

Use case of source code clones detection

A use case of source code clones detection between software projects with FLOSS tools

Luis Cañas-Díaz, Santiago Dueñas

{lcanas, sduenas}@libresoft.es

February 3rd, 2012

LibreSoft
we study libre software

GSyc



Universidad
Rey Juan Carlos

GSyC

©2012 Some rights reserved. GSyC/LibreSoft, Universidad Rey Juan Carlos

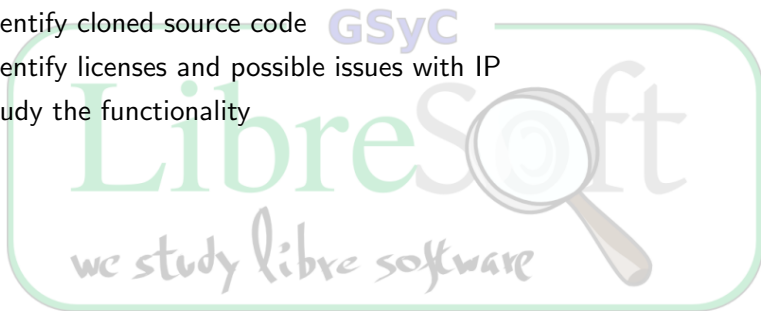
Slides licensed under a Creative Commons Attribution 3.0 Unported License



we study libre software

- Study of gisEIEL and gvSIG-EIEL , both with similar features
- gisEIEL:
 - gisEIEL was created in 2000 and paid by regional council of A Coruña
 - gisEIEL was released in 2004 as FLOSS based on gvSIG
- gvSIG-EIEL:
 - Years later .. did they reinvent the wheel?
 - Company hired by the regional council of Pontevedra performed a kind of a fork
 - gvSIG-EIEL was born with same functionality, based on gvSIG

- identify cloned source code
- identify licenses and possible issues with IP
- study the functionality



- source code to be analysed
- tools to analyse the source code
- process the raw data



CCFinder

- <http://www.ccfinder.net/>
- CCFinder allows to match similar parts of the code
- works at token level
- must be configured

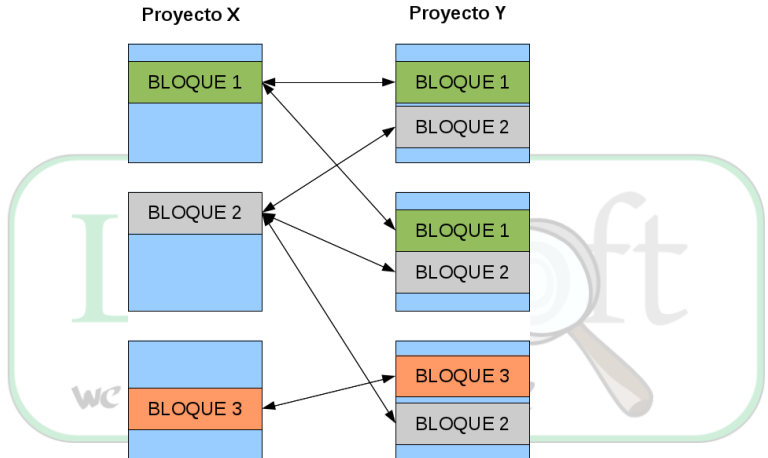
Cloc

- gets the SLOC
- 86 programming languages
- are you still using sloccount?!

Raw data from CCFinder

clone id	file id.tokens	file id.tokens
16359	476.1119-1177	2093.644-702
16359	476.1119-1177	2093.749-807
16359	476.1119-1177	2093.889-947
16359	476.1119-1177	2093.1034-1092
16359	476.1119-1177	2093.1181-1239
1207	476.1259-1310	2093.1324-1375
36	476.37-149	2094.37-149
1831	476.260-326	2094.221-287

How much code in common?



Ninka

- <http://ninka.turingmachine.org/>
- lightweight license identification tool for source code

Grep

- I guess you know it .. :P
- searches files for text strings

We combined the tools to get this ..

For this gisEIEL file:

406 gisEIEL: ExportMapTo.java (cloned files: 3) - SLOC:569 - GPLv2+
;Copyright (C) 2009 Deputación de A Coruña

we have the following clones:

1989 gvsig-eiel/./ExportSeveralTo.java (43 %) - SLOC_cloned:244 -
NONE; noC

1990 gvsig-eiel/./StopEditingToShp.java (28 %) - SLOC_cloned:159 -
NONE; noC

2046 gvsig-eiel/./ExportTo.java (47 %) - SLOC_cloned:267 - NONE; noC

gisEIEL vs. gvSIG-EIEL (1/2)

Module	SLOC	similar SLOC	%
appgvSIG	48279	483	1
EIEL-Autenticacion	1062	0	0
EIEL-DescargaMunicipiosBD	3142	0	0
EIEL-extCAD	21423	13068	61
EIEL-Formularios-Alfanumer	27224	0	0
EIEL-GeneracionScriptsInBDT	3992	0	0
EIEL-GestionDeLeyendasImpr	980	0	0
EIEL-GestionDeMapasGisEIEL	936	0	0
EIEL-GestionPermisos	776	0	0
EIEL-GestionUsuarios	1517	0	0

gisEIEL vs. gvSIG-EIEL (2/2)

Module	SLOC	similar SLOC	%
EIEL-GisEIEL	22906	687	3
EIEL-Informes	935	0	0
EIEL-Utilidades	1146	23	2
EIEL-Validaciones	3487	0	0
extJDBC	3600	36	1
extOracleSpatial	9034	90	1
_fwAndami	13886	0	0
libCorePlugin	3510	35	1
libCq CMS for java	26617	0	0
libFMap	41159	0	0

Cloned code in gisEIEL (around 6 % of total)

Módulo	SLOC	SLOC similar	%
extDBConnection	1648	0	0
ELLE	3459	35	1
OpenCADTools	36974	15899	43
NavTable	5685	57	1
exteieltable	8311	83	1
extvalidation	1160	0	0
exteieltutils	1711	0	0
exteielforms	8185	82	1

Cloned code in gvSIG-EIEL (around 25 % of total)

- 25 % of the gvSIG-EIEL code can be found in gisEIEL
- most of the cloned code in gvSIG-EIEL is part of the module *OpenCADTools*. 43 % of this code can be found in the gisEIEL module *EIEL-extCAD*
- in 91 % of the files that are derivative works of gisEIEL (in gvSIG-EIEL) the original copyright holder is not present
- some of the files of gisEIEL in the early versions were copied from the gvSIG project and the copyright holder was removed from files (already solved in a latest version)

we study libre software

