.14	\$ 922,303
iwd	26,609	6.27 (75.25)	1.08 (12.91)	5.83	\$ 847,074
tk3000e	26,267	6.19 (74.23)	1.07 (12.85)	5.78	\$ 835,647
orbe	24,495	5.75 (68.98)	1.04 (12.49)	5.52	\$ 776,556
kpreg	17,392	4.01 (48.15)	0.91 (10.90)	4.42	\$ 542,012
pslib	17,317	3.99 (47.93)	0.91 (10.88)	4.41	\$ 539,558
tk2000	17,178	3.96 (47.53)	0.90 (10.84)	4.38	\$ 535,011
lxrad	14,838	3.40 (40.75)	0.85 (10.23)	3.98	\$ 458,760
gtkdbc	11,885	2.69 (32.28)	0.78 (9.36)	3.45	\$ 363,405
docsproceedings	11,832	2.68 (32.13)	0.78 (9.34)	3.44	\$ 361,703
kompass	11,612	2.63 (31.50)	0.77 (9.28)	3.40	\$ 354,645
klavaro	9,580	2.15 (25.74)	0.72 (8.59)	3.00	\$ 289,785
linuxsoho	8,418	1.87 (22.47)	0.68 (8.16)	2.75	\$ 252,994
emailer	8,192	1.82 (21.84)	0.67 (8.07)	2.71	\$ 245,867
guarani	8,145	1.81 (21.71)	0.67 (8.05)	2.70	\$ 244,387
pila	8,141	1.81 (21.70)	0.67 (8.05)	2.70	\$ 244,261
gtk-br	8,110	1.80 (21.61)	0.67 (8.04)	2.69	\$ 243,284
gnuteca	7,814	1.73 (20.78)	0.66 (7.92)	2.62	\$ 233,969
projetoenix	7,102	1.57 (18.80)	0.64 (7.62)	2.47	\$ 211,637
sqlitebrasil	6,394	1.40 (16.84)	0.61 (7.31)	2.30	\$ 189,541
rau-tu	5,966	1.30 (15.66)	0.59 (7.11)	2.20	\$ 176,242
sgqi	5,480	1.19 (14.32)	0.57 (6.87)	2.08	\$ 161,199
apt-br	5,301	1.15 (13.83)	0.57 (6.78)	2.04	\$ 155,674
prolioli	4,306	0.93 (11.12)	0.52 (6.24)	1.78	\$ 125,147
gconjugue	4,065	0.87 (10.46)	0.51 (6.10)	1.72	\$ 117,803
ibquota	3,684	0.79 (9.44)	0.49 (5.87)	1.61	\$ 106,237
ipprotect	2,981	0.63 (7.56)	0.45 (5.39)	1.40	\$ 85,059
3ljcota	2,703	0.57 (6.82)	0.43 (5.18)	1.31	\$ 76,750
gawnuts	2,586	0.54 (6.51)	0.42 (5.09)	1.28	\$ 73,266
estante	2,516	0.53 (6.32)	0.42 (5.04)	1.26	\$ 71,185
quila	2,117	0.44 (5.27)	0.39 (4.70)	1.12	\$ 59,381
libhaircapture	2,052	0.43 (5.11)	0.39 (4.64)	1.10	\$ 57,468
afchecker	1,908	0.39 (4.73)	0.38 (4.51)	1.05	\$ 53,241
myportfolio	1,548	0.32 (3.80)	0.35 (4.15)	0.91	\$ 42,747
eplfirewall	1,304	0.26 (3.17)	0.32 (3.88)	0.82	\$ 35,701
papillon	1,301	0.26 (3.16)	0.32 (3.87)	0.82	\$ 35,615
downtube	1,146	0.23 (2.77)	0.31 (3.68)	0.75	\$ 31,173
secretarianet	1,143	0.23 (2.76)	0.31 (3.68)	0.75	\$ 31,088
egret	986	0.20 (2.36)	0.29 (3.47)	0.68	\$ 26,620

Continued on next page

Table 4.4: Basic Cocomo model applied to software releases detected in Codigo Livre forge

Project	Num Lines	Effort D/Y (D/M)	Schedule	Avg num of developers	Cost
agendapessoal	982	0.20 (2.35)	0.29 (3.46)	0.68	\$ 26,507
kdedevelbr	945	0.19 (2.26)	0.28 (3.41)	0.66	\$ 25,459
tdwebmail	936	0.19 (2.24)	0.28 (3.40)	0.66	\$ 25,205
fr0gsms	859	0.17 (2.05)	0.27 (3.28)	0.62	\$ 23,032
scanmsg	778	0.15 (1.84)	0.26 (3.15)	0.58	\$ 20,757
rwadt	673	0.13 (1.58)	0.25 (2.98)	0.53	\$ 17,826
libini	667	0.13 (1.57)	0.25 (2.97)	0.53	\$ 17,659
othelloweb	664	0.13 (1.56)	0.25 (2.96)	0.53	\$ 17,576
brnews	662	0.13 (1.56)	0.25 (2.96)	0.53	\$ 17,520
protheus	650	0.13 (1.53)	0.24 (2.94)	0.52	\$ 17,187
xlance	646	0.13 (1.52)	0.24 (2.93)	0.52	\$ 17,076
tabserver	600	0.12 (1.40)	0.24 (2.84)	0.49	\$ 15,802
lampiao	530	0.10 (1.23)	0.23 (2.71)	0.46	\$ 13,872
gnutech	505	0.10 (1.17)	0.22 (2.65)	0.44	\$ 13,186
fwqos	460	0.09 (1.06)	0.21 (2.56)	0.42	\$ 11,955
jscsseditor	457	0.09 (1.05)	0.21 (2.55)	0.41	\$ 11,873
freenas	450	0.09 (1.04)	0.21 (2.54)	0.41	\$ 11,682
p4fs	370	0.07 (0.84)	0.20 (2.34)	0.36	\$ 9,512
ingres-utility	369	0.07 (0.84)	0.20 (2.34)	0.36	\$ 9,485
vivo24hxdsl	334	0.06 (0.76)	0.19 (2.25)	0.34	\$ 8,542
investimento	292	0.05 (0.66)	0.18 (2.13)	0.31	\$ 7,418
vortaro	279	0.05 (0.63)	0.17 (2.10)	0.30	\$ 7,072
gcbq	276	0.05 (0.62)	0.17 (2.09)	0.30	\$ 6,992
easyphp	266	0.05 (0.60)	0.17 (2.06)	0.29	\$ 6,726
transpolipo	263	0.05 (0.59)	0.17 (2.05)	0.29	\$ 6,647
oeme	259	0.05 (0.58)	0.17 (2.03)	0.29	\$ 6,540
sabada	237	0.04 (0.53)	0.16 (1.96)	0.27	\$ 5,958
testarede	236	0.04 (0.53)	0.16 (1.96)	0.27	\$ 5,932
radios	235	0.04 (0.52)	0.16 (1.96)	0.27	\$ 5,906
monoqle	211	0.04 (0.47)	0.16 (1.87)	0.25	\$ 5,274
dir2lame	179	0.03 (0.39)	0.15 (1.76)	0.22	\$ 4,437
passub	158	0.03 (0.35)	0.14 (1.67)	0.21	\$ 3,893
wl80211	149	0.03 (0.33)	0.14 (1.63)	0.20	\$ 3,660
gestometro	137	0.02 (0.30)	0.13 (1.58)	0.19	\$ 3,351
servconf	112	0.02 (0.24)	0.12 (1.46)	0.17	\$ 2,712
sql2txt	100	0.02 (0.21)	0.12 (1.39)	0.15	\$ 2,408
maibhe	69	0.01 (0.14)	0.10 (1.20)	0.12	\$ 1,631
bia	46	0.01 (0.09)	0.09 (1.02)	0.09	\$ 1,065
tde	67	0.01 (0.14)	0.10 (1.19)	0.12	\$ 1,581
kurumindev	35	0.01 (0.07)	0.08 (0.92)	0.08	\$ 800
zrabbit	8	0.00 (0.02)	0.04 (0.51)	0.03	\$ 170
aquarelalinux	7	0.00 (0.01)	0.04 (0.48)	0.03	\$ 148

