

Estimación de Esfuerzo de Desarrollo en Proyectos Free/Open Source Software (FOSS) Mediante Minería de Repositorios Software

Autor: Carlos Cervigón Ávila
Tutor: Gregorio Robles Martínez

Universidad Rey Juan Carlos

Proyecto Fin de Carrera





(cc) 2014 The authors
Some rights reserved. This work licensed under Creative Commons
Attribution-ShareAlike License. To view a copy of full license, see
<http://creativecommons.org/licenses/by-sa/3.0/> or write to
Creative Commons, 559 Nathan Abbott Way, Stanford,
California 94305, USA.

Some of the figures have been taken from the Internet
Source, and author and licence if known, is specified.

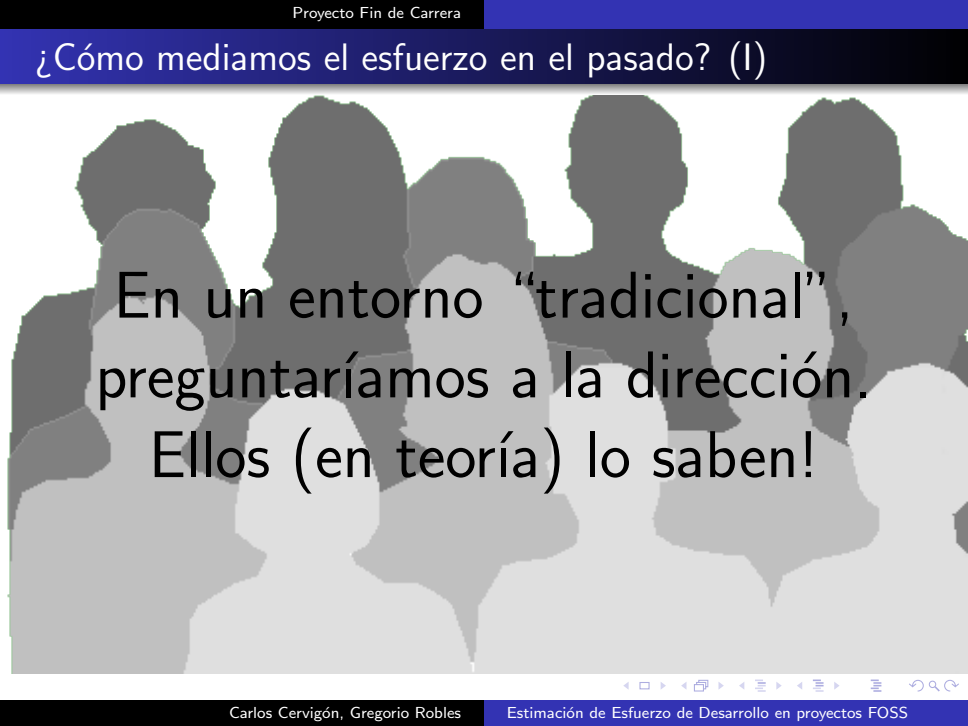
For those images, *fair use* applies.

La cuestión es...

A hand is shown holding a fan of several 500 Euro banknotes. The notes are fanned out, showing the number '500' and the European Union flag. The text '¿Cuánto dinero cuesta desarrollar un software?' is overlaid on the image.

¿Cuánto dinero cuesta desarrollar un software?

¿Cómo mediamos el esfuerzo en el pasado? (I)

The background of the slide features a collection of overlapping, semi-transparent grey silhouettes of people's heads and shoulders, representing a diverse group of individuals. The silhouettes are arranged in a way that some are in the foreground and others are behind, creating a sense of depth and a crowd.

En un entorno “tradicional”,
preguntaríamos a la dirección.
Ellos (en teoría) lo saben!

¿Cómo mediamos el esfuerzo en el pasado? (y II)



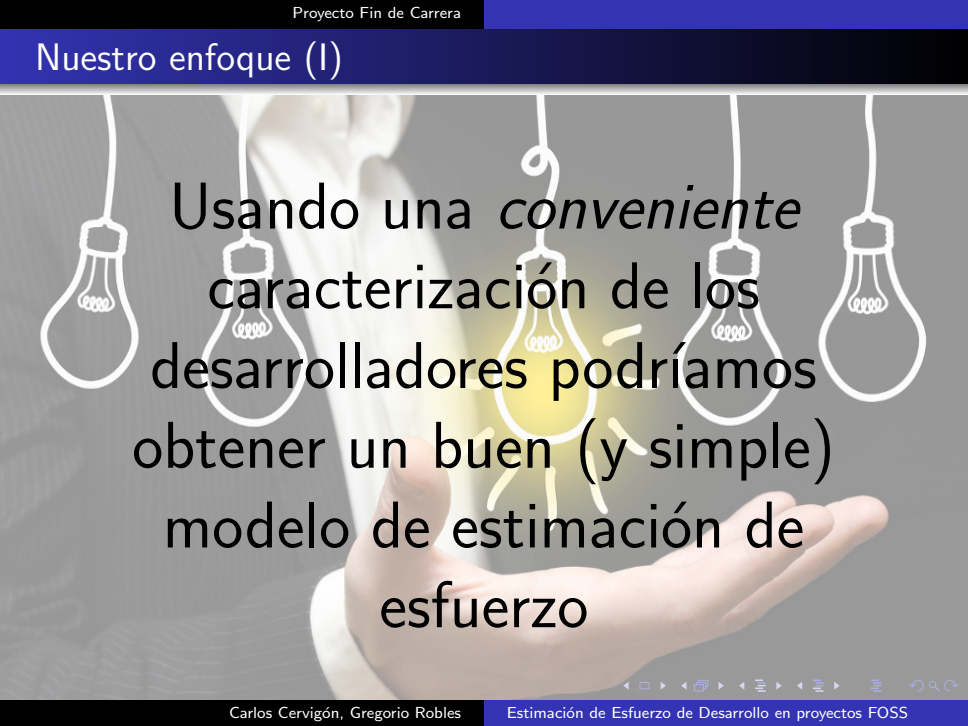
¿Pero qué hay de los proyectos
Free/Open Source Software?



Primera (probablemente mala) idea

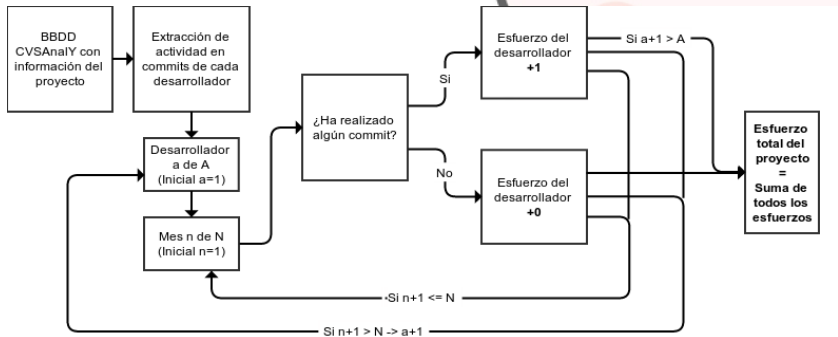
Podríamos rastrear y sumar las contribuciones de los desarrolladores en todos los repositorios

Nuestro enfoque (I)

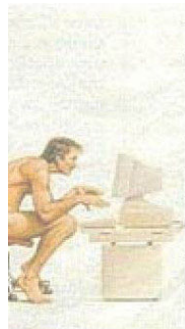
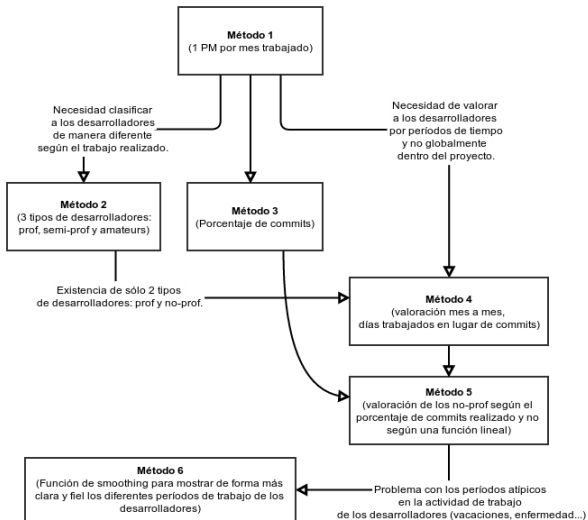
A hand in a white suit sleeve is shown holding a glowing yellow lightbulb. Several other white-outlined lightbulbs are hanging from above. The background is a blurred image of a person in a white shirt and tie.

Usando una *conveniente* caracterización de los desarrolladores podríamos obtener un buen (y simple) modelo de estimación de esfuerzo


Nuestro enfoque (II): Primeros pasos



Nuestro enfoque (y III): Evolución

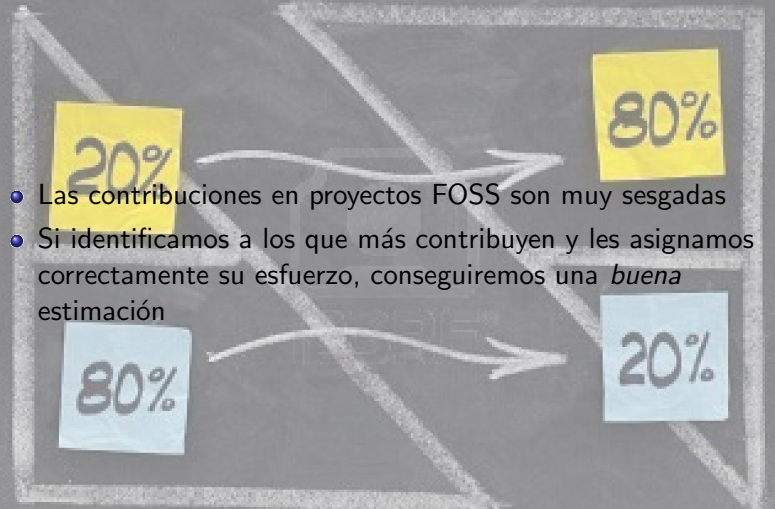


Método final: Dividir los desarrolladores en dos grupos



Desarrolladores Full-time y non-full-time

Porqué?

- 
- Las contribuciones en proyectos FOSS son muy sesgadas
 - Si identificamos a los que más contribuyen y les asignamos correctamente su esfuerzo, conseguiremos una *buena* estimación

El Modelo

- 1 Definimos un período de tiempo T
- 2 Cualquier desarrollador con más commits que un threshold t en el período T será considerado full-time:

$$\text{Esfuerzo (en personas-mes)} = \begin{cases} T & \text{si es full-time} \\ T * x/t & \text{en otro caso,} \\ & x = \text{commits en } T \end{cases}$$

Un ejemplo

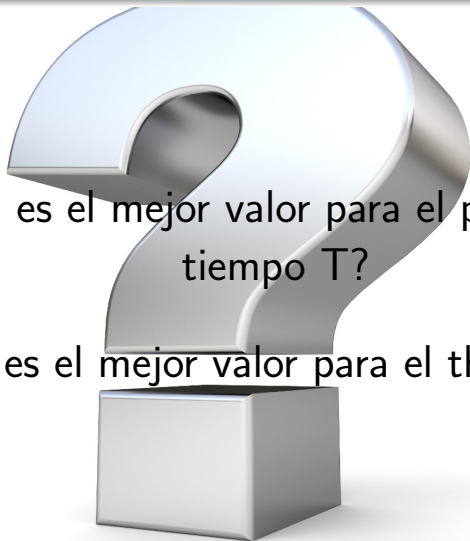
- 1 Un proyecto con 3 desarrolladores
- 2 El número de commits realizado por cada uno en los últimos 3 meses son 100, 10 y 3
- 3 Ahora asumimos un período $T = 3$ meses y un threshold $t = 30$ commits
- 4 El esfuerzo empleado en el proyecto en los últimos 3 meses es
$$3 * (1 + 1/3 + 3/30)$$
$$= 3 * 1.433 = 4.3 \text{ personas-mes}$$

Caso de estudio: OpenStack



Conocemos el proyecto muy bien y además conocemos gente en él.

¿Qué estamos buscando?



¿Cuál es el mejor valor para el período de tiempo T ?

¿Cuál es el mejor valor para el threshold t ?

Enfoque Naïve

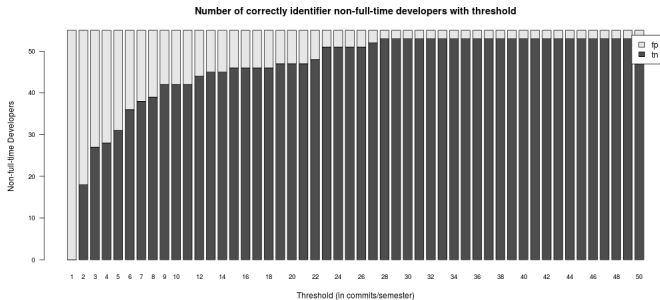
$T = 1 \text{ release} = 6 \text{ meses}$

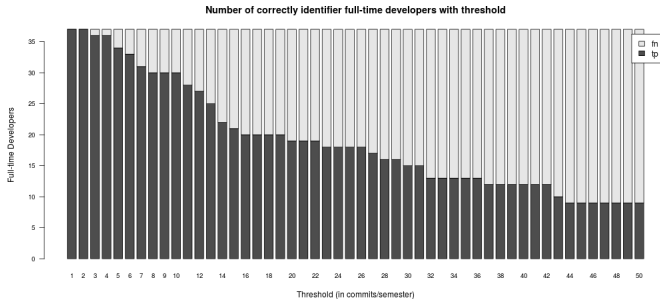
$t = 1 \text{ commit}$

(usado por la comunidad de OpenStack como estimación naïve)

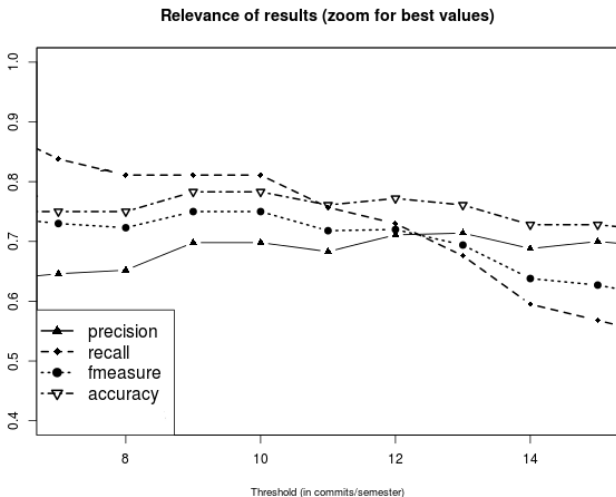
¿Cómo hemos procedido?

- Obtenemos los datos de su repositorio git
- Buscamos el período T : 6 meses (siguiendo la periodicidad de las release de OpenStack)
- Buscamos el threshold t : Preguntamos a los desarrolladores (activos) sobre su esfuerzo en una encuesta: ¿éres full-time en OpenStack?
 - Más de 100 respuestas obtenidas! ($> 10\%$ de los desarrolladores activos de OpenStack, muestra estadísticamente representativa de la población completa)

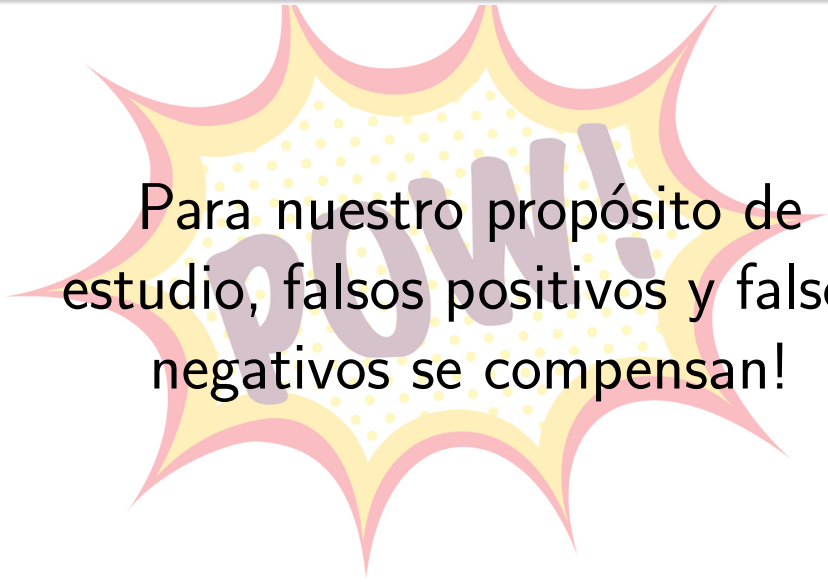




Indicadores: precision, recall,...

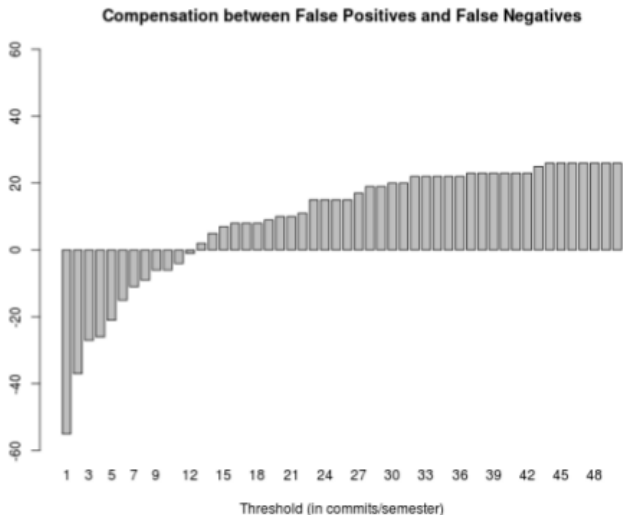


Curiosamente...

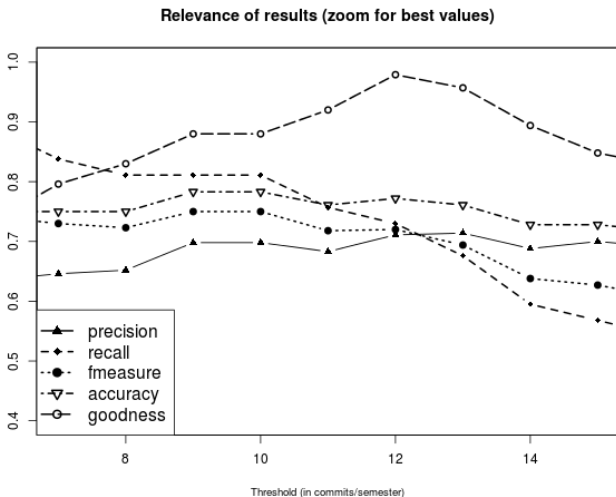


Para nuestro propósito de estudio, falsos positivos y falsos negativos se compensan!

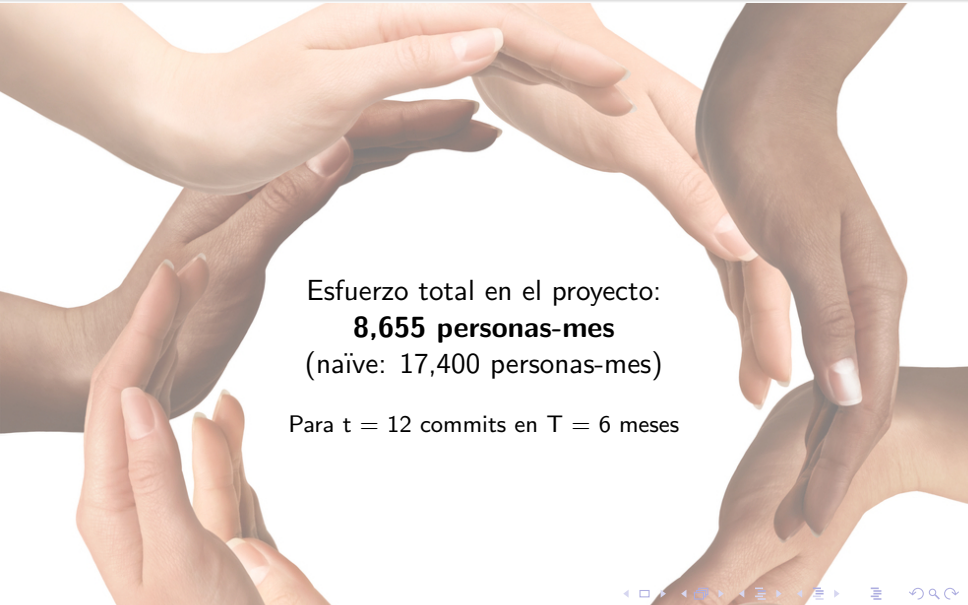
Falsos positivos y falsos negativos se compensan!



Falsos positivos y falsos negativos se compensan!



Resultados (I)



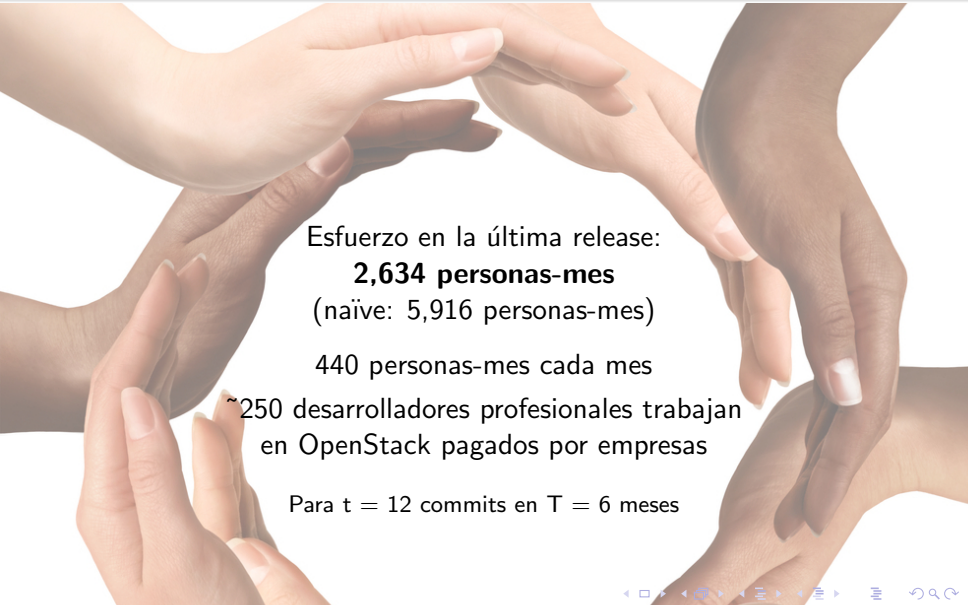
Esfuerzo total en el proyecto:

8,655 personas-mes

(naïve: 17,400 personas-mes)

Para $t = 12$ commits en $T = 6$ meses

Resultados (II)



Esfuerzo en la última release:

2,634 personas-mes


(naïve: 5,916 personas-mes)

440 personas-mes cada mes

~250 desarrolladores profesionales trabajan
en OpenStack pagados por empresas

Para $t = 12$ commits en $T = 6$ meses

Resultados (y III)



Esfuerzo total en el proyecto
(empleando días activos):
6,726 personas-mes

Desarrolladores profesionales detectados
últimos 6 meses:

260, 271, 273, 266, 240, 215

Media: **254** desarrolladores profesionales

Para $t = 2$ días activos en $T = 1$ mes

Conclusiones

Dado un repositorio, proporcionamos una vía **simple** de estimación de esfuerzo de desarrollo **realizado**

Dos parámetros tienen que ser determinados:
período de tiempo T y threshold t

Un valor de $t = 12$ commits en $T = 6$ meses,
o $t = 2$ días activos en $T = 1$ mes
han sido obtenidos como lo mejor para OpenStack

Avance del trabajo actual

Mejor valor de threshold t para los siguientes proyectos

Ceph: 24 commits cada 6 meses
24 de 135 (17.78%) desarrolladores respondieron

Linux: 17 commits cada 6 meses
652 de 3,187 (20.46%) desarrolladores respondieron

MediaWiki: 29 commits cada 6 meses
94 de 449 (20.94%) desarrolladores respondieron

WebKit: 17-25 commits cada 6 meses
85 de 661 (12.86%) desarrolladores respondieron

Moodle: 14 commits cada 6 meses
42 de 152 (27.63%) desarrolladores respondieron

Trabajo Futuro

Mecanismo de obtención de T y t óptimos

Obtención de margenes de error

¿¿¿Commits o Días activos???



Future →

Comparar resultados con métodos
tradicionales

Estimating Development Effort in Free/Open Source Software Projects by MSR

A Case Study of OpenStack

Gregorio Robles, Jesús M. González-Barahona, Carlos Cervigón,
Andrea Capiluppi, Daniel Izquierdo-Cortázar

Universidad Rey Juan Carlos, Brunel University, Bitergia S.L.

MSR 2014, Hyderabad, June 1st 2014



Estimación de Esfuerzo de Desarrollo en Proyectos Free/Open Source Software (FOSS) Mediante Minería de Repositorios Software

Autor: Carlos Cervigón Ávila
Tutor: Gregorio Robles Martínez

Universidad Rey Juan Carlos

Proyecto Fin de Carrera

