C++ Meetup Madrid

Febrero 2019



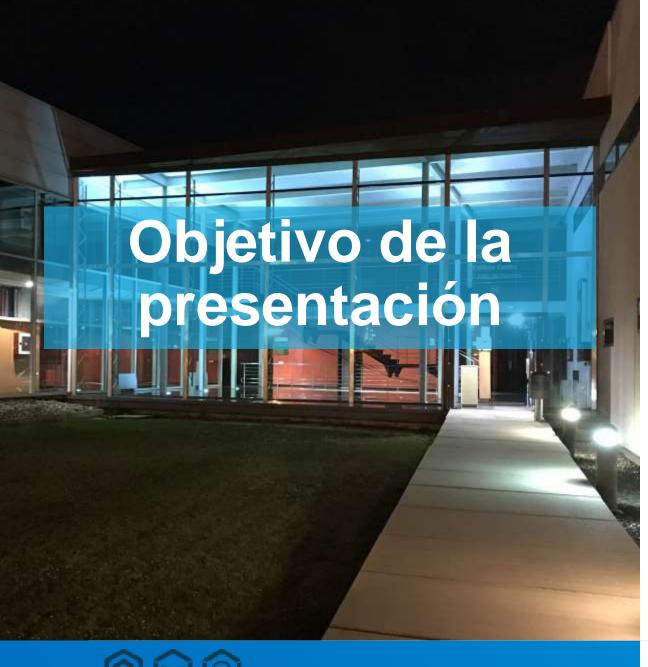


Violeta Sánchez Software Engineer @vio_woman

> Míryam Gómez Software Engineer @miryamgsm

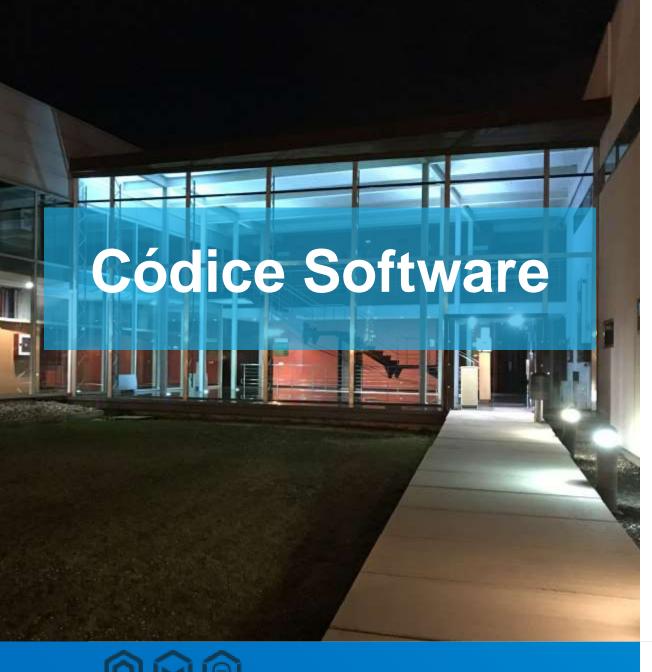






- Introducción: empresa y productos.
- Motivación del desarrollo de un parser de C++.
- Parser de C++ basado en Eclipse CDT.
- Aplicaciones prácticas:
 - Herramienta de merge: semanticmerge
 - Control de versiones: plastic y gmaster
 - Bot que resuelve conflictos de PR sin intervención manual: mergedroid





Quiénes somos y qué productos hacemos



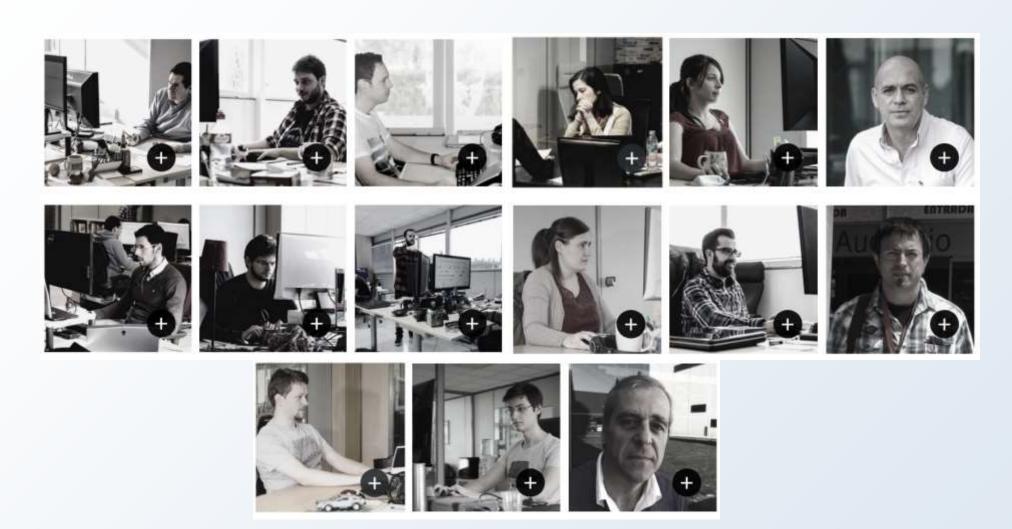
Códice Software

- Arrancamos en 2005 (llevamos en control de versiones 14 años!!)
- Situados en el Parque Tecnológico de Boecillo (Valladolid)





https://www.plasticscm.com/company/team.html





Clientes en más de 20 países

Financial Sector



Automobile & Transportation

























Defense & Gov.









Games & 3D Design











WARGAMING.NET

codemasters



Medical Devices







Utilities, Logistics & Misc











Productos









Distributed Version Control System

Essential tool for software development

Semantic Merge

Language aware diff & merge tool for any VCS

gmaster

A Git Client Tool with all UI & Semantic Technology (Beta)

Mergedroid

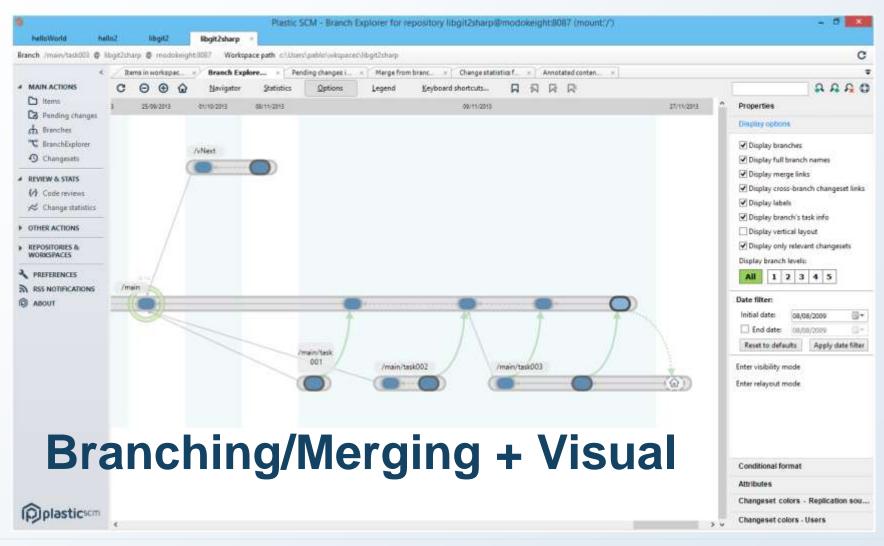
A server side semantic merge engine tool (Beta)





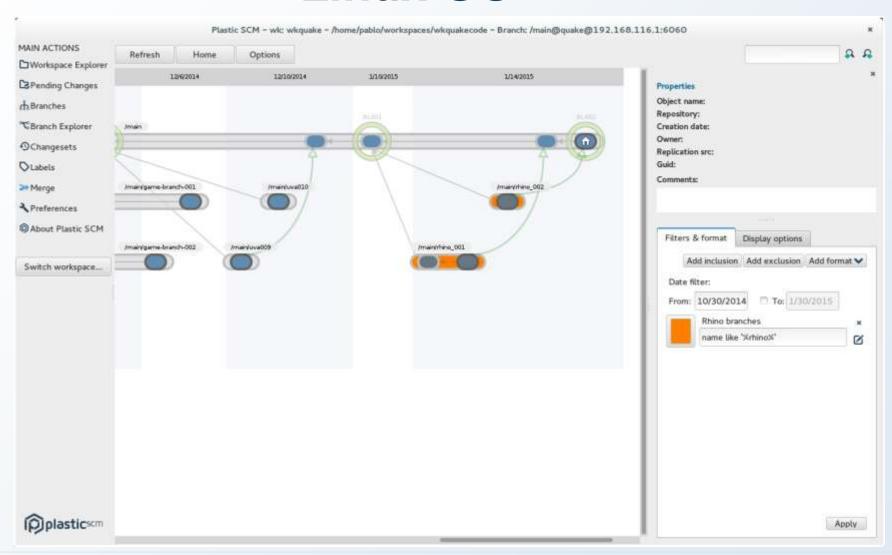






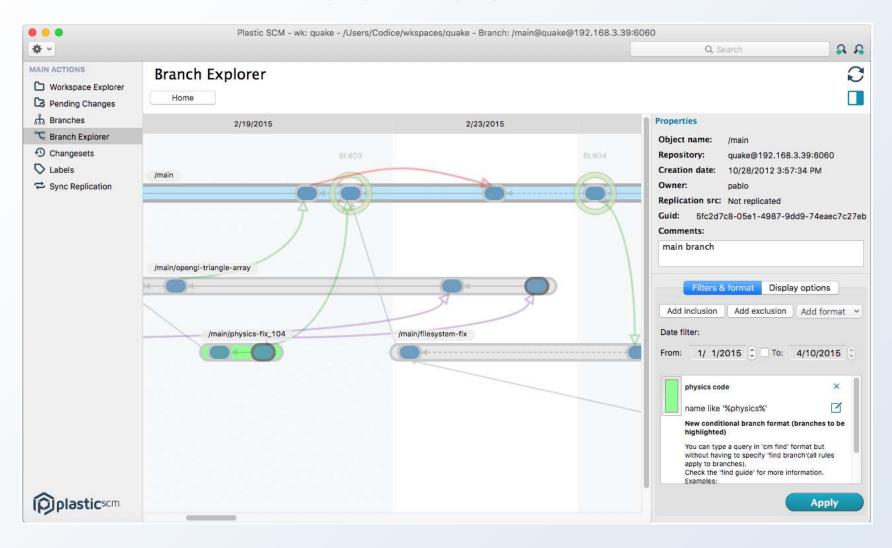


Linux GUI





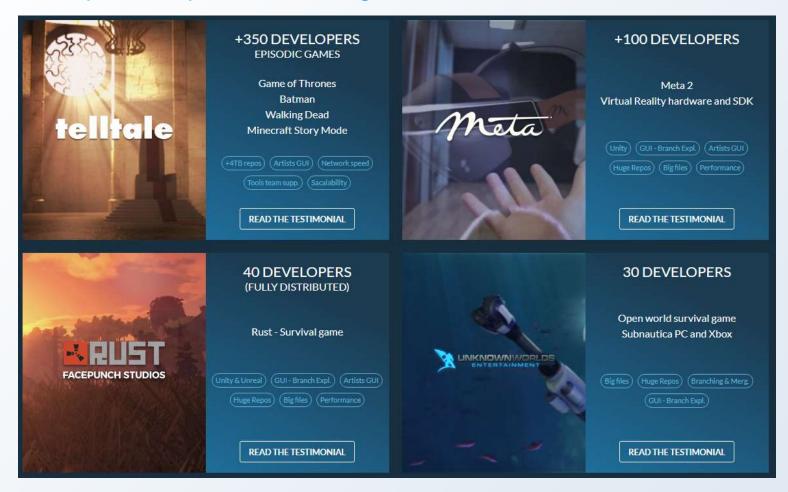
OS X GUI





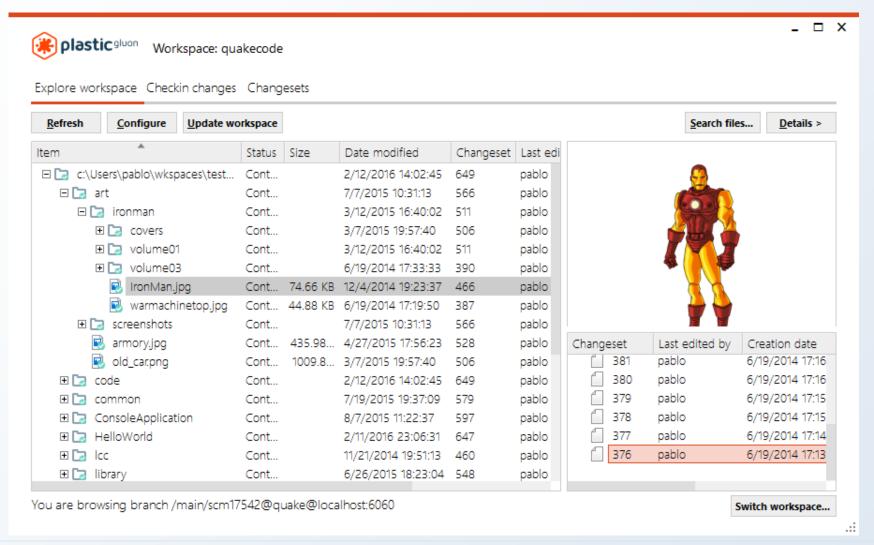
Muchos en vídeo juegos

https://www.plasticscm.com/games/success-stories/index.html





Gluon (Windows – artistas games)











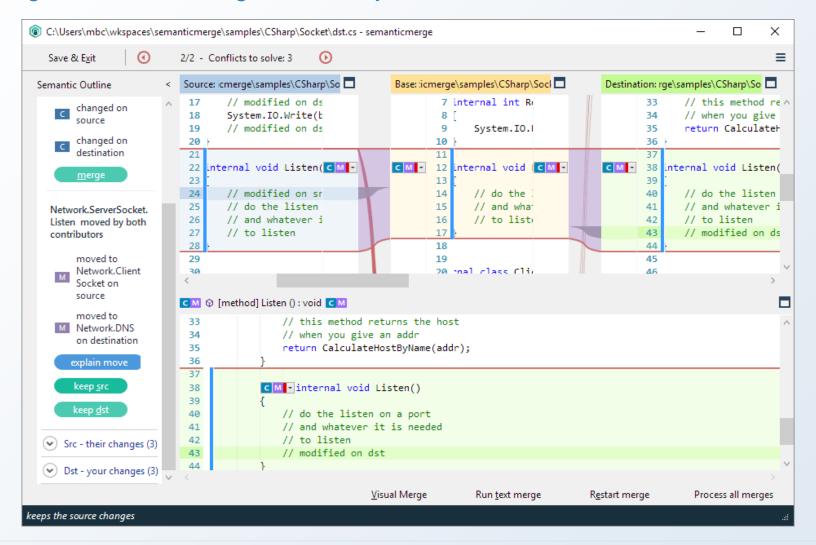






Semantic Merge

Language aware diff & merge tool for any VCS

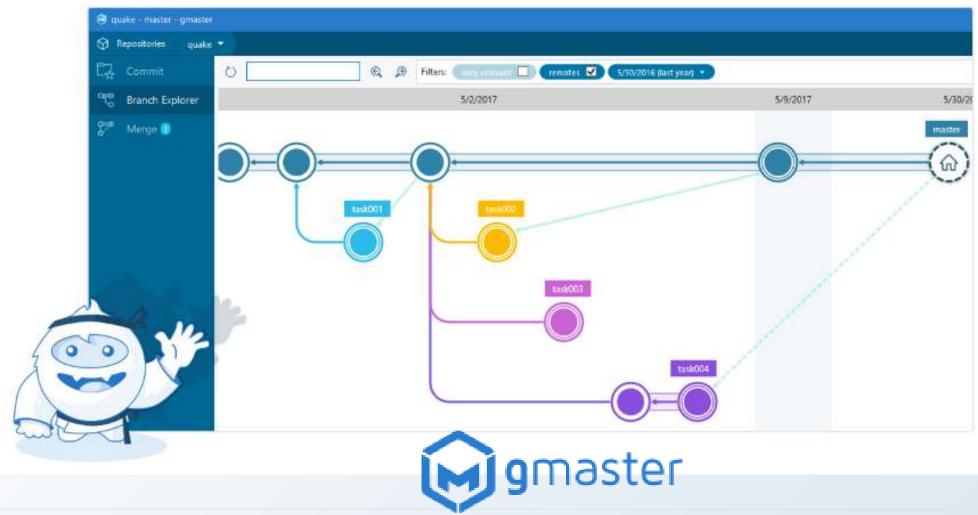




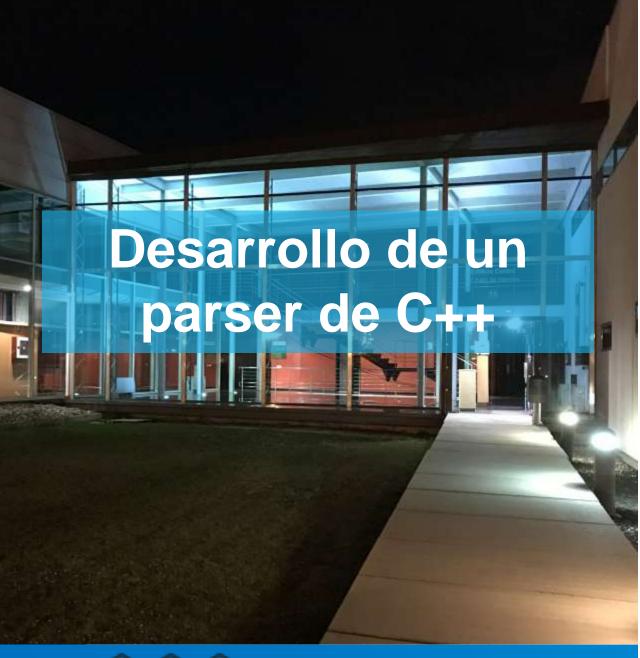


gmaster

A Git Client Tool with all UI & Semantic Technology (Beta)









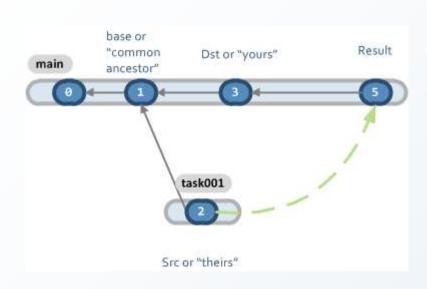
El reto a conseguir

- Principales características de PlasticSCM:
 - Gestión de ramas: patrón rama por tarea.
 - Sistema de merge: a nivel de estructura de directorios, ficheros y estructura dentro del fichero.
- Reto: crear una buena herramienta de merge.
- Objetivo: hacer el proceso de refactorización sencillo.

 Hasta ahora todas las herramientas de merge se basan en texto, reconocen línea a línea sin entender el código.

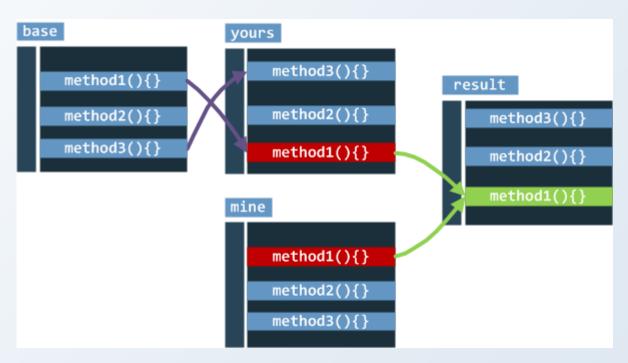


El merge a resolver



Fases del merge:

- Detección de conflictos a nivel de estructura de directorios.
- Detección de conflictos de ficheros.
- Resolución de conflictos dentro de ficheros.



Así surgen nuestras herramientas de merge:

- 1. Xmerge: basada en texto que sabe identificar fragmentos de código movido.
- 2. SemanticMerge: basada en estructura. El siguiente nivel!



SemanticMerge – cómo funciona

Es una herramienta dependiente del lenguaje a procesar

Entrada

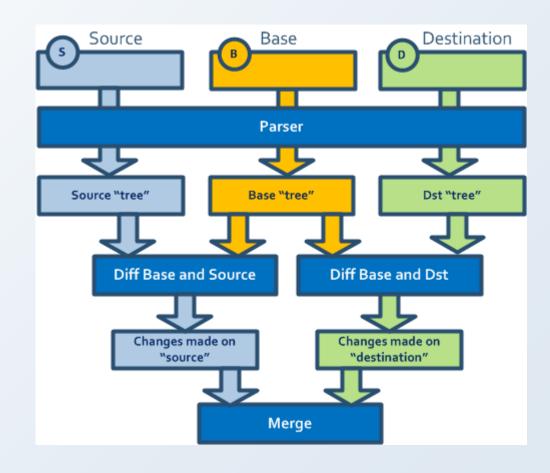
Las tres versiones del fichero, una por cada contribuidor.

Proceso

- 1. Parsear código fuente para obtener las estructuras.
- Calcular diferencias entre las estructuras de código para obtener los cambios.
- 3. Realizar el merge de los cambios para obtener los conflictos si los hubiese.

Salida

Fichero resultado del merge.



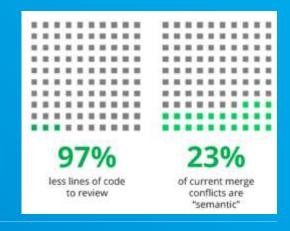


¿Por qué desarrollar un parser de C++?

La importancia de entender el código

- Motivo principal: entender el código para poder mezclarlo.
- ¿Qué nos aporta?
 - Mejorar las operaciones de diff y merge.
 - Facilita la realización de tareas que se evitan por miedo a los conflictos.
 - Incremento en la resolución de conflictos de forma automática.
 - Minimizar los tiempos de integración en un proyecto.

¿Se os ocurre alguna más? ¡Quizás sí! Lo veremos más adelante ©



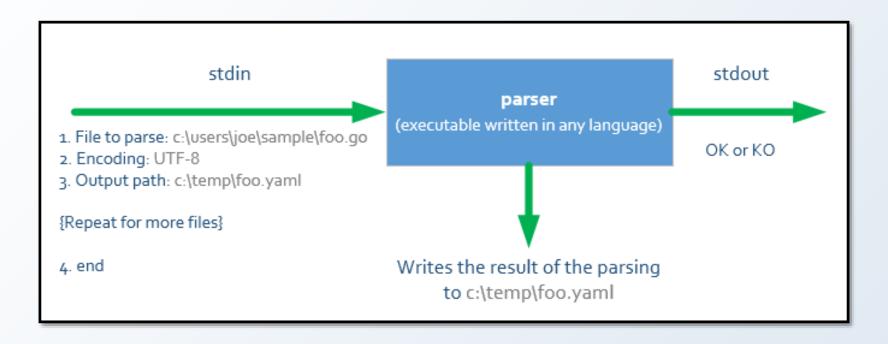






SemanticMerge - especificación de parsers

API para parser externos

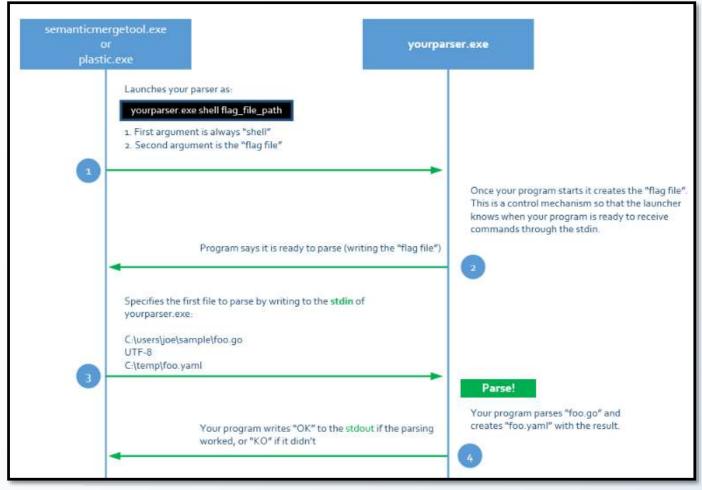


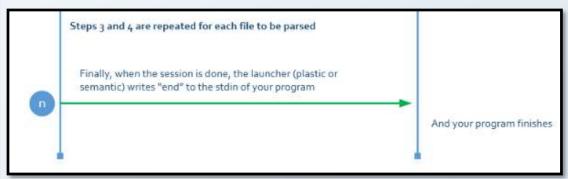
Un parser externo es una aplicación de línea de commandos que se comunica con SemanticMerge a través de un API



SemanticMerge - especificación de parsers

¿Cómo es una sesión de parseo?

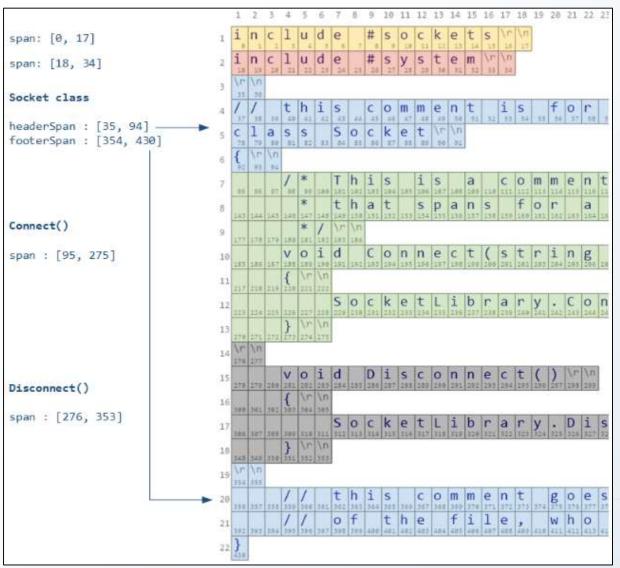




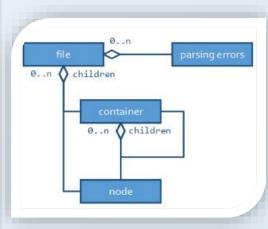


SemanticMerge - especificación de parsers

Cómo construir el árbol de declaraciones



```
type: file
name: sample\source.code
locationSpan : {start: [1,0], end: [22,1]}
footerSpan : [0,-1]
parsingErrorsDetected : false
children:
  - type : include
    name : sockets
    locationSpan : {start: [1, 0], end: [1, 18]}
    span : [0, 17]
   type : include
    name : system
    locationSpan : {start: [2, 0], end: [2,17]}
    span : [18, 34]
   type : class
    name : Socket
    locationSpan : {start: [3,0], end: [22,1]}
    headerSpan : [35, 94]
    footerSpan : [354, 430]
    children :
    - type : method
      name : Connect
      locationSpan : {start: [7, 0], end: [13,6]]
      span : [95, 275]
      type : method
      name : Disconnect
      locationSpan : {start: [14,0], end: [18,6]]
      span : [276, 353]
```



¿Por qué usar Eclipse CDT?

Evaluación de las distintas alternativas

- Parsers disponibles: C#, VB.net (Roslyn), Java (Eclipse JDT) y C (libclang).
- Pasaron unos años, hasta que en 2017... abordamos C++!
- Estudiamos distintas alternativas:
 - Antlr for C++, semanticdesigns, gccxml:
 - Fueron descartadas rápidamente (la gramática no era completa, necesidad de compilar para obtener el AST, etc.)
 - libclang:
 - Problema: el árbol AST que generaba no era completo porque no incluía los nodos con tipos desconocidos.
 - Tenía dependencia de los ficheros cabecera y de las librerías del sistema.
 - Nos encontramos con el clásico problema de parseo: A B(C) ¿declaración de función o definición de objeto?
 - Y es que C++ es un lenguaje dependiente del contexto.
 - El proceso de parseo y análisis semántico no se puede separar, cuando encuentra un tipo desconocido el nodo no se incluyen en el AST.
 - La solución pasaría por incluir las cabeceras para poder resolver los tipos.
 - En nuestro caso no es viable porque tendríamos que indicar dónde localizarlas y además estar disponibles.
 - En nuestra situación no es posible porque necesitamos los 3 contribuidores y no están en disco (control de versiones).



¿Por qué usar Eclipse CDT?

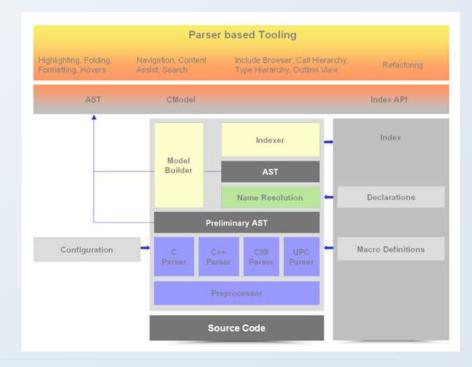
Evaluación de las distintas alternativas

- Eclipse C/C++ Development Toolkit (CDT):
 - Necesitamos el análisis semántico sintáctico del código para parsear todo el fichero y obtener su estructura.
 - Justo lo que hace Eclipse CDT, no es un compilador sino un IDE cuyo requisito principal es el rendimiento.
 - Es lo que permite al editor hacer folding, formatear un documento o mostrar la vista de outline.

Las fases incluyen:

- Escaneo y preprocesamiento: convierte caracteres a tokens y se hace el preprocesamiento (expansión de macros, compilación condicional, inclusión de cabeceras)
- 2. Parseo: convierte los tokens en un AST + location map.

- https://wiki.eclipse.org/CDT/designs/Overview of Parsing
- https://wiki.eclipse.org/images/c/c7/CDT_APIs_for_code_introspection.pdf





Parser de C++ basado en Eclipse CDT

Cómo funciona

Conceptos

TranslationUnit: fichero de código fuente + cabeceras.

El árbol AST se puede analizar recorriéndolo de forma transversal (ASTVisitor).

Entrada

Fichero de código fuente + encoding

Proceso

- Crear el parser.
- Obtener el TranslationUnit.
- Crear el ASTVisitor y establecer propiedades.
- Lanzar el recorrido a través del árbol:
 - Obtener el mapa de comentarios.
 - Obtener las macros y directivas de preprocesador.
 - Generar las declaraciones.
 - Obtener los errores de parseo.
- Convertir las directivas condicionales en contenedores.

Salida

Fichero de resultado con el árbol de declaraciones.

```
    PlasticVisitor

     PlasticVisitor(String, String)
     visit(IASTTranslationUnit): int
     • Leave(IASTTranslationUnit): int
```

- visit(ICPPASTNamespaceDefinition): int
- Leave(ICPPASTNamespaceDefinition): int
- visit(IASTDeclaration): int
- leave(IASTDeclaration) : int

```
public boolean process() {
        Parser parser = new Parser(mInputFilePath, mInputFileEncoding);
       if(parser.getCode() == null)
           return false;
       final IASTTranslationUnit translationUnit = parser.parse();
       PlasticVisitor visitor = new PlasticVisitor(mInputFilePath, parser.getCode());
       visitor.shouldVisitDeclarations = true;
        visitor.includeInactiveNodes = true;
       translationUnit.accept(visitor);
        FileDescriptor fileDescriptor = visitor.getFileDescriptor();
       ConditionalDirectiveConverter.convertToContainerDeclaration(
           fileDescriptor.getChildren());
        TreeDescriptorWriter descriptorWriter = new TreeDescriptorWriter(mOutputFilePath);
        descriptorWriter.write(fileDescriptor);
   } catch (Exception e) {
        return false;
   return true;
mNodeCommentMap = ASTCommenter.getCommentedNodeMap(mTranslationUnit);
```

```
mTranslationUnit.getMacroDefinitions();
IASTProblem[] problems = mTranslationUnit.getPreprocessorProblems();
```



Parser de C++ basado en Eclipse CDT

Problemas encontrados

Macros

Hay que manejar los distintos escenarios:

¿están dentro de una declaración? ¿definen una declaración?

Directivas condicionales

- Son un grupo contenedor.
- Hay que procesar los nodos inactivos.

Comentarios

Los comentarios son parte de la declaración.

Detección de errores de parseo

Por ejemplo, si falta un ';' al final de una instrucción

```
static void __init func(void) /* macro inside the function header
MACHINE_START(W90P910EVB, "W90P910EVB")
   /* struct declaration */
   /* Maintainer: Wan ZongShun */
                 = W90X900 PA UART,
   .io pg offst = (((u32)W90X900 VA UART) >> 18) & 0xfffc,
   boot params.
MACHINE END
#if PY MAJOR VERSION < 3
    void func(int, int, int) {
        process(a);
    void func2(int, int)
        process(a):
#else
    void func(int) {
        unprocess(b);
```



Parser de C++ basado en Eclipse CDT

Bonus track!

Regla de oro

Cualquier carácter del fichero tiene que estar contenido en una declaración de forma que a partir del conjunto de las mismas se pueda construir el fichero resultado.

Check salva-vidas:P

• Comprobación de reconstrucción del fichero a partir del árbol de declaraciones, en caso de detectar algún problema se lanza Xmerge.

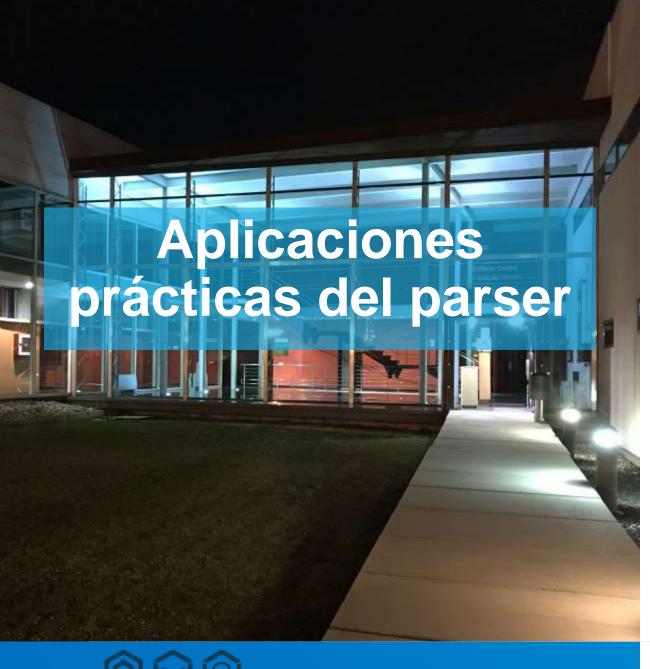
Testing

Millones de ficheros parseados de los repositorios más relevantes de GitHub.

Siguientes pasos

• Formateador de declaraciones para automatizar conflictos por cambios de formato.





- Herramienta de merge:
 semanticmerge
- Control de versiones:
 plasticscm y gmaster
- Bot que resuelve conflictos de PR sin intervención manual: mergedroid

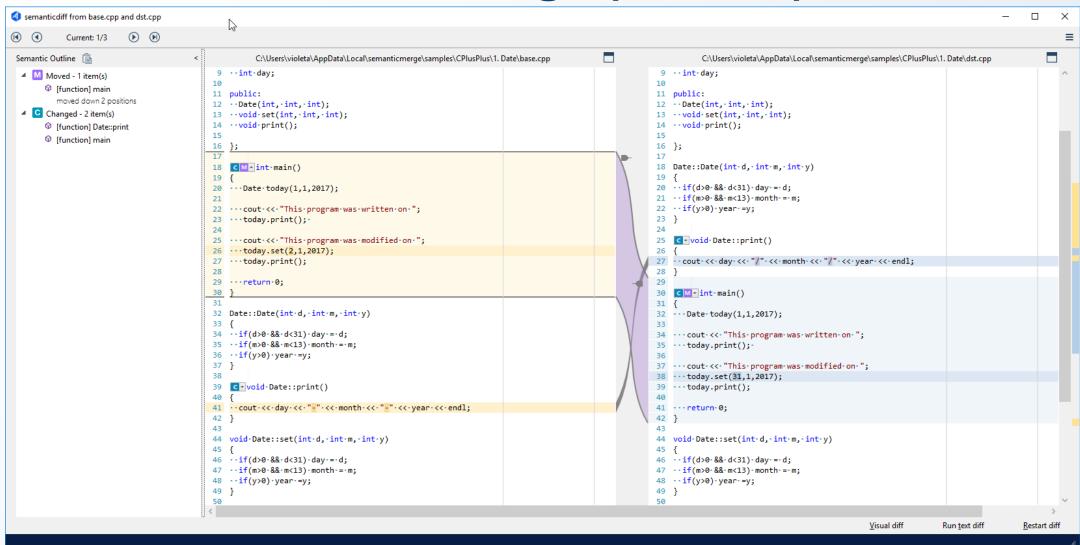


SemanticMerge (srcdiffs)

```
semanticdiff from base.cpp and src.cpp
                                                                                                                                                                                                                                        Current: 1/3 ( )
                                                                                                                                                                                                                                               \equiv
                                                                                                                                         Semantic Outline 🖺
                                                          C:\Users\violeta\AppData\Local\semanticmerge\samples\CPlusPlus\1. Date\base.cpp
                                                                                                                                                              C:\Users\violeta\AppData\Local\semanticmerge\samples\CPlusPlus\1. Date\src.cpp
                                                                                                                                                     9 ..int.day;
                                                  10
 Moved - 1 item(s)
                                                                                                                                                     10
                                                  11 public:
       @ [function] Date::print
                                                                                                                                                    11 public:
                                                  12 ··Date(int, int, int);
          moved to Date and renamed to print
                                                                                                                                                     12 ··Date(int, int, int);
                                                  13 ··void·set(int, int, int);
  Changed - 1 item(s)
                                                                                                                                                     13 ··void·set(int, int, int);
                                                  14 ··void·print();
                                                                                                                                                    14 ··void·print();
                                                  15
       @ [function] main
                                                  16 };
  Renamed - 1 item(s)
                                                                                                                                                    16 M-void-print()
       @ [function] Date::print
                                                                                                                                                    17
                                                  18 C-int·main()
          moved to Date and renamed to print
                                                                                                                                                         ··cout·<<·day·<<·"-"·<<·month·<<·"-"·<<·year·<<·endl;
                                                  20 ···Date·today(1,1,2017);
                                                                                                                                                    20
                                                                                                                                                    21 };
                                                  22 ···cout·<< "This·program·was·written·on·";</pre>
                                                                                                                                                    22
                                                  23 ···today.print();
                                                                                                                                                    23 [-int-main()
                                                                                                                                                    24 {
                                                  25 ···cout·<<·"This·program·was·modified·on·";</pre>
                                                                                                                                                        · · · Date · today(1,1,2017);
                                                  26 ...today.set(2,1,2017);
                                                  27 ···today.print();
                                                                                                                                                    27 ···cout·<< "This program was written on ";</pre>
                                                                                                                                                     28 ···today.print();
                                                  29 · · · return · 0;
                                                  30 }
                                                                                                                                                     30 ···cout·<<·"This·program·was·modified·on·";</pre>
                                                                                                                                                         ···today.set(18,1,2017);
                                                  32 Date::Date(int·d, ·int·m, ·int·y)
                                                                                                                                                    32 ···today.print();
                                                                                                                                                    33
                                                  34 ··if(d>0·&&·d<31)·dav·=·d;
                                                                                                                                                     34 ···return-0;
                                                  35 ··if(m>0·&&·m<13)·month·=·m;
                                                                                                                                                    35 }
                                                  36 ..if(y>0).year.=y;
                                                  37 }
                                                                                                                                                     37 Date::Date(int·d, int·m, int·y)
                                                  38
                                                                                                                                                     38
                                                  39 M - void · Date::print()
                                                                                                                                                     39 ··if(d>0·&&·d<31)·day·=·d;
                                                  41 ··cout·<<·day·<<·"-"·<<·month·<<·"-"·<<·year·<<·endl;
                                                                                                                                                     40 ··if(m>0.&&.m<13).month.=.m;
                                                                                                                                                     41 · · if(y>0) · year · = y;
                                                  42 }
                                                                                                                                                     42
                                                  44 void Date::set(int d, int m, int y)
                                                                                                                                                     44 void-Date::set(int-d,-int-m,-int-y)
                                                  46 ··if(d>0.&&.d<31).day.=.d;
                                                                                                                                                     46 ··if(d>0.&&.d<31).day.=.d;
                                                  47 · · if(m>0·&&·m<13)·month·=·m;
                                                                                                                                                     47 ··if(m>0·&&·m<13)·month·=·m;
                                                  48 ··if(y>0)·year·=y;
                                                                                                                                                     48 ··if(y>0)·year·=y;
                                                  49 }
                                                                                                                                                    49
                                                  50
                                                                                                                                                    50
                                                  51
                                                                                                                                                                                                   Visual diff
                                                                                                                                                                                                                    Run text diff
                                                                                                                                                                                                                                      Restart diff
```

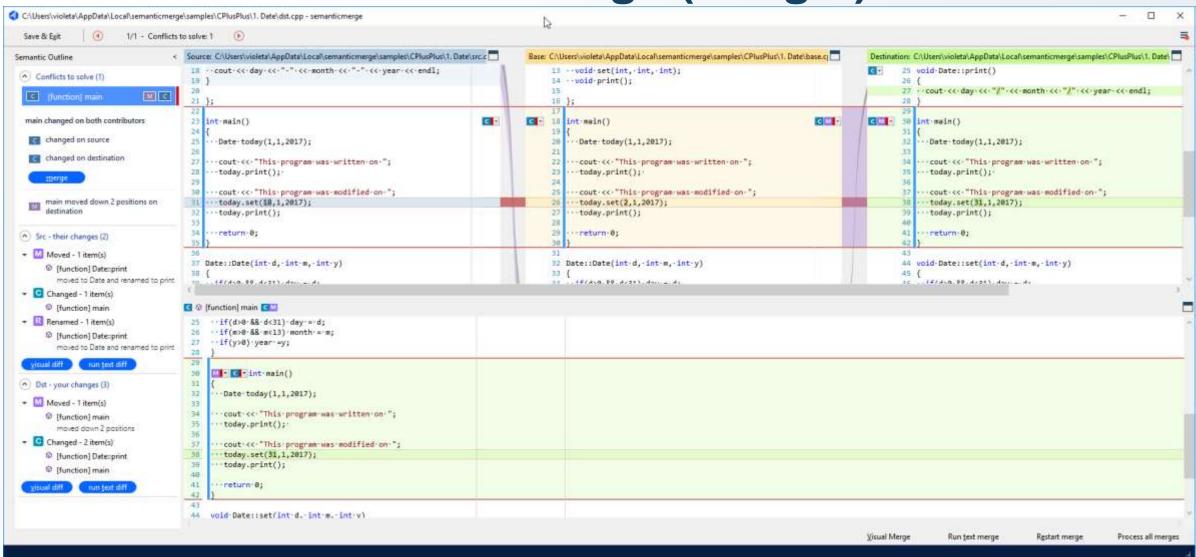


SemanticMerge (dstdiffs)



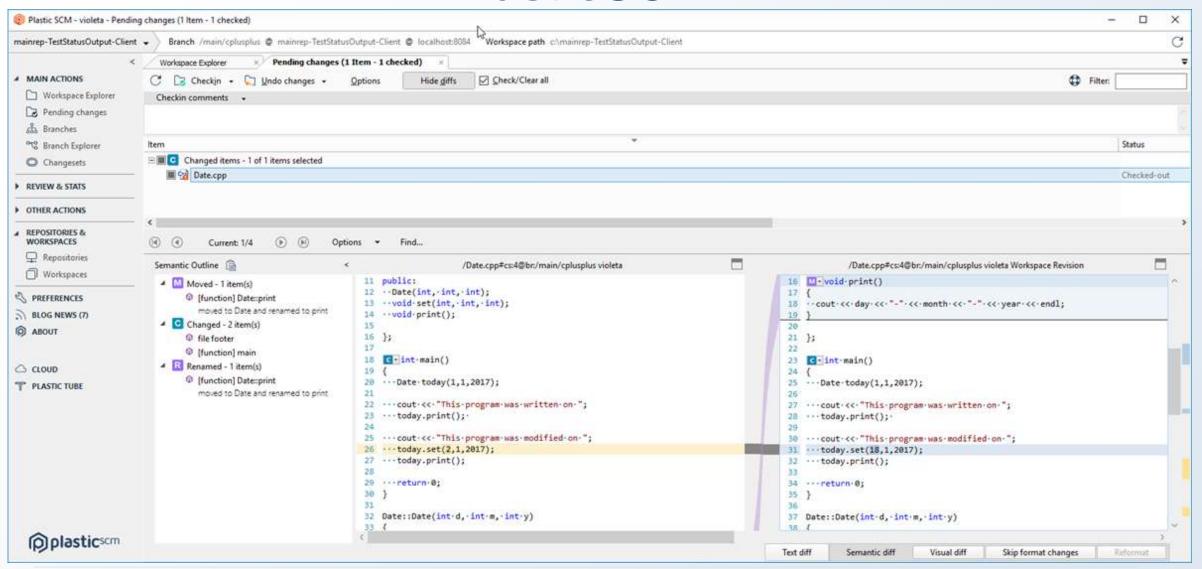


SemanticMerge (merge!)



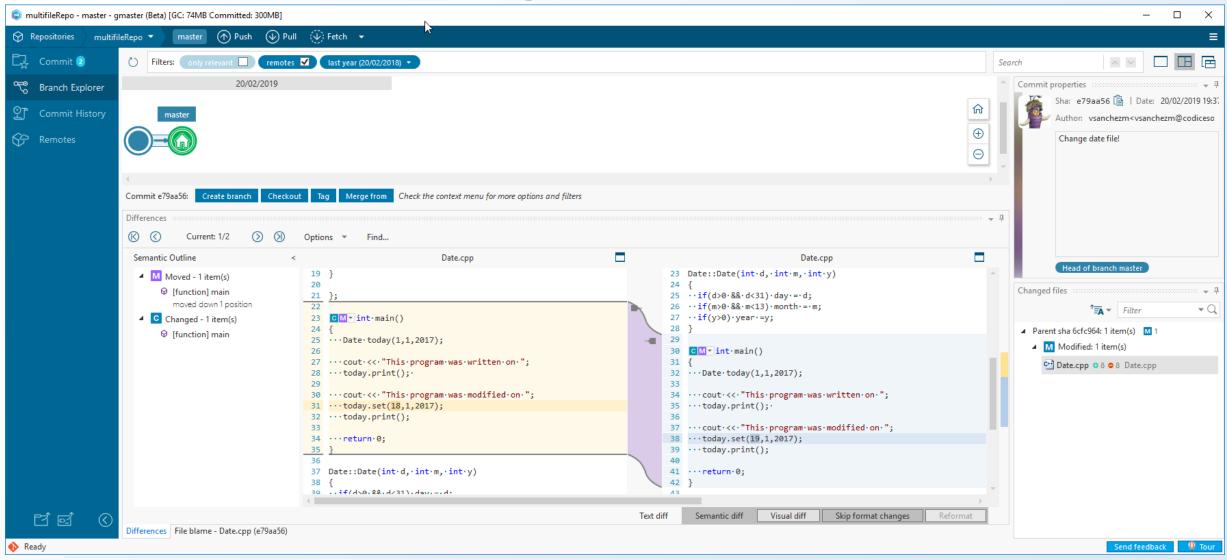


PlasticSCM



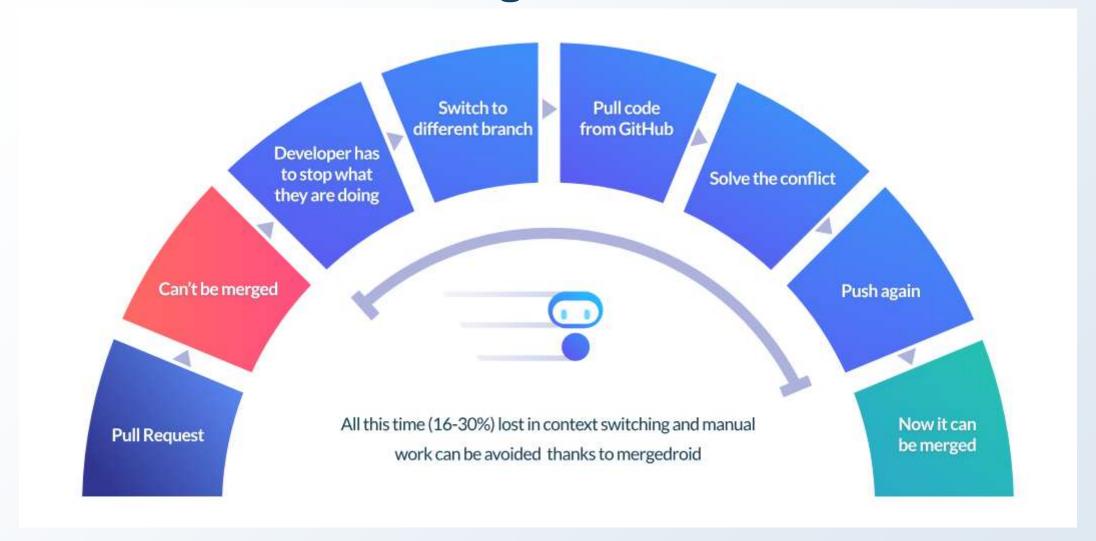


gmaster



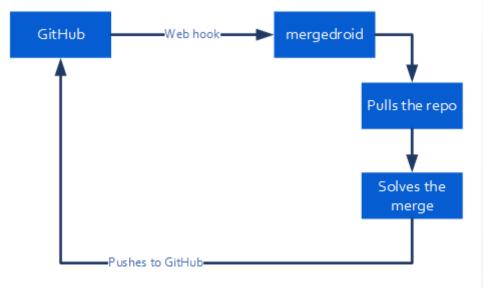


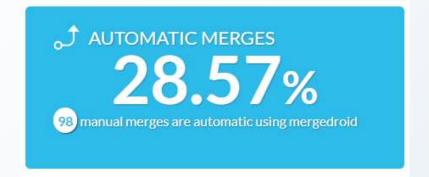
mergedroid



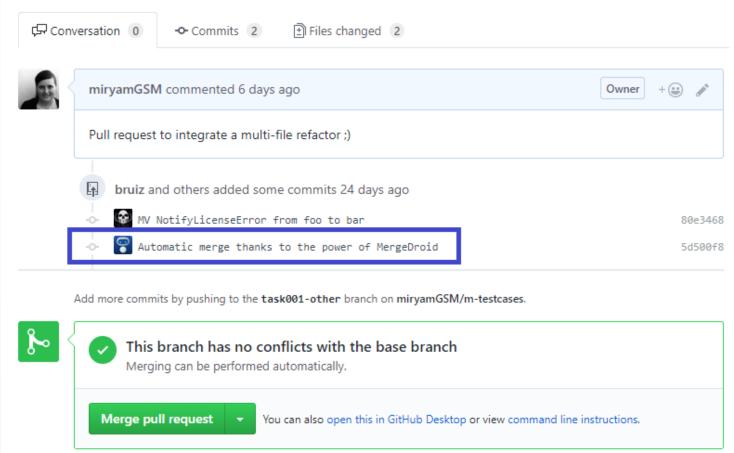


mergedroid





https://github.com/gmasterscm/gitextensions





Comentarios / Otras experiencias

