

**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 – China  
(Referred to as D25b 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

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*

Maoke Chen, China Education and Research Network

Xing Li, China Education and Research Network

Han Nai-ping, China Standard Software Co., Ltd

Liu Wenqing, China Standard Software Co., Ltd

Zou Wenjing, China Standard Software Co., Ltd

## **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

<b>1</b>	<b>Executive Summary</b>	<b>5</b>
<b>2</b>	<b>Introduction</b>	<b>6</b>
2.1	Details of the considered region . . . . .	6
2.2	Summary of results . . . . .	7
2.3	Problems and constraints found . . . . .	10
2.4	Details of the methodology . . . . .	11
2.5	Contributions . . . . .	11
<b>3</b>	<b>Results</b>	<b>12</b>
3.1	General information . . . . .	12
3.2	Forges information . . . . .	13
3.3	SCM information . . . . .	13
3.3.1	OSDN forge . . . . .	14
3.3.2	HitGforge forge . . . . .	15
3.3.3	OSS forge . . . . .	16
3.3.4	Cosoft forge . . . . .	18
3.3.5	CNForge . . . . .	19
3.4	Source code information . . . . .	19
3.4.1	OSDN . . . . .	19
3.4.2	HitGforge . . . . .	21
3.4.3	OSS . . . . .	23
3.4.4	Cosoft . . . . .	25
3.4.5	CNforge . . . . .	27
3.4.6	Aggregated data . . . . .	29
3.5	Mailing Lists . . . . .	30
3.6	Authorship data . . . . .	30
3.7	SourceForge Analysis . . . . .	32
<b>4</b>	<b>Apendixes</b>	<b>34</b>
4.1	General information . . . . .	34
4.2	Data from repositories . . . . .	36
4.3	COCOMO model applied to software releases . . . . .	43

# List of Tables

2.1	Registered users and projects in Chinese forges (data 16th April 2007). SourceForge has been included for completeness (data June, 2006). . . . .	8
2.2	Information sources that could be extracted from Chinese forges (April-May 2007). . . . .	8
3.1	Data collected by Chinese partners . . . . .	12
3.2	Chinese forges. . . . .	13
3.3	OSDN committers and commits . . . . .	14
3.4	OSDN top five developers with more commits. . . . .	14
3.5	Committers and commits for some selected projects in OSDN forge . . . . .	15
3.6	HitGforge committers and commits . . . . .	15
3.7	HitGforge developers with more commits . . . . .	15
3.8	Committers and commits for some selected projects in HitGForge . . . . .	16
3.9	OSS committers and commits . . . . .	16
3.10	OSS developers with more commits . . . . .	17
3.11	Committers and commits for some projects in OSS forge . . . . .	17
3.12	Cosoft committers and commits . . . . .	18
3.13	Cosoft developers with more commits . . . . .	18
3.14	Committers and commits for some selected projects in the Cosoft forge . . . . .	19
3.15	Local forges comparison . . . . .	19
3.16	Basic COCOMO model applied (using SLOCCount) to the largest OSDN forge software releases . . . . .	19
3.17	Programming languages used in OSDN forge . . . . .	20
3.18	Basic COCOMO model applied (using SLOCCount) to the largest HitGforge software releases . . . . .	21
3.19	Programming languages used in HitGforge . . . . .	22
3.20	Basic COCOMO model applied (using SLOCCount) to the largest OSS forge software releases . . . . .	23
3.21	Programming languages used in OSS forge . . . . .	24
3.22	Basic COCOMO model applied (using SLOCCount) to the largest Cosoft software releases . . . . .	25
3.23	Programming languages used in Cosoft forge . . . . .	26
3.24	Basic COCOMO model applied (using SLOCCount) to the largest CNforge software releases . . . . .	27
3.25	Programming languages used in CNforge forge . . . . .	28
3.26	Programming languages used in Chinese forges . . . . .	29
3.27	General mailing lists summary of results . . . . .	30
3.28	Type of authorship in Chinese forges . . . . .	31
3.29	Type of authorship in OSDN forge . . . . .	31
3.30	Type of authorship in HitGforge forge . . . . .	32
3.31	Type of authorship in OSS forge . . . . .	32
3.32	Type of authorship in Cosoft Forge . . . . .	32

3.33	Type of authorship in CNForge . . . . .	32
3.34	General data from SourceForge . . . . .	33
3.35	The most active Chinese committers detected in SourceForge . . . . .	33
4.1	Communities provided by partners . . . . .	34
4.2	Lugs provided by partners . . . . .	34
4.3	Projects provided by partners . . . . .	34
4.3	Projects provided by partners . . . . .	35
4.4	Platforms provided by partners . . . . .	36
4.5	SCM results in OSDN forge . . . . .	36
4.5	SCM results in OSDN forge . . . . .	37
4.6	SCM results in OSS forge . . . . .	37
4.6	SCM results in OSS forge . . . . .	38
4.7	SCM results in HitGforge forge . . . . .	38
4.7	SCM results in HitGforge forge . . . . .	39
4.8	SCM results in Cosoft forge . . . . .	39
4.8	SCM results in Cosoft forge . . . . .	40
4.8	SCM results in Cosoft forge . . . . .	41
4.9	SCM results in SourceForge forge . . . . .	41
4.9	SCM results in SourceForge forge . . . . .	42
4.9	SCM results in SourceForge forge . . . . .	43
4.10	Basic COCOMO model applied to OSDN software distributions . . . . .	43
4.10	Basic COCOMO model applied to OSDN software distributions . . . . .	44
4.11	Basic COCOMO model applied to OSS software distributions . . . . .	44
4.12	Basic COCOMO model applied to HitGForge software distributions . . . . .	44
4.12	Basic COCOMO model applied to HitGForge software distributions . . . . .	45
4.13	Basic COCOMO model applied to Cosoft software distributions . . . . .	45
4.13	Basic COCOMO model applied to Cosoft software distributions . . . . .	46
4.14	Basic COCOMO model applied to CNForge software distributions . . . . .	46

# List of Figures

2.1	Map of China . . . . .	7
2.2	SCM repositories, mailing lists and software releases found in forges . . . . .	9
2.3	Committers per forge . . . . .	9
2.4	Commits per forge . . . . .	10
3.1	General information obtained from surveys (comparison among countries) . . . . .	13
3.2	Distribution of commits per project in the OSDN forge . . . . .	14
3.3	Distribution of commits per project in HitGforge . . . . .	16
3.4	Relation of commits per project in OSS forge . . . . .	17
3.5	Distribution of commits per project in Cosoft forge . . . . .	18
3.6	Lines of code in OSDN forge . . . . .	20
3.7	Programming languages used in OSDN forge . . . . .	21
3.8	Lines of code in HitGforge forge . . . . .	22
3.9	Programming languages used in HitGforge forge . . . . .	23
3.10	Lines of code in OSDN forge . . . . .	24
3.11	Programming languages used in OSS forge . . . . .	25
3.12	Lines of code in Cosoft forge . . . . .	26
3.13	Programming languages used in Cosoft forge . . . . .	27
3.14	Lines of code in CNForge forge . . . . .	28
3.15	Programming languages used in CNForge forge . . . . .	29
3.16	Mailing lists, posters and messages per forge . . . . .	30
3.17	Authorship data (in software releases). . . . .	31
3.18	Commits in SourceForge . . . . .	33

# 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 China. 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 several sites hosting libre software development (forges).

Besides SourceForge (the largest forge in the world, for which the Chinese participation has been estimated), this study has analysed five Chinese forges, of which Cosoft, OSS and OSDN are the largest ones. Cosoft is the largest forge in China with more than 45,000 registered users and more than 1,450 registered projects, while OSS is the next one by number of registered users (with more than 6,500), and OSDN by number of registered projects (187). This means that there is a big gap between Cosoft and the other forges.

Considering the worldwide development community, SourceForge is also an important component of the infrastructure used by Chinese developers. More than 36,000 registered SourceForge users and 850 registered projects were estimated as Chinese.

Cosoft forge is, by far, the forge with more lines of code archived. But it must be noticed that the largest projects hosted in Cosoft are Linux distributions which have been ignored in this study. Even with this exception, Cosoft continues being the forge with more lines of code, more than 2,000,000. The next forge by lines of code is HitGforge, with about 400,000 lines of code.

Generally speaking, mailing lists are not the main channel of communication used in Chinese forges. In total, less than 150 mailing lists have been found. Only in OSDN and HitGForge forges there were found more mailing lists than software releases or SCM (source code management) repositories.

Regarding SCM repositories, no more than 250 have been found (with just 82 if SourceForge is omitted). SCM repositories are spread in similar quantities among OSDN, HitGForge, Cosoft and OSS forges, despite the enormous differences in the number of registered users and registered projects.

## Chapter 2

# Introduction

Within the context of the FLOSSWorld project, this report is devoted to the quantitative study of libre (free, open source) software development in China . The information presented here is based on public data found in the repositories of libre software projects hosted in the analysed 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 analysed, 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 “Methodology 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.** People’s Republic of China is located in the east of the Asian continent, on the western shore of the Pacific Ocean. The national capital of China is Beijing and it has Standard Chinese or Mandarin as the official language. With Hong Kong SAR (Special Administrative Region), Macao SAR and Taiwan Province excluded, the population of China mainland is estimated in 1,307.56 million (at the end of 2005) and the estimated GDP (by 2005) is RMB 13,687.59 billion Yuan, or about \$1.71 trillion (\$1,740 per capita)<sup>1</sup>

**Languages.** Chinese is the official language, although there are some unofficial languages. Four big families live in China. Sino-Tibetan, with Mandarin, Southern, Tibetan, Kam-Tai and Miao-Yao families); Indo-European family with Tajik; Austro-Asiatic with Mon-Knmer and Altai with Turkic, Mongolian, Manchu-Tungus and Korean.<sup>2</sup>

**Map.** Figure 2.1 shows a map of China<sup>3</sup>.

---

<sup>1</sup>GNI (gross national income) per capita, Atlas method, in US dollars, 2005, according to the World Bank. <http://devdata.worldbank.org/external/CPPProfile.asp?PTYPE=CP&CCODE=CHN>

<sup>2</sup>This information is based on Wikipedia [http://en.wikipedia.org/wiki/Languages\\_of\\_China](http://en.wikipedia.org/wiki/Languages_of_China)

<sup>3</sup>Official map of China as obtained in <http://www.gov.cn/>



Figure 2.1: Map of China

**Time zone.** The Chinese time zone<sup>4</sup> is (UTC +8) and it is shared with more countries such as Western Australia, Central Indonesia, Malaysia, Eastern Russia and other countries. Hence, using the timezone configured in mail clients and other information sources to identify Chinese developers is not an accurate method, although in Eastern Russia and Western Australia there is not high density of population.

**Internet top level domain.** The Chinese TLD (Domain Name System Top Level Domain, which is .cn)<sup>5</sup> is an important indicator to recover information from Chinese email addresses; so if someone has a Chinese TLD, he/she is supposed to be a Chinese user or developer.

[Both the description of the geographic area and the map of Continental China are from the Wikipedia<sup>6</sup>]

## 2.2 Summary of results

The Chinese FLOSSWorld partners have identified five forges that could be considered as Chinese local forges. These forges have been spidered in order to identify the projects they contain. Table 2.1 lists a relation of the number of projects and users registered in each of the forges. The world's most popular forge, SourceForge, has been added to the table as many Chinese developers and Chinese-driven projects have been found there. The number of registered users at SourceForge gives the estimation of Chinese developers identified as such in SourceForge<sup>7</sup>. The 851 projects in SourceForge that are Chinese-driven have a majority (i.e. more than 50%) of Chinese developers in their teams.

<sup>4</sup>Methodology report - Chapter Methodology - Section Global forge's analysis

<sup>5</sup>Methodology report - Chapter Methodology - Section Global forge's analysis

<sup>6</sup>[http://en.wikipedia.org/wiki/People's\\_Republic\\_of\\_China](http://en.wikipedia.org/wiki/People's_Republic_of_China)

<sup>7</sup>Methodology report - Chapter Methodology - Section Global forge's analysis

Forge	Registered Users	Registered Projects
OSDN	3141	187
HitGforge	565	101
Cosoft	46834	1451
CNforge	no data	no data
OSS	6583	112
Sourceforge	36517	851

Table 2.1: Registered users and projects in Chinese forges (data 16th April 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 in several kinds of repositories). This fact is shown in detail in the summary table 2.2, where the number of SCM (CVS/SVN) repositories, committers, commits, mailing lists, software releases and size of the software is given.

Forge	SCM repos	Committers	Commits	MailingLists	Releases	SLOC
OSDN	34/64	33	25,658	74	7/20	131,231
HitGforge	9/54	8	2,059	55	12/14	399,723
Cosoft	70/73	67	29,068	ND	48/59	2,090,299
CNforge	ND	ND	ND	ND	5/8	111,543
OSS	59/59	3	822	ND	8/13	127,478
SourceForge	82	82	24,053	ND	ND	ND

Table 2.2: Information sources that could be extracted from Chinese forges (April-May 2007).

The next section describes the set of difficulties that we have found during the data retrieval process and which explains the *ND* (No-Data) that appears in parts of the table 2.2.

In the results shown in table 2.2 the field *SCM repositories* (Source Code Management repositories) presents the number of non empty repositories and the number of total repositories found (including both empty and non empty repositories<sup>8</sup>). Field *Releases* shows the number of software releases which could be analysed and the number of total software releases found. For SourceForge the results presented are the number of non empty SCM repositories.

Figure 2.2 shows the number of SCM repositories, mailing lists and software releases identified and analysed. Also, figure 2.3 and figure 2.4 show number of detected committers and commits respectively.

<sup>8</sup>Empty repository: There is a SCM account open for this project, however no commit has been made

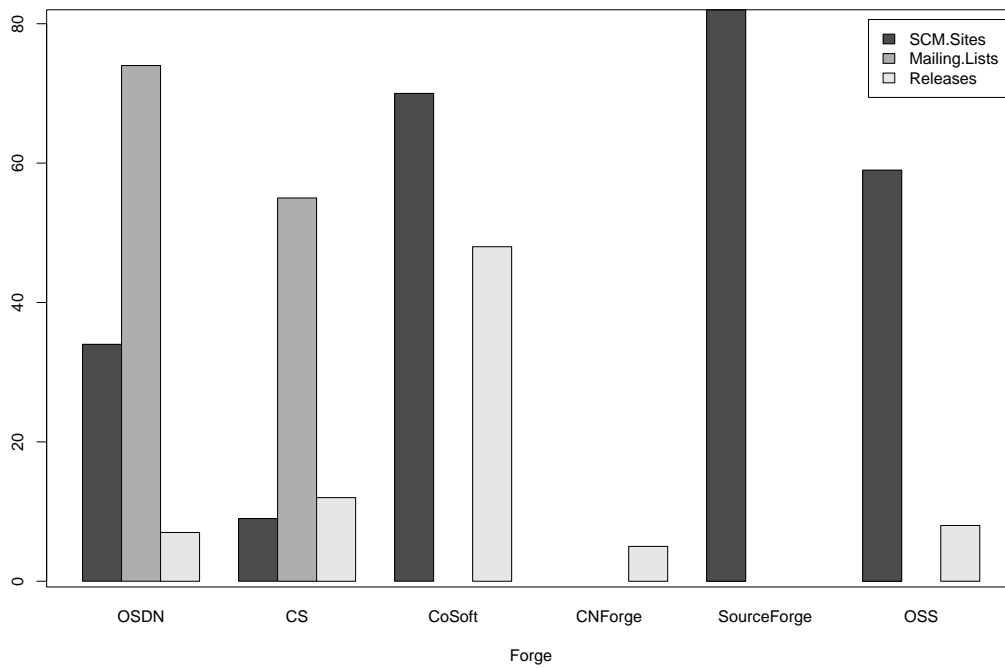


Figure 2.2: SCM repositories, mailing lists and software releases found in forges

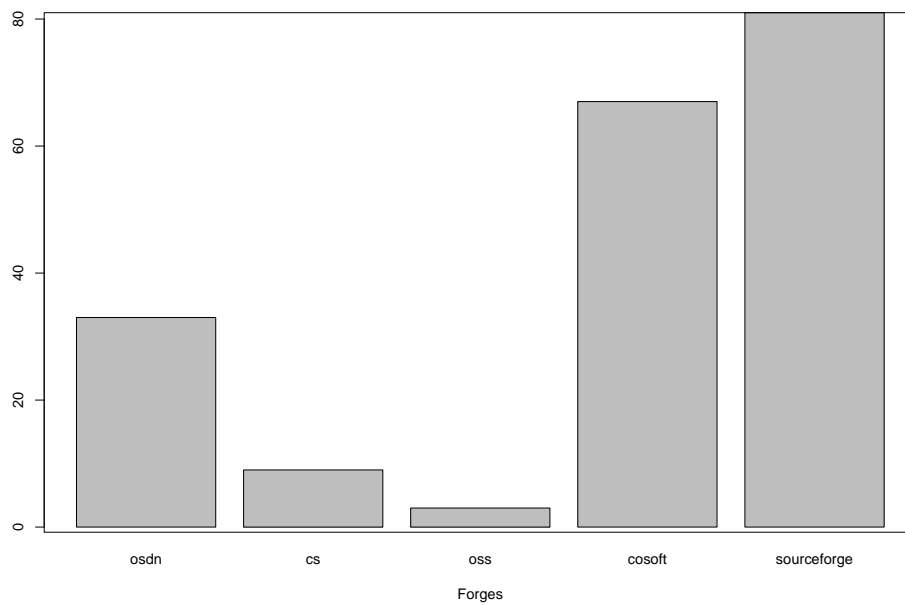


Figure 2.3: Committers per forge

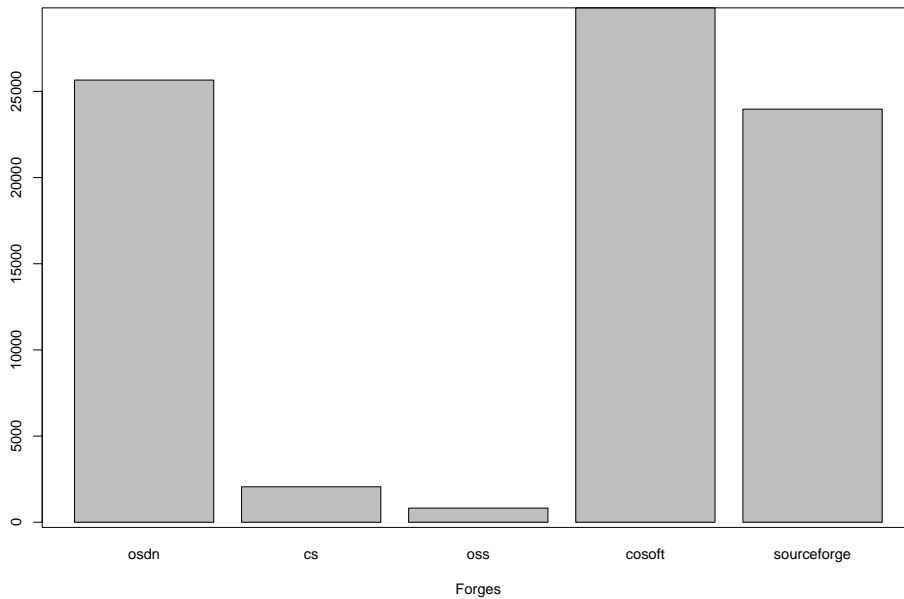


Figure 2.4: Commits per forge

## 2.3 Problems and constraints found

Several problems have been found in the data retrieval process:

1. Links to the source code management (SCM) system: the identification of the SCM systems in use by the projects was not easy as in some forges they were linked in an unclear way. After introducing some extra procedures, the spidering tool was able to find extra information.
2. Language: as most forges were in Chinese, we had problems to use our set of tools to analyze them. Support by local FLOSSWorld partners was of great help in order to fulfill the analysis.
3. The download process for some Chinese projects is slow, so we suffered several delays due to time-outs. The scripts were re-run until a complete data set was obtained.
4. Finally, two forges do not have a Gforge structure, hence they had to be analysed by directly accessing the projects and downloading software releases in a non-automatized way.

In addition to the aforementioned general problems, there have been specific issues on some of the forges which are explained next:

1. OSDN forge: Some repositories could not be downloaded because of the low download speed of the forge. In some cases, the Subversion or CVS tool failed to download the repository. Also, some of the projects listed on the main page do not have a code repository, thus, they could not be analysed. We found the same problems with mailing lists and software releases.
2. HitGforge forge: Several problems appeared, similar to the ones that affected the OSDN forge. Main difficulties could be located in the download process of repositories, and that numerous repositories, mailing lists or software releases did not exist.

3. Cosoft forge: It is not based on GForge, thus it was not possible to automate the process (using a spider tool). Nevertheless, a list of projects was elaborated and, using some scripts, repositories were finally downloaded. Secondly, mailing lists were not found, it is usual to use forums instead of mailing lists in this forge.
4. CNforge: It is not based on GForge, thus it was not possible to download data in an automatic way. For this forge a list of projects was not obtained, thus at the end there is no information related to SCMs. Regarding to mailing lists, it is usual to use forums instead of mailing lists, hence no mailing lists information could be retrieved.
5. OSS forge: Mailing lists were not found in this forge.
6. SourceForge forge: mboxs are hidden so it has not been possible to analyse mailing lists. Neither software releases for each project could be analysed.

## 2.4 Details of the methodology

In this section we provide some details about how the general methodology was applied specifically for this study. The different data retrievals and analysis were made between January 2007 and May 2007 and the following tools were used:

1. CVSanaly<sup>9</sup>: For the Chinese projects, CVSanaly was used to retrieve data from repositories in the local forges and SourceForge. Some problems appeared as some forges use a SSL-based protocols still not supported by CVSanaly. Specifically, for the HitGforge forge all SVN repositories work with this protocol, so they were downloaded manually. Also in Cosoft forge it was necessary to obtain a list of projects in order to download the SCM repositories with some little scripts.
2. MailingListStats<sup>10</sup>: It was run on the mailing lists found in the local forges. Most of the mailing lists were empty, meaning that even if a mailing list exists it could have no messages.
3. Pyternity<sup>11</sup>: It was run on the releases downloaded (not on SCMs).
4. Spider<sup>12</sup>: It was run on all URLs of the local forges. Two forges were not analysed by the Spider tool because this tools is only able to identify information in GForge based forges.

## 2.5 Contributions

This report has been drafted by the GSyC/LibreSoft team of the Universidad Rey Juan Carlos (Madrid, Spain, coordinated by Jesus M. Gonzalez-Barahona). The URJC team performed the data retrieval, mining and analysis, and produced its final version. In addition to the general collaboration by all partners, two Chinese FLOSSWorld partners have contributed specifically (CERNET<sup>13</sup> and CS2S<sup>14</sup>) by providing information about libre software developing sites, feedback about the drafts of the report, much 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.

---

<sup>9</sup>Methodology report - Chapter Tools - Section CVSanaly

<sup>10</sup>Methodology report - Chapter Tools - Section Mailing List Stats

<sup>11</sup>Methodology report - Chapter Tools - Section Pyternity

<sup>12</sup>Methodology report - Chapter Tools - Section Forge Spider

<sup>13</sup>China Education Research Network, [http://www.edu.cn/cernet/\\_1377/index.shtml](http://www.edu.cn/cernet/_1377/index.shtml)

<sup>14</sup>China Standard Software Co. Ltd, <http://www.cs2c.com.cn/>

## Chapter 3

# Results

### 3.1 General information

Chinese partners collaborated with us identifying different URL's where data could be found<sup>1</sup>. For more information access section 4.1 where this data is detached.

There is a summary in table 3.1.

<b>Region</b>	<b>Communities</b>	<b>Developers</b>	<b>Lugs</b>	<b>Media</b>	<b>Projects</b>	<b>Platforms</b>
China	5	0	2	0	43	2

Table 3.1: Data collected by Chinese partners

Where *Communities* means group of users interested in libre software, *LUGs* are Oficial Linux User Groups (with a physical address), *Media* are any kind of journal magazine related to Linux 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.

---

<sup>1</sup>Methodology report - Chapter Data Sources - Section Primary Data Sources

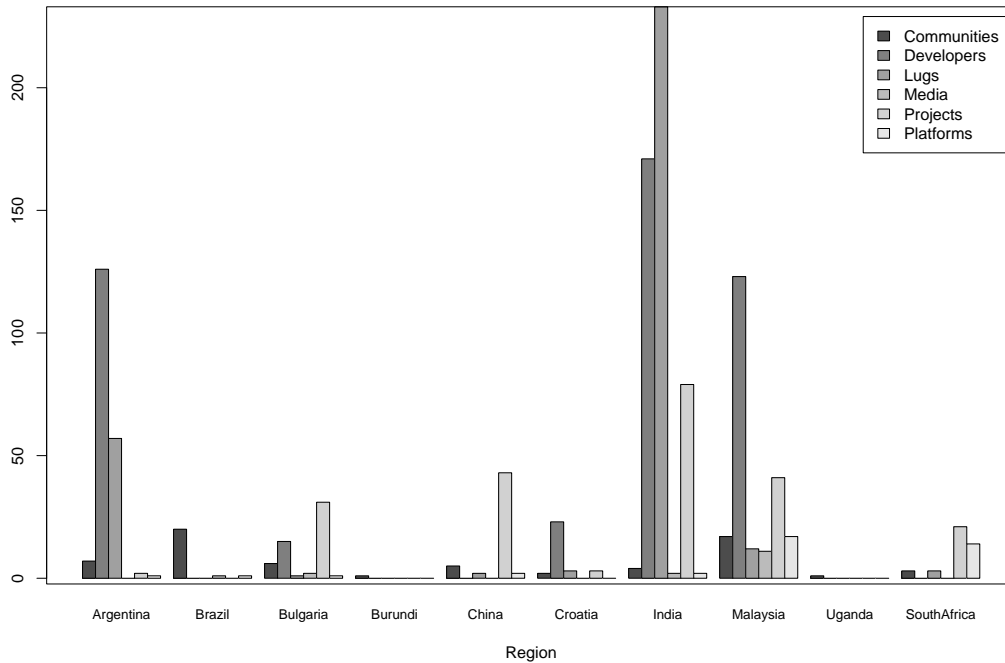


Figure 3.1: General information obtained from surveys (comparison among countries)

### 3.2 Forges information

We have been working with five forges. Table 3.2 gives a small summary of the forges under study.

Forge	Forge name
<a href="http://gforge.osdn.net.cn/">http://gforge.osdn.net.cn/</a>	OSDN
<a href="http://gf.cs.hit.edu.cn/">http://gf.cs.hit.edu.cn/</a>	HitGforge
<a href="http://cosoft.org.cn/">http://cosoft.org.cn/</a>	Cosoft
<a href="http://www.cnforge.org/">http://www.cnforge.org/</a>	CnForge
<a href="http://gforge.oss.org.cn/">http://gforge.oss.org.cn/</a>	OSS

Table 3.2: Chinese forges.

### 3.3 SCM information

Although five forges have been studied, the SCM information for one of them (CNforge) could not be analysed. CNForge could not be analysed because it does not have an homogeneous format to download repositories. On the other hand, forges like OSDN, HitGforge, OSS and Cosoft do not need and account and they have repositories, although some difficulties appeared. HitGforge uses the SSL protocol to connect to repositories and CVSanaly does not support this yet. The SVN repositories had to be downloaded manually in order to be analyzed by CVSanaly. The next paragraphs explain the results for each forge. Finally, in Cosoft forge it was necessary to obtain a list of projects (accessing the root directory of CVS) and trying to generate all the URLs to download the projects.

### 3.3.1 OSDN forge

64 projects have been found in this forge, but only 34 of them have a repository. There are 34 active developers who have ever made a commit and in total we can account up to 25658 commits (see table 3.3).

Forge	No. committers	No. commits
<a href="http://gforge.osdn.net.cn/">http://gforge.osdn.net.cn/</a>	34	25,658

Table 3.3: OSDN committers and commits

Table 3.4 shows the five most active developers in the OSDN forge.

Project name	Committer	# Commits
xreal3d	nhsoft	4,586
cjavatutorial	leasun	3,325
jetnuke	ginger547	3,160
dduoftspider	y97523szb	1,999
abc	leejingbo	1,592

Table 3.4: OSDN top five developers with more commits.

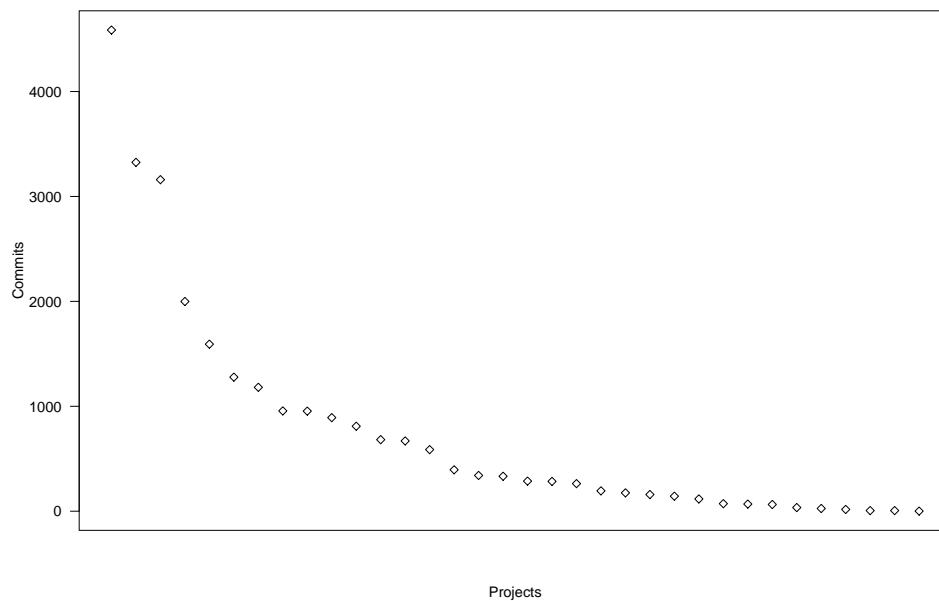


Figure 3.2: Distribution of commits per project in the OSDN forge

Project name	No. of committers	No. of commits
xreal3d	1	4586
cjavatutorials	1	3325
jetnuke	1	3160
dduoftspider	1	1999
abc	1	1592
clking	1	1278
iboke	1	1181

Table 3.5: Committers and commits for some selected projects in OSDN forge

### 3.3.2 HitGforge forge

64 projects have been found in this forge, but only 28 of them have a repository. There are 9 active developers who have ever made a commit and in total there are 2077 commits. 16 root “users” have been found; this is because there is an automated script which sets up the repository for every project and that is run by root.

Forge	No. committers	No. commits
<a href="http://gf.cs.hit.edu.cn/">http://gf.cs.hit.edu.cn/</a>	9	2077

Table 3.6: HitGforge committers and commits

Table 3.7 shows the number of committers and commits for this forge.

Project name	Committer	No. commits
lilacwebex	robaggio	888
xnews	btliu	476
csamples	sun	363
lightning	cowoo	106
freeip	hexing	17

Table 3.7: HitGforge developers with more commits

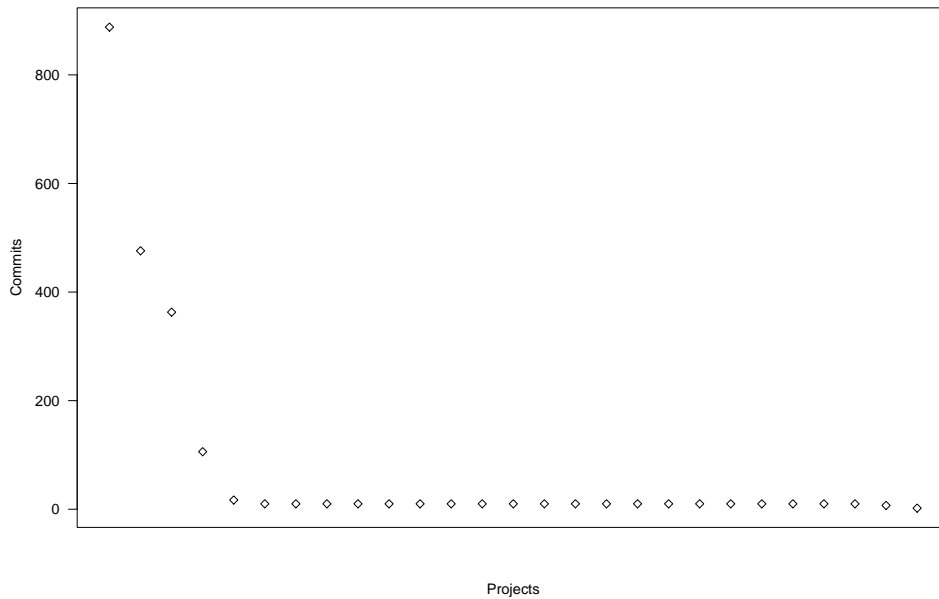


Figure 3.3: Distribution of commits per project in HitGforge

Project name	No. of committers	No. of commits
lilacwebex	1	888
xnews	1	476
csamples	1	363
lightning	1	106

Table 3.8: Committers and commits for some selected projects in HitGForge

### 3.3.3 OSS forge

58 projects have been found in this forge and all of them have an SCM repository. However, most of them did not have important information which could be analyzed by CVSanaly: 53 of these repositories had only the user “liy1”, with 10 commits. There is a total of 822 commits in the whole forge.

Forge	No. committers	No. commits
<a href="http://gforge.oss.org.cn/">http://gforge.oss.org.cn/</a>	59	822

Table 3.9: OSS committers and commits

Table 3.10 shows committers and commits for some projects.

Project name	Committer	No. commits
dget	balancesli	197
p2pdeliver	balancesli	79
Rest of the projects	liyl	10
elfparser	balancesli	4
sharelib	jacob	2

Table 3.10: OSS developers with more commits

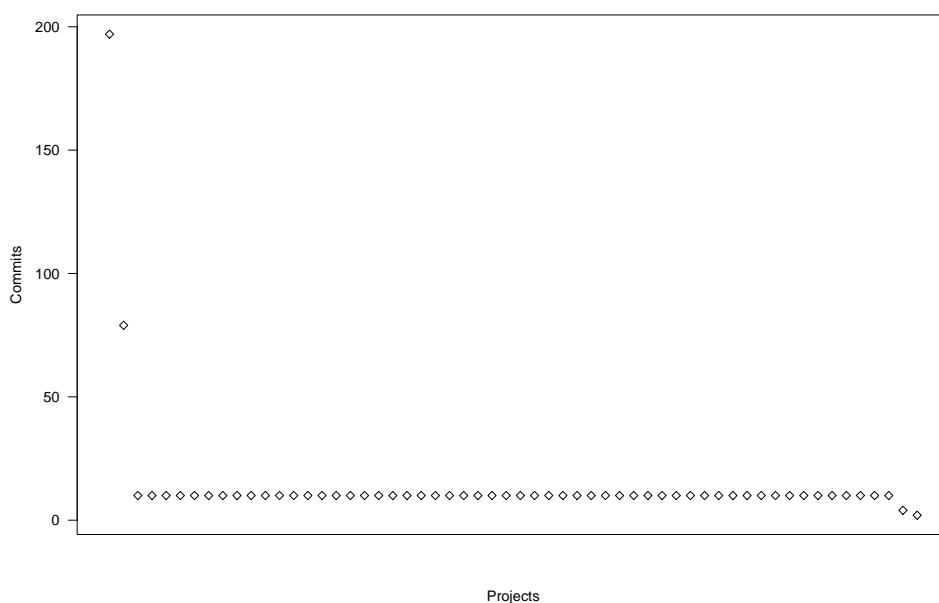


Figure 3.4: Relation of commits per project in OSS forge

Project name	No. of committers	No. of commits
dget	1	197
p2pdeliver	1	79
ioogamestar	1	10
seekstorage	1	10
codebeautifiers	1	10
rpl_input	1	10
radish	1	10
liktron	1	10
thuthesis	1	10
frenotify	1	10

Table 3.11: Committers and commits for some projects in OSS forge

### 3.3.4 Cosoft forge

73 projects have been found in this forge and all of them had a repository. However, three of them did not have important information which could be analysed by CVSanaly. In total, there are 29,068 commits (29,849 when adding root commits).

Forge	No. committers	No. commits
<a href="http://www.cosoft.org.cn/">http://www.cosoft.org.cn/</a>	67	29,068

Table 3.12: Cosoft committers and commits

Table 3.13 shows committers and commits for some selected projects.

Project name	Committer	No. commits
psis	iamfoolberg	3227
jvc	rufujian2005	3166
javahope	etxiaohei	3105
xtelsim	xtel	2960
myblog	somebody	1721

Table 3.13: Cosoft developers with more commits

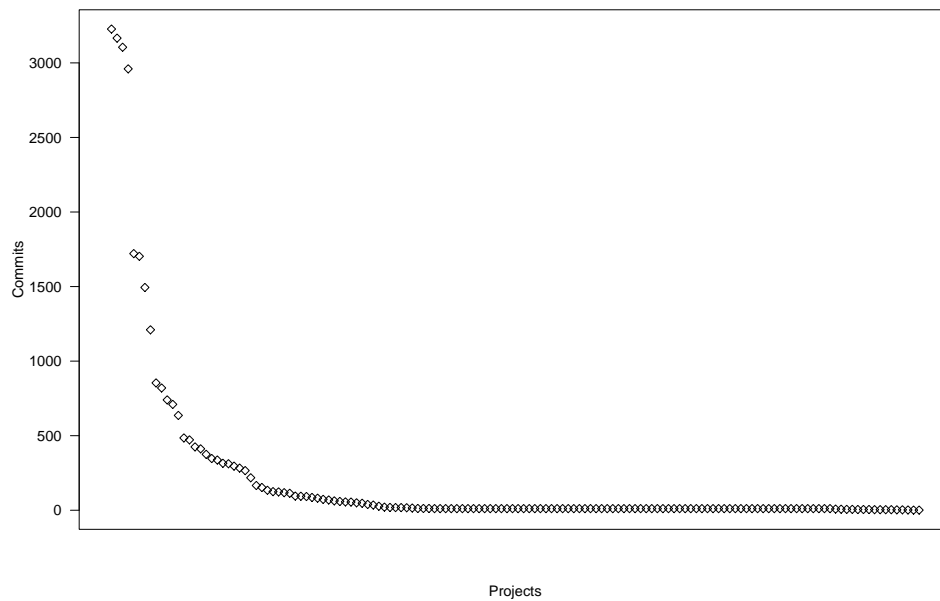


Figure 3.5: Distribution of commits per project in Cosoft forge

Project name	No. of committers	No. of commits
psis	1	3227
jvc	2	3178
javahope	1	3105
xtelsim	1	2960
granite	3	1919
myblog	1	1721
revolutionbiz	1	1494
fitframework	1	1210

Table 3.14: Committers and commits for some selected projects in the Cosoft forge

### 3.3.5 CNForge

As stated before, it was not possible to carry out a proper SCM analysis for this forge. Each project hosted in CNforge had its own repository, but there was no automatic method for downloading the information.

## 3.4 Source code information

The data for the analysis of the source code was obtained by downloading releases of software and then running SLOCCount over them. This whole process was automatic and carried out by our spider software. The cases of Cosoft and CNforge were special. These two forges could not be analysed by the spider tool. In these cases, software releases had to be downloaded by hand.

The difference between the “number of analyzed projects” and the “number of projects with data” is due to non-analyzable files (mainly binary files, such as dynamic libraries -.dll- or executable -.exe- files). Table 3.15 shows projects and analyzed projects for each forge.

Forge	No. detected releases	No. analysed releases
<a href="http://gforge.osdn.net.cn/">http://gforge.osdn.net.cn/</a>	20	7
<a href="http://gf.cs.hit.edu.cn/">http://gf.cs.hit.edu.cn/</a>	14	12
<a href="http://gforge.oss.org.cn/">http://gforge.oss.org.cn/</a>	13	8
<a href="http://cosoft.org.cn/">http://cosoft.org.cn/</a>	59	50
<a href="http://www.cnforge.org/">http://www.cnforge.org/</a>	8	8

Table 3.15: Local forges comparison

### 3.4.1 OSDN

Table 3.16 and figure 3.6 show results from applying the basic COCOMO model for cost estimation, to releases obtained from OSDN.

Project name	No. of SLOC	Effort	Schedule	No. developers	Estimated cost
xreal3d	75,705	18.80 (225.58)	1.63 (19.60)	11.51	\$ 2,539,349
ylib	38,820	9.32 (111.87)	1.25 (15.01)	7.45	\$ 1,259,360
afw	13,243	3.01 (36.17)	0.81 (9.77)	3.70	\$ 407,125
rttim	1,219	0.25 (2.95)	0.31 (3.77)	0.78	\$ 33,262
msnchatlog	936	0.19 (2.24)	0.28 (3.40)	0.66	\$ 25,205

Table 3.16: Basic COCOMO model applied (using SLOCCount) to the largest OSDN forge software releases

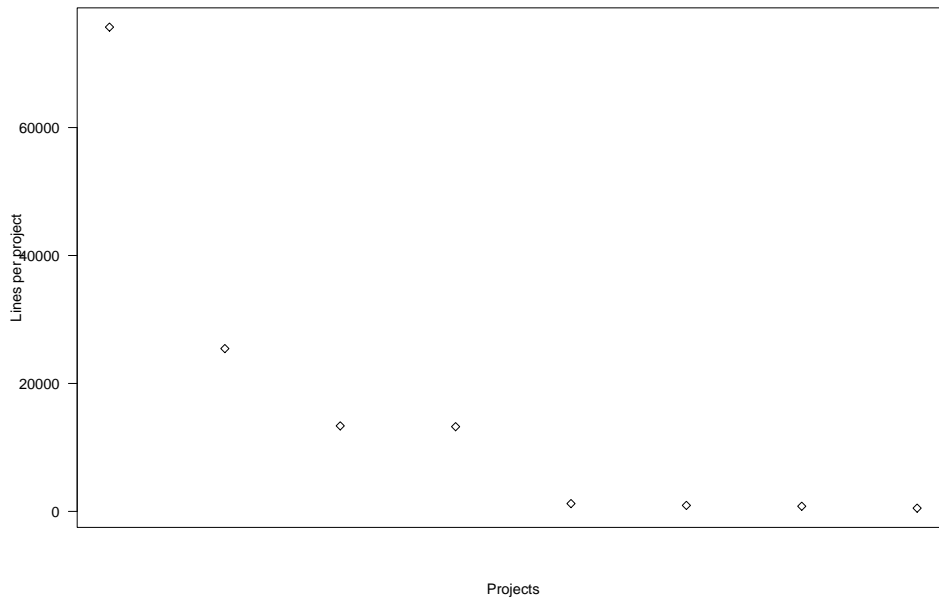


Figure 3.6: Lines of code in OSDN forge

Also, table 3.17 and figure 3.7 contain information regarding the percentage of use of programming languages OSDN.

Programming language	No. of lines	Percentage
C ++	103816	79.1093
C	14172	10.7992
C #	13243	10.0913
Total	131231	100.0

Table 3.17: Programming languages used in OSDN forge

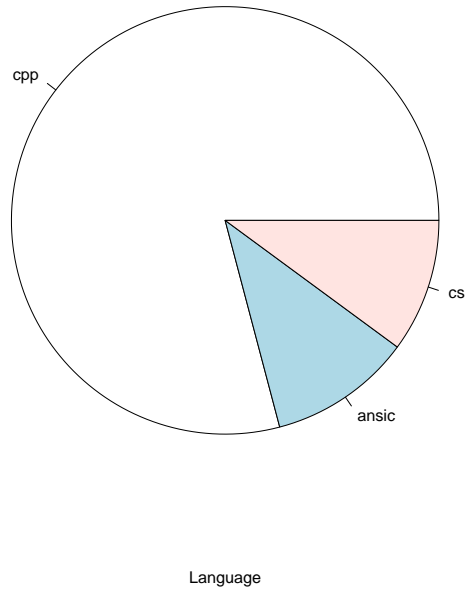


Figure 3.7: Programming languages used in OSDN forge

### 3.4.2 HitGforge

Table 3.18 and figure 3.8 show results from applying the basic COCOMO model for cost estimation, to releases obtained from HitGforge.

Project name	No. of SLOC	Effort	Schedule	No. developers	Estimated cost
cn-moodle	306,114	81.51 (978.11)	2.85 (34.22)	28.58	\$ 11,010,803
ck-erp	84,656	21.14 (253.66)	1.71 (20.49)	12.38	\$ 2,855,501
irvp	2,540	0.53 (6.39)	0.42 (5.06)	1.26	\$ 71,898
xnews	1,330	0.27 (3.24)	0.33 (3.91)	0.83	\$ 36,449
mystar	1,274	0.26 (3.09)	0.32 (3.84)	0.81	\$ 34,839

Table 3.18: Basic COCOMO model applied (using SLOCCount) to the largest HitGforge software releases

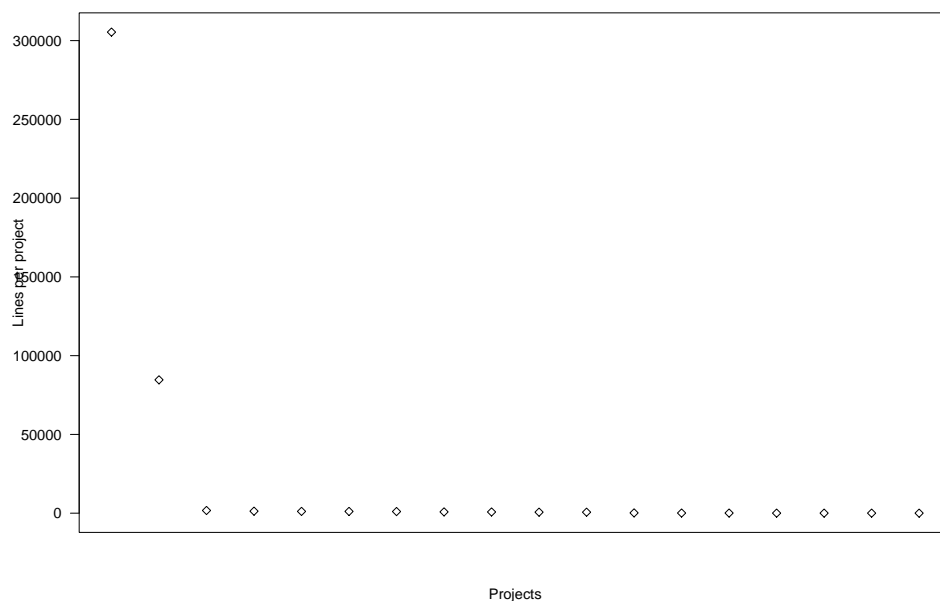


Figure 3.8: Lines of code in HitGforge forge

Also, table 3.19 and figure 3.9 contain information regarding the percentage of use of programming languages in the HitForge.

Programming language	No. of lines	Percentage
PHP	390145	97.6038
C ++	3601	0.9008
C	2959	0.7402
Java	1176	0.2942
Shell	960	0.2401
Perl	638	0.1596
JSP	191	0.0477
Pascal	36	0.0090
Awk	17	0.0042
Total	399723	100.0

Table 3.19: Programming languages used in HitGforge

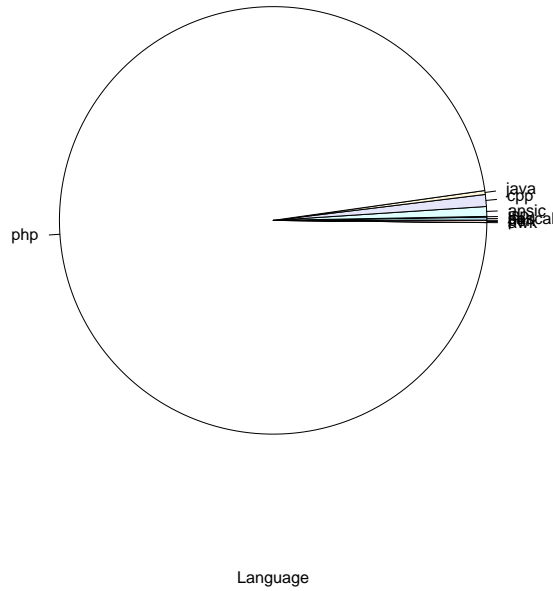


Figure 3.9: Programming languages used in HitGforge forge

### 3.4.3 OSS

Table 3.20 and figure 3.10 show results from applying the basic COCOMO model for cost estimation to releases obtained from OSS forge.

Project name	No. of SLOC	Effort	Schedule	No. developers	Estimated cost
ck-erp	84,656	21.14 (253.66)	1.71 (20.49)	12.38	\$ 2,855,501
dget	28,977	6.86 (82.29)	1.11 (13.36)	6.16	\$ 926,398
yam	7,243	1.60 (19.19)	0.64 (7.68)	2.50	\$ 216,051
imageviewer	2,629	0.55 (6.62)	0.43 (5.13)	1.29	\$ 74,546
elfparser	2,329	0.49 (5.83)	0.41 (4.89)	1.19	\$ 65,640

Table 3.20: Basic COCOMO model applied (using SLOCCount) to the largest OSS forge software releases

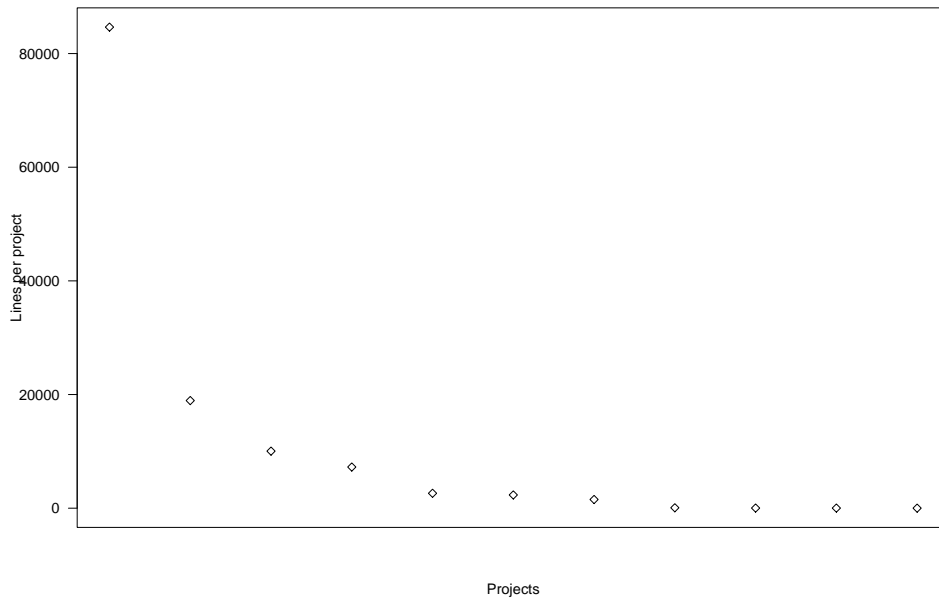


Figure 3.10: Lines of code in OSDN forge

Also, table 3.21 and figure 3.11 contain information regarding the percentage of use of programming languages in the OSS forge.

Programming language	No. of lines	Percentage
PHP	84656	66.4083
C ++	28509	22.3638
C	10040	7.8758
Java	4152	3.2570
Shell	121	0.0949
Total	127478	100.0

Table 3.21: Programming languages used in OSS forge

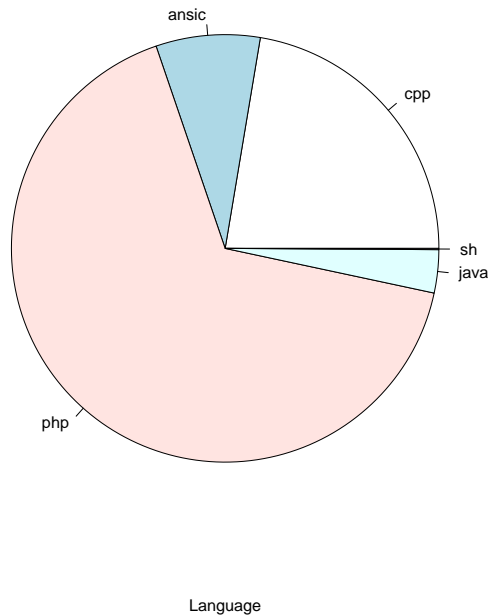


Figure 3.11: Programming languages used in OSS forge

### 3.4.4 Cosoft

In Cosoft forge three linux distribution have been detected. Since these have been not generated in this Chinese forge they have not been included in the tables. However, these distributions have around four million of SLOC each one. Precisely, these distributions are cc-linux, ccBSD and RedOffice whose packages are obtained from ordinary packages from distributions more popular in the linux world, such as .deb or .rpm based distributions.

Other projects have the same problem, however they do not have so many lines. Table 3.22 and figure 3.12 show the results from applying the Basic COCOMO model to the software releases obtained from Cosoft Forge

Project name	No. of SLOC	Effort	Schedule	No. developers	Estimated cost
todde-works	664,855	184.03 (2,208.38)	3.89 (46.63)	47.36	\$ 24,860,217
chinesepython	423,972	114.75 (1,376.94)	3.25 (38.97)	35.33	\$ 15,500,498
cnwizards	168,458	43.54 (522.43)	2.25 (26.96)	19.37	\$ 5,881,090
isee	147,601	37.89 (454.73)	2.13 (25.58)	17.78	\$ 5,119,003
dusk	129,275	32.97 (395.64)	2.02 (24.26)	16.31	\$ 4,453,812

Table 3.22: Basic COCOMO model applied (using SLOCCount) to the largest Cosoft software releases

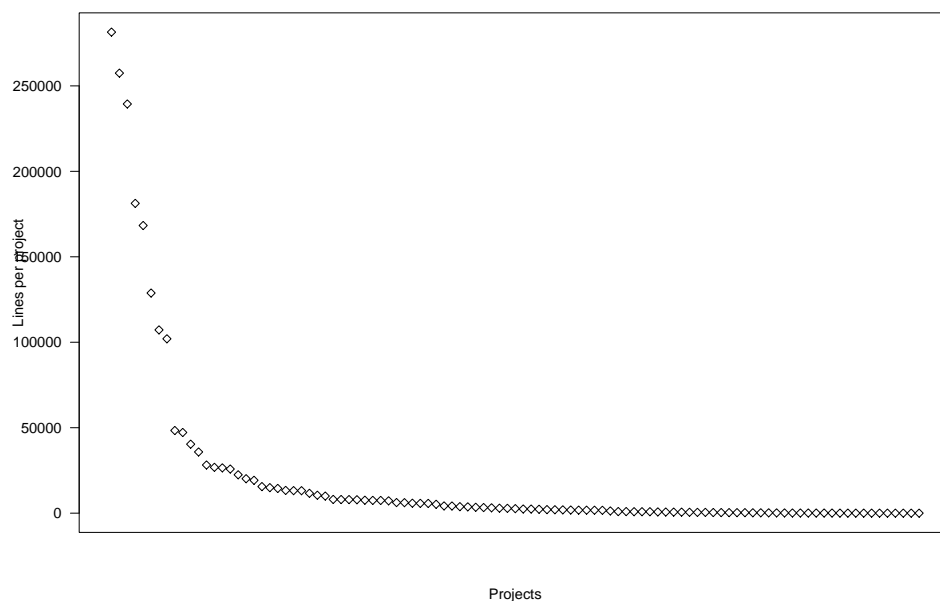


Figure 3.12: Lines of code in Cosoft forge

Also, the table 3.23 and figure 3.4.4 contain information regarding the percentage of use of programming languages in Cosoft (once again, dismissing results from the aforementioned Linux distributions).

Programming language	No. of lines	Percentage
C	718286	34.3628
C ++	326152	15.6031
PHP	304171	14.5515
Perl	261675	12.5185
Python	202526	9.6888
Pascal	170160	8.1404
Shell	52231	2.4987
Java	38942	1.8629
Asm	8475	0.4054
JSP	4689	0.2243
Lisp	2474	0.1183
Yacc	316	0.0151
Sed	202	0.0096
Total	2090299	100.0

Table 3.23: Programming languages used in Cosoft forge

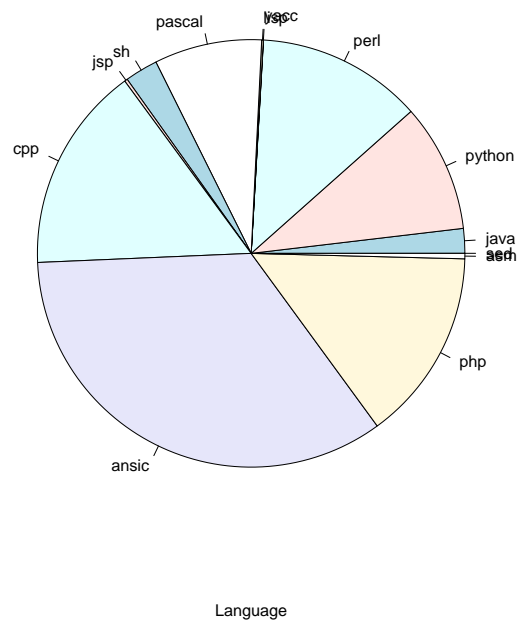


Figure 3.13: Programming languages used in Cosoft forge

### 3.4.5 CNforge

Table 3.24 and figure 3.4.5 show the results from applying the Basic COCOMO model to software releases obtained from CNForge.

Project name	No. of SLOC	Effort	Schedule	No. developers	Estimated cost
ck-erp	84,656	21.14 (253.66)	1.71 (20.49)	12.38	\$ 2,855,501
geosings	15,329	3.51 (42.17)	0.86 (10.36)	4.07	\$ 474,713
spadger	6,827	1.50 (18.04)	0.63 (7.50)	2.40	\$ 203,041
mybomber	3,897	0.83 (10.01)	0.50 (6.00)	1.67	\$ 112,696
pmvc	834	0.17 (1.98)	0.27 (3.24)	0.61	\$ 22,329

Table 3.24: Basic COCOMO model applied (using SLOCCount) to the largest CNforge software releases

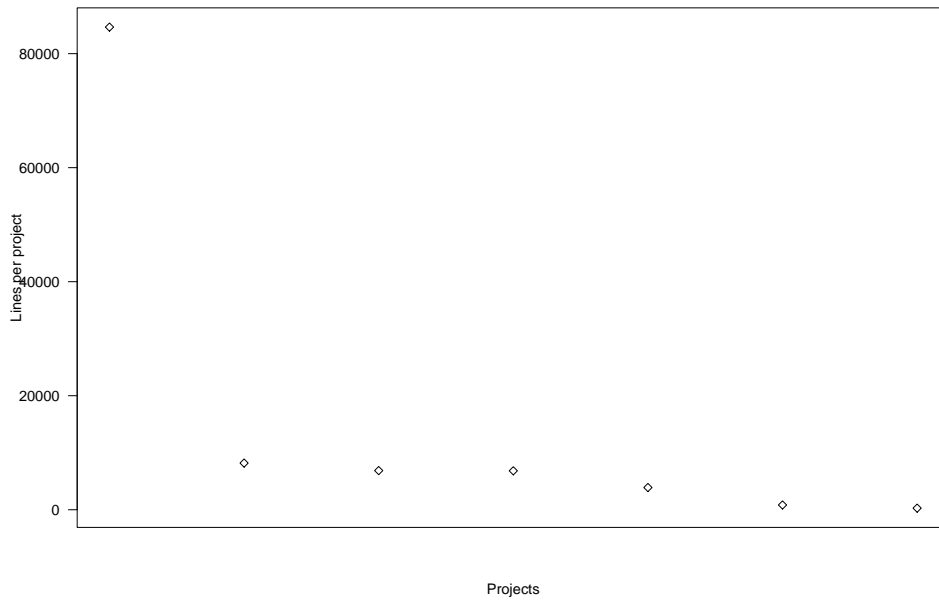


Figure 3.14: Lines of code in CNForge forge

Also, the table 3.25 and figure 3.4.5 contain information regarding the percentage of use of programming languages in CNForge.

Programming language	No. of lines	Percentage
PHP	85490	76.6430
C ++	11000	9.8616
Shell	8185	7.3379
Python	6868	6.1572
Total	111543	100.0

Table 3.25: Programming languages used in CNforge forge

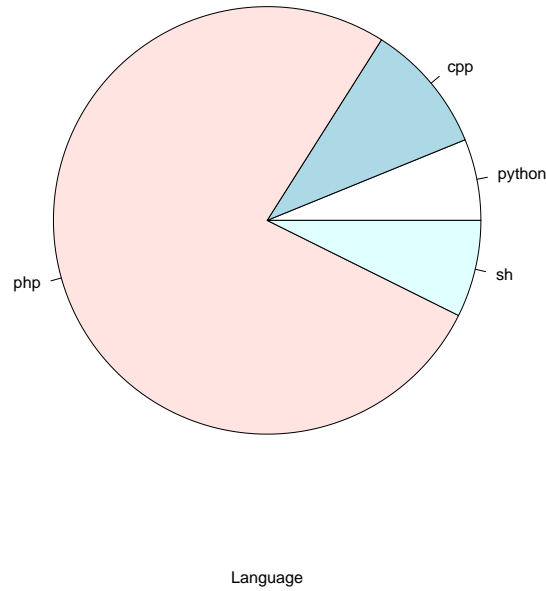


Figure 3.15: Programming languages used in CNForge forge

### 3.4.6 Aggregated data

Table 3.26 shows aggregated data obtained in all the forges in the study: Number of lines of code per programming language and percentage of use for each of them.

Programming language	No. of lines	Percentage
PHP	867590	30.2992
C	745457	26.0339
C ++	473078	16.5215
Perl	262313	9.1608
Python	209394	7.3127
Pascal	170196	5.9438
Shell	61497	2.1476
Java	44270	1.5460
C #	13243	0.4624
Asm	8475	0.2959
JSP	4880	0.1704
Lisp	2474	0.08640
Yacc	316	0.0110
Sed	202	0.0070
Awk	17	0.0005
Total	2863402	100.0

Table 3.26: Programming languages used in Chinese forges

### 3.5 Mailing Lists

There is a total of 129 mailing lists among the forges in this study. But and all of them have been found and analysed in two Chinese forges: <http://gforge.osdn.net.cn/> and <http://lists.gf.cs.hit.edu.cn/>.

Many of these mailing lists are empty mailing lists. That is mailing lists that have not been created ever, but have a link in the project's page. There are also a good number of commit mailing lists, which are those that contain automatic messages from the SCM system, usually generated when a developer commits changes to the repository.

Forge	No. mailing lists	No. posters	No. messages
<a href="http://gforge.osdn.net.cn/">http://gforge.osdn.net.cn/</a>	74	11	23
<a href="http://lists.gf.cs.hit.edu.cn/">http://lists.gf.cs.hit.edu.cn/</a>	55	5	26

Table 3.27: General mailing lists summary of results

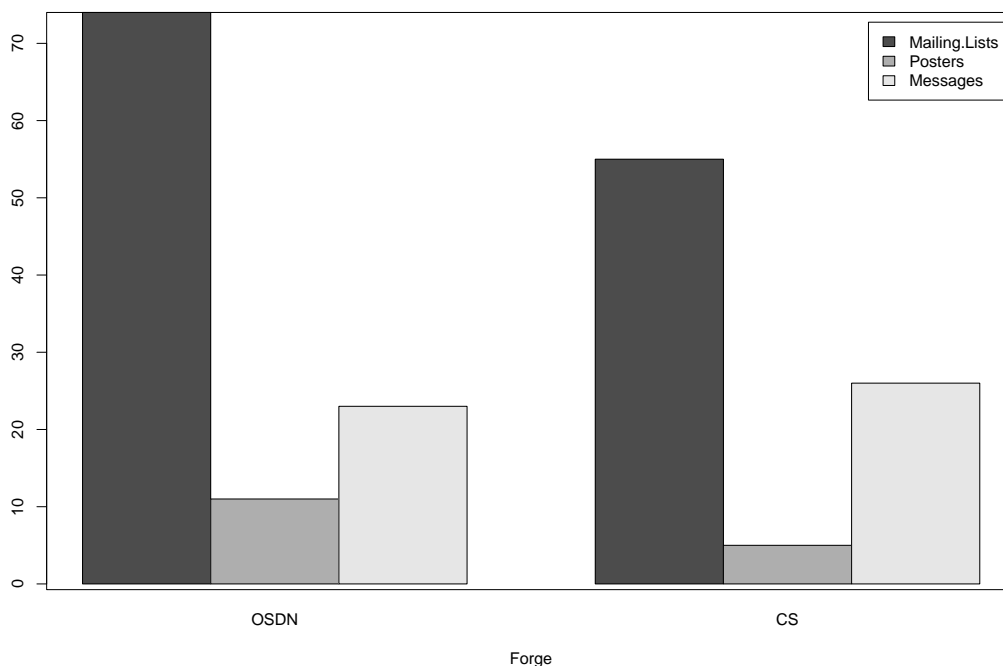


Figure 3.16: Mailing lists, posters and messages per forge

### 3.6 Authorship data

There are different types of authors in each project. Some of them, work for companies that are interested in participating in the project. The results in this section have been obtained from software releases (not from SCM repositories), by analysing copyright attributions in source files. Different authors identified have been classified by categories. Table 3.28 and figure 3.17 show these results.

Type of author	Detected number
Individual authors	267
Teams and Groups	22
Universities	17
Enterprises	90
Foundation or public entities	5
Anonymous/Unknown	11

Table 3.28: Type of authorship in Chinese forges

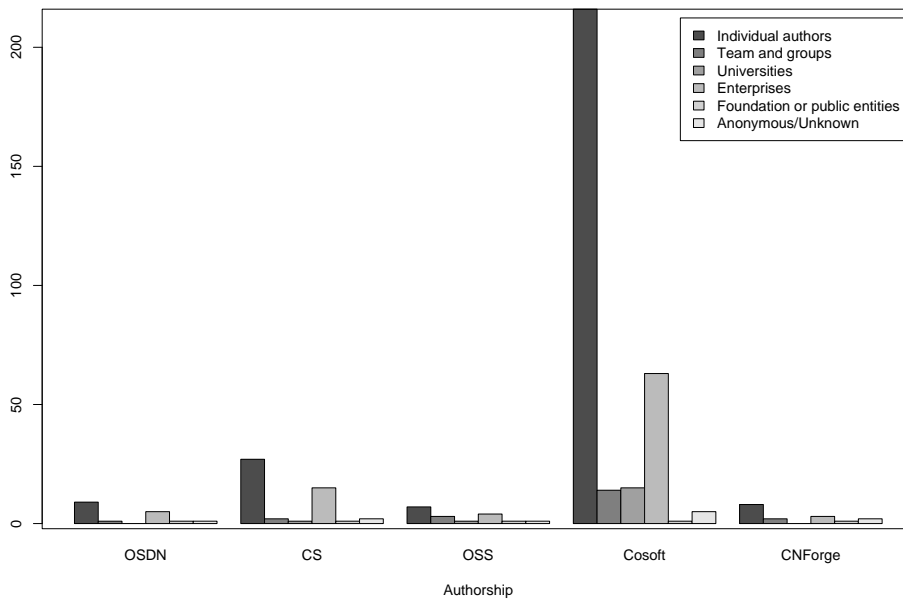


Figure 3.17: Authorship data (in software releases).

Type of author	Detected number
Individual authors	9
Teams and Groups	1
Universities	0
Enterprises	5
Foundation or public entities	1
Anonymous/Unknown	1

Table 3.29: Type of authorship in OSDN forge

<b>Type of author</b>	<b>Detected number</b>
Individual authors	27
Teams and Groups	2
Universities	1
Enterprises	15
Foundation or public entities	1
Anonymous/Unknown	2

Table 3.30: Type of authorship in HitGforge forge

<b>Type of author</b>	<b>Detected number</b>
Individual authors	7
Teams and Groups	3
Universities	1
Enterprises	4
Foundation or public entities	1
Anonymous/Unknown	1

Table 3.31: Type of authorship in OSS forge

<b>Type of author</b>	<b>Detected number</b>
Individual authors	216
Teams and Groups	14
Universities	15
Enterprises	63
Foundation or public entities	1
Anonymous/Unknown	5

Table 3.32: Type of authorship in Cosoft Forge

<b>Type of author</b>	<b>Detected number</b>
Individual authors	8
Teams and Groups	2
Universities	0
Enterprises	3
Foundation or public entities	1
Anonymous/Unknown	2

Table 3.33: Type of authorship in CNForge

### 3.7 SourceForge Analysis

Using the methodology outlined in the introduction of this report, several SourceForge developers have been identified as Chinese. This section includes data from the analysis of the activity of these developers in SourceForge.

Country	No. of committers	No. of commits
China	82	24053

Table 3.34: General data from SourceForge

Next table shows committers with the highest number of commits:

Project name	Committer	No. of commits
xmlucene	bitiboy	4085
dlog4j	liusoft	3437
garnet3d	chenlee	1080
zhdotemacs	unidevel	1066
kylin	kylinwym	1062

Table 3.35: The most active Chinese committers detected in SourceForge

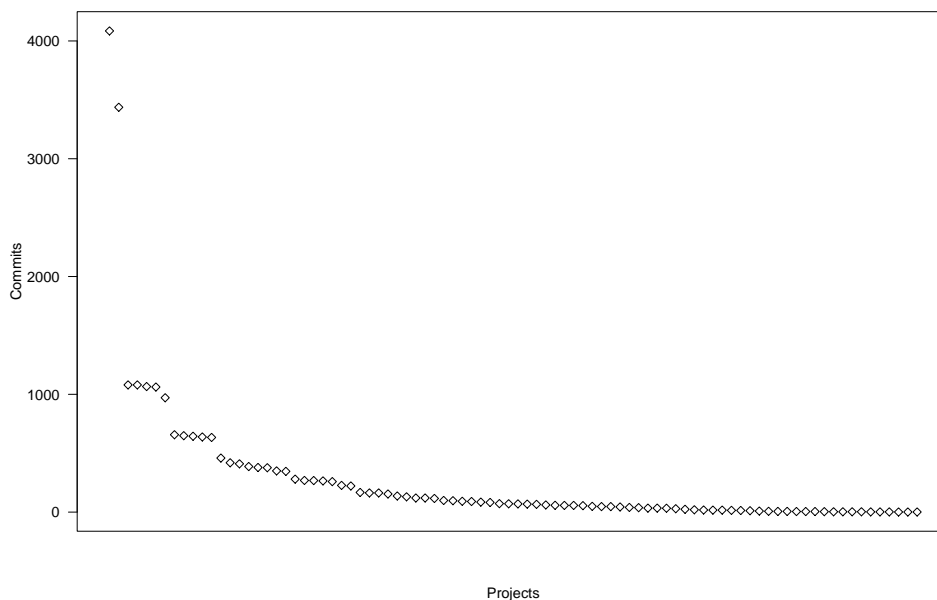


Figure 3.18: Commits in SourceForge

## Chapter 4

# Apendixes

### 4.1 General information

The following is a list of tables showing a relation of information provided by our partners.

#### Communities

Community name	Size	Web site	Comments
China OSS	huge > 1000	<a href="http://oss.org.cn/">http://oss.org.cn/</a>	Under the China OSS promotion union
LUPA	huge > 1000	<a href="http://www.lupaworld.com/">http://www.lupaworld.com/</a>	
Co-create Software League	huge > 1000	<a href="http://cosoft.org.cn/html/">http://cosoft.org.cn/html/</a>	
Huihoo	huge > 1000	<a href="http://www.huihoo.com/">http://www.huihoo.com/</a>	
Aka	very large < 1000	<a href="http://www.aka.org/index.shtml">http://www.aka.org/index.shtml</a>	

Table 4.1: Communities provided by partners

#### Lugs

Lug name	Address	Web site	Type
Beijing GNU/Linux User Group	Beijing	<a href="http://www.beijinglug.org/en/index.php">http://www.beijinglug.org/en/index.php</a>	Local
SHLUG (GeekBone)	Shanghai	<a href="http://www.shlug.org/">http://www.shlug.org/</a>	Local

Table 4.2: Lugs provided by partners

#### Projects

Table 4.3: Projects provided by partners

Project name	Project created	Size
Thesis Template for Tsinghua University	2005	1.54M
Code Beautifier Collection	2005	6.15M
dget - download get for Linux	2006	893K

Continued on next page

Table 4.3: Projects provided by partners

Project name	Project created	Size
BugFreeJ	2006	11.73M
Elf Parser	2006	129K
CK-ERP	2006	1.51M
Image Viewer	2006	587K
Yet Another MATLAB	2006	3.17M
Minsh - a Mini Shell	2006	26K
Chinese Chess under GPL	2001	85,236
Web PM - Online Project Manager	2001	4.1M
ISeeExplorer	2001	2,654,615
PHOENIX Dynamic DNS	2002	15,831
gafSolution GameDev Solution	2002	158,072
General ERP	2001	779,949
andrew software	2001	307,830
recite word easily	2002	1,480,338
CXXUint	2002	58,870
kpop3	2001	1,584,784
Script C	2002	3,069,826
pmail	2001	198,991
Time Sync	2002	9,853
Openext	2002	806,287
DuSK 3D	2001	2,749,113
Fava	2001	151,368
PHPshop Chinese Patch	2001	1,225
ch2000-redoffice-linux	2001	523,025,655
Rexip	2001	7,010,474
yq personal information manage	2002	443,735
Candy Player	2001	333,292
SoftOS	2001	40.5M
King of Translator	2001	13,205,296
Garnet 3D Game Library	2001	1,770,831
Soulogic - Guestbook	2002	21,176
Embedded Database in Linux Minidr	2001	11,369,718
NewEdit	2001	133,171
CNOS	2002	1,051,314
jclientbyyanqh	2002	232,955
easyapache (for win)	2002	24,690
Cosoft Development Platform	2001	2,365,010
MiniGUI	1999	4,594,590
Diablo2 Realm Emulator	2000	365,753

## Platforms/Forges

Platform name	Web site	Registered users	Registered projects
China Gforge	<a href="http://gforge.oss.org.cn/">http://gforge.oss.org.cn/</a>	5986	104
Cosoft	<a href="http://cosoft.org.cn/index.php">http://cosoft.org.cn/index.php</a>	46432	1437

Table 4.4: Platforms provided by partners

## 4.2 Data from repositories

The following is a list of tables showing data from the repository mining with CVSAly, for each forge analysed in the study.

### OSDN forge

Table 4.5: SCM results in OSDN forge

Project name	Committer	No. of commits
xreal3d	nhsoft	4586
cjavatutorials	leasun	3325
jetnuke	ginger547	3160
dduoftpspider	y97523szb	1999
abc	leejingbo	1592
clking	ninebullscom	1278
iboke	hnkelely	1181
jsdatatable	yidinghe	956
sylvan	lonelyp	954
shutters	duyanning	892
angelpm	anticlockwise	810
ndoc	percyboy	682
sarum	xinyu_dev	670
filebox	interhui	587
tradeage	yahle	395
enuke	duxu2004	341
hybridwin32	richardsundusky	333
comframe	netlinux	287
shootsearch	shootsoft	284
openwsn	openwsn	263
hanos	ablo_zhou	194
pbhelper	truewaylee	175
lanim	bineon	159
hakunamatata	interhui	143
anc	hmichaelchen	117
freepyce	backstrokefish	72
magicos	viewstar	68
cnmatrix	outlaw666	64
dragon	cpsoft	35
rorext	cpunion	26
hick	hickwu	17

Continued on next page

Table 4.5: SCM results in OSDN forge

Project name	Committer	No. of commits
minilist	tangtao2000	6
dipteran	wiwolf	6
csqq	topiemie	1

## OSS forge

Table 4.6: SCM results in OSS forge

Project name	Committer	N. of commits
dget	balancesli	197
p2pdeliver	balancesli	79
p2pgridroot	liyl	10
yalog	liyl	10
txge	liyl	10
rplinux64	liyl	10
rmath	liyl	10
frenotify	liyl	10
gorpweblog	liyl	10
debiandoc	liyl	10
unzip_rplinux	liyl	10
rplinuxserver	liyl	10
rpl_livecd	liyl	10
ipv6	liyl	10
rorext	liyl	10
rpl_gcc4	liyl	10
kernel_upgrade	liyl	10
kerner_security	liyl	10
rpl_mount	liyl	10
rpl_unionfs	liyl	10
rpl_packaging	liyl	10
rpl_installer	liyl	10
clg	liyl	10
rpl_depcheck	liyl	10
rpl_input	liyl	10
rpl_kde	liyl	10
rpl_gb18030	liyl	10
rpl_selinux	liyl	10
chn_char	liyl	10
milwj	liyl	10
rss	liyl	10
ipsearcher	liyl	10

Continued on next page

Table 4.6: SCM results in OSS forge

Project name	Committer	N. of commits
ucms	liyl	10
gsearch	liyl	10
seekstorage	liyl	10
imageviewer	liyl	10
textunix	liyl	10
wemacs	liyl	10
bugfreej	liyl	10
thuthesis	liyl	10
codebeautifiers	liyl	10
siteadmin	liyl	10
liktron	liyl	10
xinglottery	liyl	10
radish	liyl	10
yam	liyl	10
oodb	liyl	10
openlib	liyl	10
lumaqq	liyl	10
sobar	liyl	10
iogamestar	liyl	10
opennic	liyl	10
rplinux	liyl	10
zzdd	liyl	10
delstrong	liyl	10
putty	liyl	10
elfparser	balancesli	4
sharelib	jacob	2

### HitGforge forge

Table 4.7: SCM results in HitGforge forge

Project name	Committer	N. of Commits
lilacwebex	robaggio	888
xnews	btliu	476
csamples	sun	363
lightning	cowoo	106
freeip	hexing	17
mystar	root	10
wtdic	chunyu	10
bhtrayicon	root	10
batp	chunyu	10

Continued on next page

Table 4.7: SCM results in HitGforge forge

Project name	Committer	N. of Commits
rss	root	10
irvp	chunyu	10
jfontchooser	chunyu	10
hitwebsite	root	10
rkey	root	10
guoyilau	root	10
egg	root	10
munnel	root	10
ime4mn	root	10
cscantoolkit	root	10
minidb	root	10
hitgooglecamp	root	10
mscmobile	root	10
mwbash	root	10
gpcms	root	10
lilacbbs	root	10
tiddlywiki	jjason	7
faintsendmail	cymacs	2

**Cosoft forge**

Table 4.8: SCM results in Cosoft forge

Project name	Committer	Commits
psis	iamfoolberg	3227
jvc	rufujian2005	3166
javahope	etxiaohei	3105
xtelsim	xtel	2960
myblog	somebody	1721
granite	renchen	1703
revolutionbiz	redlly	1494
fitframework	lixia	1210
ossdk	thiswind	854
freejava	jljljl	820
onlydownload	lu_zi	740
cleanmine	badqiu	710
wonder	badqiu	636
rtie	redlly	485
subteam	windshell	472
cjarticle	zjpsh	425
jave	samwise	412

Continued on next page

Table 4.8: SCM results in Cosoft forge

Project name	Committer	Commits
ogl2dsdk	kaikai	375
gavator	ajaxchelsea	347
menc	stanleyhuang	337
dpane	xinyuz	315
dbexplorer	bensonchen	312
sunrise	shinji1984	296
cntrac	xleasun	283
wxhowto	cnwesleywang	266
webstipe	nbhopson	218
mayflowererp	xthmpro_cn	166
granite	myporter	153
wicketbridge	julius_caesar	134
beex	pianopanx	125
xpdlparser	ajaxchelsea	123
srdgame	changch84	118
citydrugstore	snowmy	114
icecube	fastbit	95
papachong	papachong	94
osooncrm	jsonic	92
jfos	dato	86
free3d	strong_di	81
iriskernel	martin_25	73
sdaf	jljljl	69
granite	xiang_wang898	63
cgis	assiss	59
snmap	ardencom	56
citydrugstore	stlxv	55
sedit	sattva	50
srggame	microsoftxiao	47
gtbot	chunlinyao	39
fipai	dwfhz	35
xmlgbserv	z33	27
prototypex	dpsnet	21
replace	sattva	19
dtc	jljljl	18
delrpt	zbluecn	18
bloghome	jakieyoung	17
pomegranate	dwfhz	15
sunwb	felixsun	12
jvc	somebody	12
fanguanqun	fanguanqun	12
atl4unix	fallen_angel	11
share	herberthb	8
tomcat_res_mgr	cm4ever	7

Continued on next page

Table 4.8: SCM results in Cosoft forge

Project name	Committer	Commits
xspringdoclet	jxy	6
ccq	cc0cc	6
downovel	kalya	5
tide	akang	5
citydrugstore	rhode	5
e_mobile	joybyte	4
lanchat	ree	4
laszlo	xinye0123	4
cmss	robinfanuss	4
booklib	xinye0123	3
mlc3d	fortruth	3
eccn	xplucky	2
aio	moke	2
irss	mosane	1
citydrugstore	wanming	1

## SourceForge

Table 4.9: SCM results in SourceForge forge

Project name	Committer	N.of commits
xmlucene	bitiboy	4085
dlog4j	liusoft	3437
garnet3d	chenlee	1080
garnet3d	chenlee	1080
zhdotemacs	unidevel	1066
kylin	kylinwym	1062
gina	whistler_wmz	971
qterm	hephooey	657
ebf	funnyhao	649
openq	hzhr	644
golc	wrxzzj	638
qterm	sidos	634
greenvalley	bladesn	459
lazylucy	steeven	419
powindex	aegirsun	410
bluewater	bladesn	387
stronger	arthurkkk	379
nbt	redforks	377
opccomu	bobbercheng	350
njlug	zhaoway	346

Continued on next page

Table 4.9: SCM results in SourceForge forge

Project name	Committer	N.of commits
bugfree	wangcs	280
anole	javacat	269
freejava	jljljl	268
testtracking	chicochen	265
fiwestone	hzhzh0725	259
nvwa	adah	227
p4ulm	feeling	222
qmailadminweb	loya	167
golp	allanxu	163
vparm	ssddn	163
molee	cyfer	154
bookbot	qjzhou	137
bugfree	liuzf	130
ddutils	zweigzhang	120
olft	trydofor	120
netbbs	houzhitao	116
openq	joymarquis	99
dotnut	pkretek	97
bugfree	wwcss	92
cchess	wrxzzj	90
w3replib	soar	85
runner	xiaohu_dong	83
cvsphpview	rockeychu	82
cvsphpview	CZR	72
newsseed	newsseed	71
upstand	wilonguo	70
mssqladmin	php_sql_admin	68
cmpg	chiesaitaly	66
lrc	dasn	61
bind4j	steeven	58
openq	henryouly	57
easyswing	whl77	57
abuky	lei-ding	55
ezproject	bonami	49
buildfiledesign	powerise	47
thckpt	littleleo	47
linenumvb	redguard	44
regress	savorjava	40
msb	jljljl	38
emix	ericmiao	34
feednow	eddyxu	34
registry	yckaby	33
mec	snnn	29
xiadupdb	caihead	24

Continued on next page

Table 4.9: SCM results in SourceForge forge

Project name	Committer	N.of commits
atapi	lvjian	20
xsltgenerator	smliu	18
ued	ericmiao	17
icct	pennyge	17
chsid	jljljl	15
dlog4j	xfliang	14
tsinghua	bistoury	12
dlog4j	yhetoday	9
dlog4j	aeonsun	7
feednow	albertlee	6
zqlib	zoomq	6
oooxsltfilters	gavinlu	5
dlog4j	thatway	5
ltrue	ccden	5
vparm	jensoncox	4
kylin	dhuwym	3
upstand	stone3815	3
lincoserver	rickyu	3
evide	neosun	3
odsclientcompl	complystill	2
easyq	shaver_lee	2
msb	jack_spec	2
t9tr	t9tr	1
tarfs	liudy	1
stronger	schua	1

### 4.3 COCOMO model applied to software releases

The following is a list of tables showing data obtained after applying basic COCOMO model to projects found in the forges.

#### OSDN forge

Table 4.10: Basic COCOMO model applied to OSDN software distributions

Project name	SLOC	EPM(EPY)	Est. Years	Est Dev.	Est. Cost
xreal3d	75,705	18.80 (225.58)	1.63 (19.60)	11.51	\$ 2,539,349
ylib	38,820	9.32 (111.87)	1.25 (15.01)	7.45	\$ 1,259,360
afw	13,243	3.01 (36.17)	0.81 (9.77)	3.70	\$ 407,125
rttim	1,219	0.25 (2.95)	0.31 (3.77)	0.78	\$ 33,262

Continued on next page

Table 4.10: Basic COCOMO model applied to OSDN software distributions

Project name	SLOC	EPM(EPY)	Est. Years	Est Dev.	Est. Cost
msnchatlog	936	0.19 (2.24)	0.28 (3.40)	0.66	\$ 25,205
minilist	803	0.16 (1.91)	0.27 (3.19)	0.60	\$ 21,458
opencad	505	0.10 (1.17)	0.22 (2.65)	0.44	\$ 13,186

### OSS forge

Table 4.11: Basic COCOMO model applied to OSS software distributions

Project name	SLOC	EPM(EPY)	Est. Years	Est Dev.	Est. Cost
ck-erp	84,656	21.14 (253.66)	1.71 (20.49)	12.38	\$ 2,855,501
dget	28,977	6.86 (82.29)	1.11 (13.36)	6.16	\$ 926,398
yam	7,243	1.60 (19.19)	0.64 (7.68)	2.50	\$ 216,051
imageviewer	2,629	0.55 (6.62)	0.43 (5.13)	1.29	\$ 74,546
elfparser	2,329	0.49 (5.83)	0.41 (4.89)	1.19	\$ 65,640
yalog	1,547	0.32 (3.79)	0.35 (4.15)	0.91	\$ 42,718
sobar	74	0.01 (0.16)	0.10 (1.23)	0.13	\$ 1,755
ipsearcher	23	0.00 (0.05)	0.06 (0.77)	0.06	\$ 515

### HitGforge forge

Table 4.12: Basic COCOMO model applied to HitGForge software distributions

Project name	SLOC	EPM(EPY)	Est. Years	Est Dev.	Est. Cost
cn-moodle	306,114	81.51 (978.11)	2.85 (34.22)	28.58	\$ 11,010,803
ck-erp	84,656	21.14 (253.66)	1.71 (20.49)	12.38	\$ 2,855,501
irvp	2,540	0.53 (6.39)	0.42 (5.06)	1.26	\$ 71,898
xnews	1,330	0.27 (3.24)	0.33 (3.91)	0.83	\$ 36,449
mystar	1,274	0.26 (3.09)	0.32 (3.84)	0.81	\$ 34,839
cvtenki	1,113	0.22 (2.69)	0.30 (3.64)	0.74	\$ 30,232
csamples	1,038	0.21 (2.50)	0.29 (3.54)	0.71	\$ 28,096
sfsearch	769	0.15 (1.82)	0.26 (3.14)	0.58	\$ 20,505
egg	647	0.13 (1.52)	0.24 (2.93)	0.52	\$ 17,104
miniotter	116	0.02 (0.25)	0.12 (1.48)	0.17	\$ 2,814
pmwiki- phpbb	89	0.02 (0.19)	0.11 (1.33)	0.14	\$ 2,131

Continued on next page

Table 4.12: Basic COCOMO model applied to HitGForge software distributions

Project name	SLOC	EPM(EPY)	Est. Years	Est Dev.	Est. Cost
jfontchooser	37	0.01 (0.08)	0.08 (0.94)	0.08	\$ 848

### Cosoft forge

Table 4.13: Basic COCOMO model applied to Cosoft software distributions

Project name	SLOC	EPM(EPY)	Est. Years	Est Dev.	Est. Cost
cc-linux	4,998,903	1,530.55 (18,366.64)	8.69 (104.30)	176.10	\$ 206,756,927
ccBSD	3,942,371	1,192.82 (14,313.85)	7.91 (94.87)	150.88	\$ 161,133,926
todde-works	664,855	184.03 (2,208.38)	3.89 (46.63)	47.36	\$ 24,860,217
chinesepython	423,972	114.75 (1,376.94)	3.25 (38.97)	35.33	\$ 15,500,498
cnwizards	168,458	43.54 (522.43)	2.25 (26.96)	19.37	\$ 5,881,090
isee	147,601	37.89 (454.73)	2.13 (25.58)	17.78	\$ 5,119,003
dusk	129,275	32.97 (395.64)	2.02 (24.26)	16.31	\$ 4,453,812
atheos	104,812	26.45 (317.43)	1.86 (22.31)	14.23	\$ 3,573,330
garnet	52,669	12.84 (154.11)	1.41 (16.96)	9.09	\$ 1,734,900
itreesurf	37,123	8.90 (106.74)	1.23 (14.75)	7.24	\$ 1,201,619
reciteword	35,740	8.55 (102.57)	1.21 (14.53)	7.06	\$ 1,154,659
gnucleus	33,379	7.96 (95.47)	1.18 (14.13)	6.75	\$ 1,074,703
ebus	26,519	6.25 (74.98)	1.07 (12.89)	5.81	\$ 844,066
candyplayer	25,168	5.91 (70.97)	1.05 (12.63)	5.62	\$ 798,974
cosoft	25,056	5.89 (70.64)	1.05 (12.61)	5.60	\$ 795,241
linqt	21,387	4.99 (59.82)	0.99 (11.83)	5.05	\$ 673,440
jini	19,232	4.46 (53.51)	0.95 (11.34)	4.72	\$ 602,375
kinggate	16,306	3.75 (45.00)	0.89 (10.62)	4.24	\$ 506,531
libnet	16,265	3.74 (44.88)	0.88 (10.61)	4.23	\$ 505,194
mydict	15,701	3.60 (43.24)	0.87 (10.46)	4.13	\$ 486,817
baow	13,328	3.03 (36.41)	0.82 (9.80)	3.72	\$ 409,869
at	13,209	3.01 (36.07)	0.81 (9.76)	3.69	\$ 406,027
pku-edb	13,194	3.00 (36.03)	0.81 (9.76)	3.69	\$ 405,543
xbkconf	12,320	2.79 (33.52)	0.79 (9.50)	3.53	\$ 377,383
xml- platform	10,499	2.36 (28.34)	0.74 (8.91)	3.18	\$ 319,041
webpm	9,974	2.24 (26.85)	0.73 (8.73)	3.08	\$ 302,311
fava	7,592	1.68 (20.16)	0.65 (7.83)	2.58	\$ 226,995
yqpim	6,208	1.36 (16.32)	0.60 (7.22)	2.26	\$ 183,756
flyedit	5,841	1.28 (15.31)	0.59 (7.05)	2.17	\$ 172,367

Continued on next page

Table 4.13: Basic COCOMO model applied to Cosoft software distributions

Project name	SLOC	EPM(EPY)	Est. Years	Est Dev.	Est. Cost
cxxunit	5,681	1.24 (14.87)	0.58 (6.97)	2.13	\$ 167,412
isirwe	4,409	0.95 (11.40)	0.53 (6.30)	1.81	\$ 128,292
gaf	4,251	0.91 (10.97)	0.52 (6.21)	1.77	\$ 123,469
rexp	3,551	0.76 (9.08)	0.48 (5.78)	1.57	\$ 102,214
fractalstudio	3,500	0.75 (8.94)	0.48 (5.75)	1.56	\$ 100,673
libnids	3,009	0.64 (7.63)	0.45 (5.41)	1.41	\$ 85,898
mantischess	2,298	0.48 (5.75)	0.40 (4.86)	1.18	\$ 64,723
see	1,802	0.37 (4.45)	0.37 (4.41)	1.01	\$ 50,140
vphttp	1,762	0.36 (4.35)	0.36 (4.37)	1.00	\$ 48,972
webdict	1,386	0.28 (3.38)	0.33 (3.97)	0.85	\$ 38,062
dictionary	861	0.17 (2.05)	0.27 (3.28)	0.62	\$ 23,088
uidesigner	613	0.12 (1.44)	0.24 (2.87)	0.50	\$ 16,161
mpegplayer	334	0.06 (0.76)	0.19 (2.25)	0.34	\$ 8,542
h2x	315	0.06 (0.71)	0.18 (2.20)	0.32	\$ 8,033
monster	250	0.05 (0.56)	0.17 (2.01)	0.28	\$ 6,302
soulogic-guestbook	237	0.04 (0.53)	0.16 (1.96)	0.27	\$ 5,958
burnmp3	234	0.04 (0.52)	0.16 (1.95)	0.27	\$ 5,879
wage	91	0.02 (0.19)	0.11 (1.34)	0.14	\$ 2,181
vimdoc	32	0.01 (0.06)	0.07 (0.88)	0.07	\$ 728

### CNforge forge

Table 4.14: Basic COCOMO model applied to CNForge software distributions

Project name	SLOC	EPM(EPY)	Est. Years	Est Dev.	Est. Cost
ck-erp	84,656	21.14 (253.66)	1.71 (20.49)	12.38	\$ 2,855,501
geosings	15,329	3.51 (42.17)	0.86 (10.36)	4.07	\$ 474,713
spadger	6,827	1.50 (18.04)	0.63 (7.50)	2.40	\$ 203,041
mybomber	3,897	0.83 (10.01)	0.50 (6.00)	1.67	\$ 112,696
pmvc	834	0.17 (1.98)	0.27 (3.24)	0.61	\$ 22,329