



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería
Informática**

**Comparador de métricas de
evolución en repositorios
software**



Presentado por Miguel Ángel León Bardavío
en Universidad de Burgos — 27 de julio
de 2019

Tutor: Carlos López Nozal



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



D. profesor del departamento de Ingeniería Civil, Área de Lenguajes y Sistemas Informáticos.

Expone:

Que el alumno D. Miguel Ángel León Bardavío, con DNI 71362165L, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado *Comparador de métricas de evolución en repositorios software*.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 27 de julio de 2019

Vº. Bº. del Tutor:

D. Carlos López Nozal

Resumen

El proceso de un proyecto software es un conjunto de actividades o fases por las que pasa el producto software que se está desarrollando como, por ejemplo: la especificación, el diseño, la implementación, pruebas, el aseguramiento de calidad, la configuración del proyecto, etc.

Los repositorios de código, además de almacenar el código fuente de un proyecto software, pueden incluir sistemas que faciliten las actividades del proceso de software: sistemas de control de versiones, sistemas de seguimiento de incidencias, sistemas de revisión de código, sistemas de despliegue de ejecutables, etc. En la última década han surgido forjas de repositorios útiles tanto para proyectos empresariales como para proyectos open source (SourceForge ¹, Github ², GitLab ³, Bitbucket ⁴).

Las métricas de evolución ayudan a cuantificar características de los procesos software. Un ejemplo de este tipo de medidas es el *número de días de cierre*, en la que se mide el número de días que pasan desde que se abre una incidencia hasta su cierre. Estas métricas se pueden obtener gracias a los datos estadísticos que proporcionan los repositorios.

En este TFG se diseña un software para calcular métricas de evolución sobre distintos repositorios. En el diseño se ha optado por implementar una aplicación web en Java que toma como entrada un conjunto de repositorios públicos o privados de GitLab y calcula métricas de evolución que permiten comparar los proyectos. Además, se ha procurado un diseño extensible a otras forjas de repositorios y a nuevas métricas. La aplicación ha sido probada con Trabajos Fin de Grado presentados en la Universidad de Burgos y que han sido almacenados en repositorios públicos de GitLab.

Descriptores

Métricas de evolución, repositorios de código, proceso de desarrollo de software, ciclo de vida de desarrollo de software, gestión de calidad, forjas de repositorios, comparación de proyectos software, aplicación web.

¹<https://sourceforge.net/>

²<https://github.com/>

³<https://about.gitlab.com/>

⁴<https://bitbucket.org/>

Abstract

The software development process is a set of activities or phases through which the software product that is being developed passes, such as: specification, design, implementation, testing, quality assurance, project management, etc.

The source code repositories, in addition to storing the source code of a software project, may include systems that ease the activities of the software development process: version control systems, issue tracking systems, code review systems, deployment systems, etc. Forges of source code repositories have emerged in the last decade, both for business projects and for open source projects (SourceForge ⁵, Github ⁶, GitLab ⁷, Bitbucket ⁸).

Evolution metrics helps to quantify features of a software development process. An example of this type of measure is the *days to close an issue*, in which the number of days that pass from when an incident is opened until its closure is measured. These metrics can be obtained from the statistics provided by the source code repositories.

In this project, a software is designed to calculate evolution metrics on different source code repositories. The design has chosen to implement a web application in Java language that takes as input a set of GitLab public or private repositories and calculates evolution metrics that allow the repositories to be compared. In addition, an extensible design to other repositories forges and new metrics has been sought. The application has been tested with Final Degree Projects presented at the University of Burgos and that have been stored in public repositories of GitLab.

Keywords

Evolution metrics, source code repositories, software development process, software development life cycle, quality management, forge of repositories, comparison of software projects, web application.

⁵<https://sourceforge.net/>

⁶<https://github.com/>

⁷<https://about.gitlab.com/>

⁸<https://bitbucket.org/>

Índice general

| | |
|---|------------|
| Índice general | III |
| Índice de figuras | V |
| Índice de tablas | VI |
| Introducción | 1 |
| 1.1. Estructura de la memoria | 3 |
| Objetivos del proyecto | 5 |
| 2.1. Objetivos generales | 5 |
| 2.2. Objetivos técnicos | 6 |
| Conceptos teóricos | 7 |
| 3.1. Calidad de proceso y producto software | 7 |
| 3.2. Evolución de software | 9 |
| 3.3. Framework de medición | 12 |
| 3.4. Forjas de proyectos software | 13 |
| Técnicas y herramientas | 15 |
| 4.1. Herramientas utilizadas | 15 |
| Aspectos relevantes del desarrollo del proyecto | 19 |
| 5.1. Motivación de la elección | 19 |
| 5.2. Evolución del proyecto | 19 |
| 5.3. Documentación | 20 |
| 5.4. Configuración del flujo de trabajo del desarrollo software . . | 21 |
| 5.5. Automatización de tareas de desarrollo | 22 |

| | |
|--|-----------|
| Trabajos relacionados | 25 |
| 6.1. Activiti-API | 25 |
| 6.2. Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones | 25 |
| 6.3. Software Project Assessment in the Course of Evolution - Jacek Ratzinger | 25 |
| Conclusiones y Líneas de trabajo futuras | 27 |
| 7.1. Conclusiones | 27 |
| 7.2. Líneas de trabajo futuras | 27 |
| Bibliografía | 29 |

Índice de figuras

| | |
|---|----|
| 1.1. Comparativa de tendencia de búsqueda de Google desde 2004 con los términos de distintas forjas de proyectos software. | 3 |
| 3.2. Principales factores de calidad del producto de software[11] . . . | 8 |
| 3.3. Métricas de control y métricas de predicción[11] | 8 |
| 3.4. Diagrama del framework para el cálculo de métricas con perfiles que almacena valores umbrales. | 13 |
| 5.5. Milestone iniciales del proyecto en GitLab. | 20 |

Índice de tablas

Introducción

El desarrollo de software es un proceso muy complejo que depende de múltiples factores: equipo de desarrollo, tipo de producto software, estabilidad de los requisitos funcionales, importancia de los requisitos no funcionales como escalabilidad, seguridad, licencias, lenguaje de programación, tipo de arquitectura de computación...

Existen varias metodologías que ayudan a definir las actividades y artefactos del proceso de desarrollo de software. Los artefactos son las salidas de las actividades y el conjunto de artefactos conforman el producto software. En el caso de Unified Process (UP) [5] se identifican las siguientes actividades o flujos de trabajo: recolección de requisitos, diseño e implementación, pruebas, despliegue. Además en UP se añaden tres flujos de trabajo de soporte: configuración de cambios, gestión de proyecto y entorno. Estos flujos de trabajo se aplican iterativamente durante varias fases del desarrollo en cada una de las cuales se incrementa el producto software con algún artefacto resultado de la actividad. La característica de iteración e incremental es recogida en otros métodos o buenas prácticas de desarrollo ágiles[10]: Scrum, eXtreme Programming, Lean...

Una suposición subyacente de la administración de la calidad es que la calidad del proceso de desarrollo afecta directamente a la calidad de los productos a entregar [11, pág 543]. La calidad del proceso es uno de los factores que determinan la calidad del producto software, tal y como expone Sommerville en *Ingeniería de software*[11].

Los repositorios de código son espacios virtuales donde los equipos de desarrollo generan los artefactos colaborativos procedentes de las actividades de un proceso de desarrollo. En los repositorios además de guardar los artefactos, versión final y versiones previas, se almacena la interacción de

los miembros del equipo justificando el cambio de version. Dependiendo del artefacto generado se utiliza distintos sistemas: foros de comunicación, sistemas de control de versiones, sistemas de documentación, sistemas de gestión de issues, gestión de pruebas, de revisiones de calidad, de integración y despliegue continuo [3].

En la última década han surgido forjas de proyectos software de fácil acceso tanto para proyectos empresariales como para proyectos open source (SourceForge ⁹, Github ¹⁰, GitLab ¹¹, Bitbucket ¹²). Estas forjas suelen integrar múltiples sistemas para dar soporte a los flujos de trabajo y registrar las interacciones entre los miembros del equipo. Además dan la posibilidad de extensión funcional con sistemas de terceros para gestionar otras actividades no soportadas directamente, por ejemplo Travis CI ¹³ para gestionar la integración continua o Codacy ¹⁴ para gestionar las revisiones automáticas de calidad. Actualmente estas forjas han tenido una gran aceptación entre la comunidad de desarrolladores y existen muchos desarrollos de software de tendencia que las utilizan. La forjas de proyectos software proporcionan interfaces de programación específicas que permiten acceder a toda la información registrada. En la Fig. 1.1 se aprecia como cambia la tendencia de utilización de dichas forjas en el tiempo. Actualmente la forja predominante es claramente GitHub pero se ve un incremento en el uso de Gitlab.

Parece lógico considerar como hipótesis que la calidad de un artefacto software tenga alguna relación con la manera en la que el equipo aplica las actividades del proceso de desarrollo dentro del repositorio. La validación empírica de estas hipótesis ha abierto una nueva línea de aplicación con los conjuntos de datos extraídos de los repositorios. Estas forjas de proyectos software están en constante evolución, tanto en sus estructuras estáticas como en sus interacciones dinámicas en los proyectos. Se registran grandes conjuntos de datos difíciles de procesar. El desafío a la comunidad científica y empresarial es constante mostrando un incremento en el interés en las aplicaciones que mejoren sus sistemas de decisión.

Este trabajo presenta el diseño y desarrollo de una aplicación Java Web que: calcula un conjunto de métricas de evolución del proceso, de proyectos software alojados en GitLab, para posteriormente, compararlos en relación con otros proyectos.

⁹<https://sourceforge.net/>

¹⁰<https://github.com/>

¹¹<https://about.gitlab.com/>

¹²<https://bitbucket.org/>

¹³<https://travis-ci.org/>

¹⁴<https://www.codacy.com/>



Figura 1.1: Comparativa de tendencia de búsqueda de Google desde 2004 con los términos de distintas forjas de proyectos software.

1.1. Estructura de la memoria

La memoria se estructura de la siguiente manera¹⁵[12]:

Introducción. Introducción al trabajo realizado, estructura de la memoria y listado de materiales adjuntos.

Objetivos del proyecto. Objetivos que se persiguen alcanzar con la realización del proyecto.

Conceptos teóricos. Conceptos clave para comprender los objetivos, el proceso y el producto del proyecto.

Técnicas y herramientas. Técnicas y herramientas utilizadas durante el desarrollo del proyecto.

Aspectos relevantes del desarrollo. Aspectos destacables durante el proceso de desarrollo del proyecto.

Trabajos relacionados. Otros proyectos de la misma naturaleza y los cuales han ayudado a la realización de este.

Conclusiones y líneas de trabajo futuras. Conclusiones tras la realización del proyecto y posibilidades de mejora o expansión.

¹⁵<https://github.com/ubutfgm/plantillaLatex>

Se incluyen también los siguientes anexos:

Plan del proyecto software. Planificación temporal y estudio de la viabilidad del proyecto.

Especificación de requisitos del software. Análisis de los requisitos.

Especificación de diseño. Diseño de los datos, diseño procedimental y diseño arquitectónico.

Manual del programador. Aspectos relevantes del código fuente.

Manual de usuario. Manual de uso para usuarios que utilicen la aplicación.

Objetivos del proyecto

En este capítulo se detallarán los objetivos generales que se desean alcanzar en este proyecto, así como los objetivos más técnicos.

2.1. Objetivos generales

El objetivo general de este TFG es diseñar una aplicación Web en Java que permita obtener un conjunto de métricas de evolución del proceso software [8] reflejado en los repositorios de GitLab, para permitir comparar los distintos procesos de desarrollo software de cada repositorio. La aplicación se probará con datos reales para comparar los repositorios de Trabajos Fin de Grado del Grado de Ingeniería Informática presentados en GitLab.

A continuación se desglosa el objetivo general en objetivos más detallados.

- Se desea obtener medidas de métricas de evolución de uno o varios proyectos alojados en repositorios de GitLab.
- Las métricas que se desean calcular de un repositorio son algunas de las especificadas en [8] y adaptadas a los repositorios software.
 - Número total de incidencias (*issues*)
 - Cambios(*commits*) por incidencia
 - Porcentaje de incidencias cerrados
 - Media de días en cerrar una incidencia
 - Media de días entre cambios
 - Días entre primer y último cambio
 - Rango de actividad de cambios por mes
 - Porcentaje de pico de cambios

- El objetivo de obtener las métricas es poder evaluar el estado de un proyecto comparándolo con otros proyectos de la misma naturaleza. Para ello se deberán establecer unos valores umbrales de cada métrica basados en el cálculo de los cuartiles Q1 y Q3, respectivamente.
- Se desea poder calcular dinámicamente los valores umbrales en perfiles de configuración de medidas. Esta información sobre los valores umbrales se puede almacenar de manera persistente para permitir comparaciones futuras. Un ejemplo de utilidad es guardar los valores umbrales de repositorios por lenguaje de programación, o en el caso de repositorios de TFG de la UBU por cursos académico.

2.2. Objetivos técnicos

Este apartado recoge los requisitos más técnicos del proyecto.

- Diseñar la aplicación de manera que se puedan extender con nuevas métricas con el menor coste de mantenimiento posible. Se aplicará un diseño basado en frameworks y en patrones de diseño [2].
- El diseño de la aplicación debe facilitar la extensión a otras plataformas de desarrollo colaborativo como GitHub o Bitbucket. Se aplicará un diseño basado en frameworks y en patrones de diseño [2].
- Aplicar el *frameworks* modelo vista controlador para separar la lógica de la aplicación para el cálculo de métricas y la interfaz de usuario. Se aplicará un diseño basado en frameworks y en patrones de diseño [2].
- Crear una batería de pruebas automáticas con cobertura del por encima del 90 % en los subsistemas de lógica de la aplicación .
- Utilizar una plataforma de desarrollo colaborativo que incluya un sistema de control de versiones, un sistema de seguimiento de incidencias y que permita una comunicación fluida entre el tutor y el alumno.
- Utilizar un sistema de integración y despliegue continuo.
- Diseñar una correcta gestión de errores definiendo excepciones de biblioteca y registrando eventos de error e información en ficheros de *log*.
- Aplicar nuevas estructuras del lenguaje Java para el desarrollo, como son expresiones lambda.
- Utilizar sistemas que aseguren la calidad continua del código que permitan evaluar la deuda técnica del proyecto.
- Probar la aplicación con ejemplos reales y utilizando usos avanzados de bibliotecas, como la conjuntos de datos entrada de test en ficheros con formato tabulado tipo csv (*comma separated values*).

Conceptos teóricos

En este capítulo se explican conceptos relevantes para la comprensión de este proyecto y su contexto.

3.1. Calidad de proceso y producto software

Todo proyecto de software tienen como objetivo desarrollar software de alta calidad, que cumpla con los requisitos acordados. Sin embargo, la calidad de un producto de software no tiene que ver solo con que se cumplan todos los requisitos funcionales, sino también otros requerimientos no funcionales que no se incluyen en la especificación como los de mantenimiento.

Sommerville enumera en *Software Engineering* [11] los principales factores que afectan a la calidad del producto, como se puede observar en la figura 3.2:

- Calidad del proceso
- Tecnología de desarrollo
- Calidad del personal
- Costo, tiempo y duración

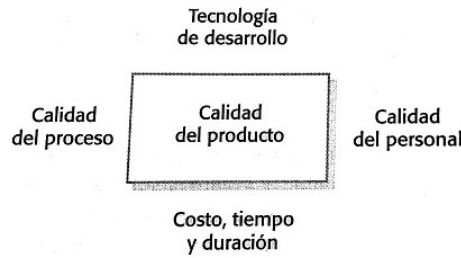


Figura 3.2: Principales factores de calidad del producto de software[11]

Para llegar a tener un software de calidad hay que tener en cuenta todos los factores mencionados anteriormente en cada una de las tres fases de la administración de la calidad: Aseguramiento, planeación y control.

La fase de control es la que se encarga de que el equipo de desarrollo cumpla los estándares y procedimientos definidos en el plan de calidad del proyecto. Esta fase puede realizarse mediante un proceso de medición.

La medición del software es un proceso en el cual se asignan valores numéricos o simbólicos a atributos de un producto o proceso software. Una métrica es una medida relacionada con un atributo de un proceso o producto software (por ejemplo). Las métricas son de control o de predicción. Las **métricas de control** se asocian al proceso de desarrollo del software (p. ej. Media de días ue se tarda en cerrar una incidencia)y las **métricas de predicción** se asocian a productos software (p. ej. Complejidad ciclomática de una función). Y ambos tipos de métricas influyen en la toma de decisiones administrativas como se observa en la figura 3.3.

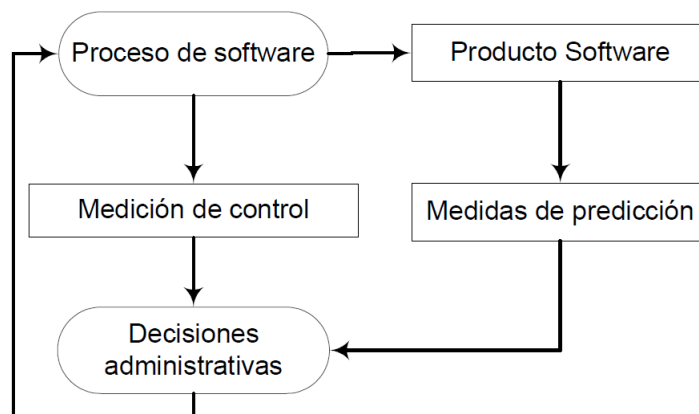


Figura 3.3: Métricas de control y métricas de predicción[11]

3.2. Evolución de software

Este proyecto se centra en las métricas de control, es decir las que están asociadas al proceso de desarrollo o de evolución del software.

Conceptos de evolución

Una plataforma de desarrollo colaborativo como GitLab puede presentar herramientas para controlar la evolución de un proyecto software. Por ejemplo un sistema de control de versiones (VCS - *Version Control System*) como Git, un sistema de seguimiento de incidencias (*issue tracking system*), un sistema de integración continua (CI - *Continuous Integration*), un sistema de despliegue continuo (CD - *Continuous deployment*), etc. Todas estas herramientas sirven para llevar un control sobre la evolución de un proyecto software ya que facilitan la comunicación entre los miembros de un equipo de desarrollo, ayudan a gestionar los cambios que producen cada uno de los miembros y proporcionan mediciones de proceso. Estas mediciones se pueden utilizar para obtener métricas de proceso que ayuden a mejorar la evolución del proyecto comparandolas con otras mediciones de otros proyectos.

Métricas de evolución o de proceso

Como se ha explicado previamente, este proyecto calcula métricas de proceso. Las métricas que se utilizan en este proyecto provienen una Master Tesis titulada *sPACE: Software Project Assessment in the Course of Evolution* [8]. A continuación se describen las métricas que se implementan en este proyecto usando la plantilla de definicion de la norma ISO 9126 [4].

I1 - Número total de issues(incidencias)

- **Categoría:** Proceso de Orientación
- **Descripción:** Número total de issues creadas en el repositorio
- **Propósito:** ¿Cuántas issues se han definido en el repositorio?
- **Fórmula:** NTI . $NTI = \text{número total de issues}$
- **Fuente de medición:** Proyecto en una plataforma de desarrollo colaborativo.
- **Interpretación:** $NTI \geq 0$. Valores bajos indican que no se utiliza un sistema de seguimiento de incidencias, podría ser porque el proyecto acaba de comenzar
- **Tipo de escala:** Absoluta

- **Tipo de medida:** $NTI = Contador$

I2 - Commits(cambios) por issue

- **Categoría:** Proceso de Orientación
- **Descripción:** Número de commits por issue
- **Propósito:** ¿Cuál es el volumen medio de trabajo de las issues?
- **Fórmula:** $CI = \frac{NTC}{NTI}$. $CI = Cambios por issue$, $NTC = Número total de commits$, $NTI = Numero total de issues$
- **Fuente de medición:** Proyecto en una plataforma de desarrollo colaborativo.
- **Interpretación:** $CI \geq 0$, Si se acerca a 1 se definen bien las issues, si alto: no se definen bien las issues, si bajo: desarrollo del proyecto lento
- **Tipo de escala:** Ratio
- **Tipo de medida:** $NTC, NTI = Contador$

I3 - Porcentaje de issues cerradas

- **Categoría:** Proceso de Orientación
- **Descripción:** Porcentaje de issues cerradas
- **Propósito:** ¿Qué porcentaje de issues definidas en el repositorio se han cerrado?
- **Fórmula:** $PIC = \frac{NTIC}{NTI} * 100$. $PIC = Porcentaje de issues cerradas$, $NTIC = Número total de issues cerradas$, $NTI = Numero total de issues$
- **Fuente de medición:** Proyecto en una plataforma de desarrollo colaborativo.
- **Interpretación:** $0 \leq PIC \leq 100$. Cuanto más alto mejor
- **Tipo de escala:** Ratio
- **Tipo de medida:** $NTI, NTIC = Contador$

TI1 - Media de días en cerrar una issue

- **Categoría:** Constantes de tiempo
- **Descripción:** Media de días en cerrar una issue
- **Propósito:** ¿Cuánto se suele tardar en cerrar una issue?
- **Fórmula:** $MDCI = \frac{\sum_{i=0}^{NTIC} DCI_i}{NTIC}$. $MDCI = Media de días en cerrar una issue$, $NTIC = Número total de issues cerradas$, $DCI = Días en cerrar la issue$

- **Fuente de medición:** Proyecto en una plataforma de desarrollo colaborativo.
- **Interpretación:** $MDCI \geq 0$. Cuanto más pequeño mejor.
- **Tipo de escala:** Ratio
- **Tipo de medida:** NTI , $NTIC = Contador$

TC1 - Media de días entre commits

- **Categoría:** Constantes de tiempo
- **Descripción:** Media de días que pasan entre dos commits consecutivos
- **Propósito:** ¿Cuántos días suelen pasar desde un commit hasta el siguiente?
- **Fórmula:** $MDC = \frac{\sum_{i=1}^{NTC} TC_i - TC_{i-1}}{NTC}$. $TC_i - TC_{i-1}$ en días; $MDC =$ Media de días entre cambios, $NTC =$ Número total de commits, $TC =$ Tiempo de Commit
- **Fuente de medición:** Proyecto en una plataforma de desarrollo colaborativo.
- **Interpretación:** $MDEC \geq 0$. Cuanto más pequeño mejor.
- **Tipo de escala:** Ratio
- **Tipo de medida:** $NTC = Contador$; $TC = Tiempo$

TC2 - Días entre primer y último commit

- **Categoría:** Constantes de tiempo
- **Descripción:** Días transcurridos entre el primer y el ultimo commit
- **Propósito:** ¿Cuántos días han pasado entre el primer y el último commit?
- **Fórmula:** $DEPUC = TC2 - TC1$. $TC2 - TC1$ en días; $DEPUC =$ Días entre primer y último commit, $TC2 =$ Tiempo de último commit, $TC1 =$ Tiempo de primer commit
- **Fuente de medición:** Proyecto en una plataforma de desarrollo colaborativo.
- **Interpretación:** $DEPUC \geq 0$. Mejor valores altos
- **Tipo de escala:** Absoluta
- **Tipo de medida:** $TC = Tiempo$

TC3 - Ratio de actividad de commits por mes

- **Categoría:** Constantes de tiempo

- **Descripción:** Muestra el número de commits relativos al número de meses
- **Propósito:** ¿Cuál es el número medio de cambios por mes?
- **Fórmula:** $RCM = \frac{NTC}{12}$. $RCM = \text{Ratio de cambios por mes}$, $NTC = \text{Número total de commits}$
- **Fuente de medición:** Proyecto en una plataforma de desarrollo colaborativo.
- **Interpretación:** $RCM > 0$. Cuanto más alto mejor
- **Tipo de escala:** Ratio
- **Tipo de medida:** $NTC = \text{Contador}$

C1 - Cambios pico

- **Categoría:** Constantes de tiempo
- **Descripción:** Número de commits en el mes que más commits se han realizado en relación con el número total de commits
- **Propósito:** ¿Cuál es la proporción de trabajo realizado en el mes con mayor número de cambios?
- **Fórmula:** $CP = \frac{NCMP}{NTC}$. $CP = \text{Cambios pico}$, $NCMP = \text{Número de commits en el mes pico}$, $NTC = \text{Número total de commits}$
- **Fuente de medición:** Proyecto en una plataforma de desarrollo colaborativo.
- **Interpretación:** $0 \leq CCP \leq 1$. Mejor valores intermedios
- **Tipo de escala:** Ratio
- **Tipo de medida:** $NCMP, NTC = \text{Contador}$

3.3. Framework de medición

Para la implementación de las métricas se ha seguido la solución basada en frameworks propuesta en *Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones* [6]. El framework cuyo diseño se muestra en la Figura 3.4 es independiente del repositorio concreto (*GitLab*, *GitHub*, *Bitbucket*, *SourceForge*) y su objetivo es la reutilización en la implementación del cálculo de métricas.

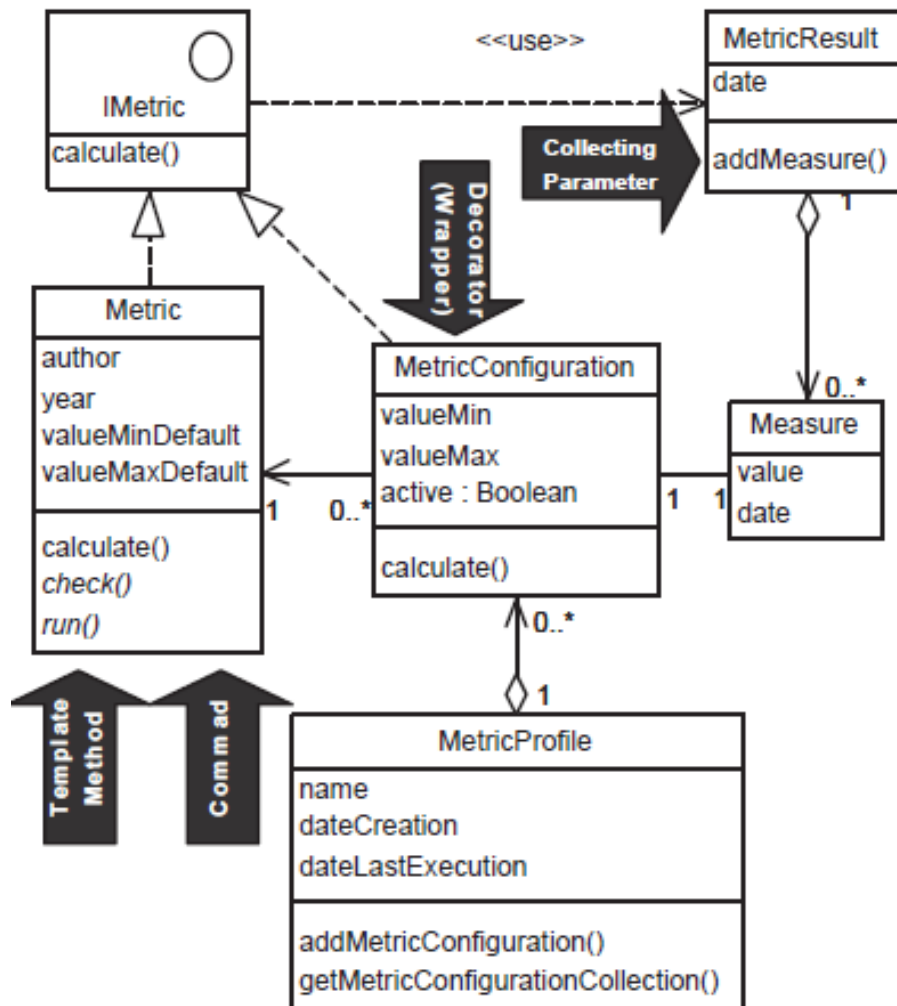


Figura 3.4: Diagrama del framework para el cálculo de métricas con perfiles que almacena valores umbrales.

En la Figura 3.4 se muestra el diagrama de clases UML que identifica los conceptos principales del framework de medición y la relación entre ellos, especificando la navegabilidad y la multiplicidad. Las anotaciones en forma de flecha especifican los patrones de diseño aplicados en el framework [2].

3.4. Forjas de proyectos software

Comentario de revisión...Este párrafo está repetido en la introducción.

Pero tiene que ser reducido en este apartado se puede detallar más extensamente con alguna captura de pantalla que ayude a describir las equivalentes en múltiples repositorios. **Comentario de revisión...**revisar

En la última década han surgido forjas de proyectos software de fácil acceso tanto para proyectos empresariales como para proyectos open source (SourceForge ¹⁶, Github ¹⁷, GitLab ¹⁸, Bitbucket ¹⁹). Estas forjas suelen integrar múltiples sistemas para dar soporte a los flujos de trabajo y registrar las interacciones entre los miembros del equipo. Además dan la posibilidad de extensión funcional con sistemas de terceros para gestionar otras actividades no soportadas directamente, por ejemplo Travis CI ²⁰ para gestionar la integración continua o Codacy ²¹ para gestionar las revisiones automáticas de calidad. Actualmente estas forjas han tenido una gran aceptación entre la comunidad de desarrolladores y existen muchos desarrollos de software de tendencia que las utilizan. La forjas de proyectos software proporcionan interfaces de programación que permiten acceder a toda la información registrada.

Estas forjas de proyectos software están en constante evolución, tanto en sus estructuras estáticas como en sus interacciones dinámicas en los proyectos. Se registran grandes conjuntos de datos difíciles de procesar. El desafío a la comunidad científica y empresarial es constante mostrando un incremento en el interés en las aplicaciones de minería que mejoren sus sistemas de decisión [3]:

Github vs. GitLab

Comentario de revisión...introducir la sección **Comentario de revisión...** Usa la información de esta página <https://about.gitlab.com/devops-tools/github-vs-gitlab.html> para recoger las características de los repositorios y su comparación.

¹⁶<https://sourceforge.net/>

¹⁷<https://github.com/>

¹⁸<https://about.gitlab.com/>

¹⁹<https://bitbucket.org/>

²⁰<https://travis-ci.org/>

²¹<https://www.codacy.com/>

Técnicas y herramientas

Este capítulo muestra las herramientas que se han utilizado para la alcanzar los objetivos del proyecto.

4.1. Herramientas utilizadas

Entorno de desarrollo

Comentario de revisión...En todas las herramientas describir brevemente las funcionalidades de que se utilizan concretamente en el proyecto. **Comentario de revisión...**Se pueden acompañar con alguna captura de pantalla, porción de código y a una referencia al manual del programador para obtener más detalle. **Comentario de revisión...**Por ejemplo Java SE v11.0.1 Expresiones Lambda en el uso de stream. **Comentario de revisión...**Apache Maven indicar brevemente el flujo de trabajo utilizado de compilación, testing, calidad, package y despliegue

Eclipse IDE for Java EE Developers . Entorno de programación Java para aplicaciones web. Se ha utilizado la version: 2018-09 (4.9.0).

Enlace a página de descarga:

<https://www.eclipse.org/downloads/>

Java SE 11 (JDK) . *Java Development Kit*. Se ha utilizado la versión v11.0.1.

Enlace a página de descarga:

<https://www.oracle.com/technetwork/java/javase/downloads/index.html>

Apache Maven . Gestor de proyectos software que ayuda en la construcción del proyecto, la generación de documentación, generación de informes, gestión de dependencias, integración con un sistema de control de versiones, etc. Se ha utilizado la versión v3.6.0.

Enlace a página de descarga:

<https://maven.apache.org/download.cgi>

Apache Tomcat . Contenedor de aplicaciones web con soporte de servlets Java. Sirve para desplegar la aplicación. Se ha utilizado la versión v9.0.13.

Enlace a página de descarga:

<https://tomcat.apache.org/download-90.cgi>

Logging

SLF4J . Fachada de logging.

Enlace a página de descarga:

<https://www.slf4j.org/download.html>

Log4j 2 . Logger. Se ha utilizado la versión v2.11.2.

Enlace a página de descarga:

<https://logging.apache.org/log4j/2.x/download.html>

Pruebas

JUnit5 . Conjunto de bibliotecas para el desarrollo de pruebas unitarias. Se ha utilizado la versión v5.3.1.

Enlace a página de descarga:

<https://junit.org/junit5/>

Frameworks y librerías específicas para el proyecto

Comentario de revisión... Además de los enlaces de la aplicación añade los enlaces de tu repositorio

gitlab4j-api . Framework de conexión a GitLab API. Se ha utilizado la versión v4.9.14.

Enlace:

<https://github.com/gitlab4j/gitlab4j-api>

Se ha preferido frente a `timols/java-gitlab-api`²² tras realizar un estudio sobre sus métricas de evolución y una comparativa sobre la documentación. Concluyendo en que **gitlab4j-api** contiene mejor documentación, mejor evolución y a día de hoy se sigue desarrollando.

Apache Commons Math . Librería que se utiliza para matemáticas descriptivas utilizada para el cálculo de cuartiles. Se ha utilizado la versión v3.6.1.

Enlace a página de descarga:

https://commons.apache.org/proper/commons-math/download_math.cgi

Interfaz gráfica

Vaadin . Framework para desarrollo de interfaces Web con Java. Se ha utilizado la versión v13.0.0

Enlace:

<https://vaadin.com/>

CI/CD y Calidad de código

GitLab . Plataforma de desarrollo colaborativo en la que se ha almacenado el proyecto en un repositorio Git.

Enlace:

https://gitlab.com/users/sign_in

Codacy . Herramienta de generación automática de informes de calidad de código.

Enlace:

<https://www.codacy.com/>

JaCoCo . Librería para cobertura de código en Java. Se ha utilizado la versión v0.8.3.

Enlace:

<https://www.eclemma.org/jacoco/>

Heroku . Herramienta para despliegue continuo.

Enlace:

<https://id.heroku.com/login>

²²<https://github.com/timols/java-gitlab-api>

Documentación

LaTeX . Sistema de composición de textos.

Enlace:

<https://www.latex-project.org/>

TeXstudio . Entorno de desarrollo de documentos LaTeX.

Enlace:

<https://www.texstudio.org/>

Zotero . Herramienta de gestión de fuentes bibliográficas.

Enlace:

<https://www.zotero.org/>

Técnicas

A lo largo del proyecto se han utilizado numerosos patrones de diseño [2] como Singleton, Factory Method, Wrapper, Builder, Listener, etc.

Para el motor de métricas se ha utilizado como base el framework propuesto en *Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones* [6]. Ver figura 3.4.

Aspectos relevantes del desarrollo del proyecto

Este capítulo recoge los aspectos destacables durante el desarrollo del proyecto.

5.1. Motivación de la elección

La elección de este trabajo fue motivada por su relación con las asignaturas de la mención de ingeniería del software, destacando las asignaturas: *Gestión de Proyectos*, *Diseño y Mantenimiento del Software*, *Desarrollo Avanzado de Sistemas Software*, en las que se enseña como desarrollar software de calidad usando repositorio Software.

5.2. Evolución del proyecto

Tras la elección del proyecto se acordó definir una evolución que siga las bases del modelo Scrum. Tomando un proceso de desarrollo incrementa con revisión de las iteraciones cada dos semanas. Se han definido sprints de dos semanas. Estas reuniones constaban de dos partes:

- Revisión del sprint: En las que se revisaba el incremento generado, los problemas que hubo durante su desarrollo, las soluciones que se han implementado o que se plantean para el siguiente sprint.
- Planificación del siguiente sprint: Se definían las nuevas tareas.

Las primeras fases del desarrollo fueron de investigación y configuración. Luego se planteó un diseño inicial, que fue la base para la implementación de nuevas funcionalidades aunque, con el tiempo, se fue modificando el diseño inicial para adaptarlo a las nuevas funcionalidades o para resolver ciertos problemas que aparecían. En las siguientes secciones se detalla más a fondo cada una de estas etapas de desarrollo.

En la Fig. 5.5 se muestran los primeros *sprints* del proyecto. Por cada sprint se observa su descripción, la fecha de apertura y cierre y el número de issues asociadas.

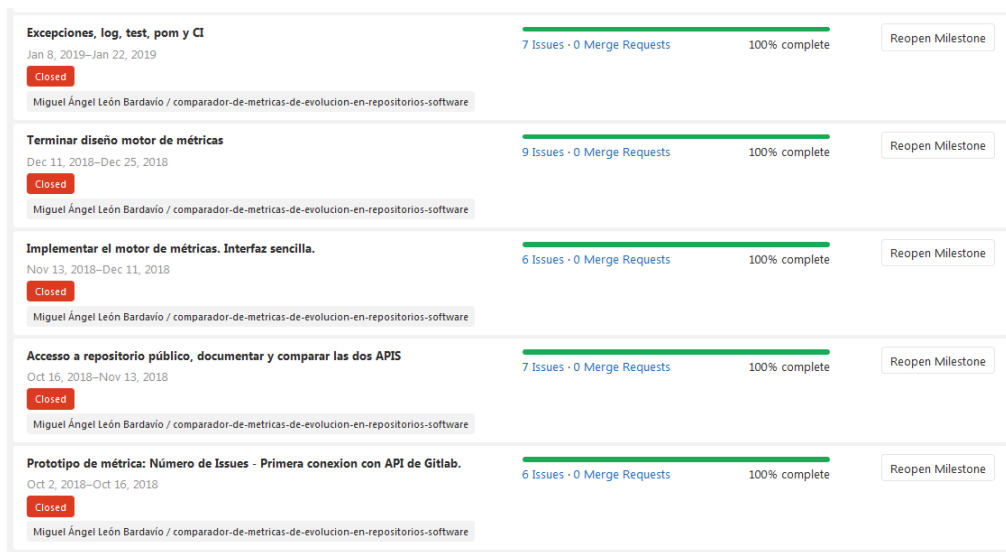


Figura 5.5: Milestone iniciales del proyecto en GitLab.

5.3. Documentación

La fase de documentación duró un mes y fue a la par con la etapa de configuración.

Se recopiló información sobre trabajos relacionados como *Activiti-Api*[9], *Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones* [6] y *sPACE: Software Project Assessment in the Course of Evolution* [8] y se estudió en qué entornos y qué herramientas se utilizarían para el desarrollo del proyecto.

Uno de los estudios más relevantes fue la elección de un API Java que permitiese la conexión a GitLab. Había tres opciones:

- Crear un framework propio de conexión a GitLab a partir de GitLab API. Añadía cierta complejidad al proyecto al tener que desarrollar otro módulo más, pero permitía poder definir las funciones que se necesitaban.
- Usar [timols/java-gitlab-api](https://github.com/timols/java-gitlab-api)²³[7]. Al principio fue la solución que se escogió, pero posteriormente se descubrió otro API bastante mejor. La documentación es bastante pobre y la evolución del proyecto software estaba parada o no evolucionaba bien, tenían demasiadas incidencias abiertas y no ofrecía gran parte de la funcionalidad que aportaba GitLab API.
- Usar [24](https://github.com/gitlab4j/gitlab4j-api)[1]. Es un proyecto bastante decente y, a día de hoy, sigue creciendo. Tiene un alto porcentaje de incidencias cerradas, un gran número de releases, y evolución constante. Este es el API con el que se ha desarrollado este proyecto.

Otro de las decisiones más difíciles fue la versión de Java. Recientemente se lanzó la versión de Java 11 y es la que inicialmente se utilizó en el proyecto. Ha habido bastantes problemas de compatibilidad, pero se han ido solucionando a lo largo del tiempo.

5.4. Configuración del flujo de trabajo del desarrollo software

Compilación y empaquetado

El proyecto se iba a desarrollar en Java desde el principio, era un requisito no funcional. La versión más moderna que había cuando se comenzó el proyecto era Java 11. La decisión de utilizar esta versión de Java generó muchos problemas de compatibilidad de versiones con otras bibliotecas en el desarrollo. **Comentario de revisión...**Enumerar cuales tenían problemas ¿Vaadin?

En el proyecto se ha utilizado alguna de las nuevas versiones de Java no vista durante los curso. En el siguiente fragmento de código se describe un ejemplo de código que obtiene una colección de repositorios del usuario actual.

```
...  
if (connectionType != EnumConnectionType.NOT_CONNECTED) {
```

²³<https://github.com/timols/java-gitlab-api>

²⁴<https://github.com/gitlab4j/gitlab4j-api>

```

return gitLabApi.getProjectApi().getUserProjectsStream(
    currentUser.getId(),
    new ProjectFilter())
    .map(p ->
        new Repository(p.getWebUrl(), p.getName(), p.getId()))
    .collect(Collectors.toList());
}
...

```

5.5. Automatización de tareas de desarrollo

Comentario de revisión... Describir el uso de Maven Ventajas e inconvenientes

Pruebas

Para las pruebas se disponían de datos de otros repositorios de GitLab que se han presentado como TFG en el Grado de Ingeniería Informática en la Universidad de Burgos. Esto ha facilitado en gran medida el trabajo de pruebas. Es destacable la funcionalidad de acceso a los repositorios por el concepto de grupo definido en Gitlab. Esto fue posible gracias a que la empresa Hewlett Packard SCDS en su colaboración con TFG con la UBU organiza su propuestas de TFG en GitLab en grupos para organizarlos por cursos académicos.

Otro aspecto destacable es la uso con funciones avanzadas en la implementación de test unitarios. Como ejemplo en el siguiente fragmento de código se recoge como parametrizar un test para especificar un fichero que recojan los múltiples casos de prueba. De esta forma se simplifica mucho las prueba.

Comentario de revisión... se puede formatear mejor el código

```

@ParameterizedTest(name = "Run with User = \"{0}\" and
    Password = \"{1}\" must throw an exception.")
@CsvFileSource(resources = "/testConnectUserPasswordWrong.csv",
    numLinesToSkip = 1, delimiter = ';', encoding = "UTF-8")

public void testConnectUserPasswordWrong(String user, String password){
    assertThrows(RepositoryDataSourceException.class, () -> {
        repositoryDataSource.connect(user, password);},
        getErrorMsg("testConnectUserPasswordWrong",
            "Wrong user-password should throw an exception"));
    assertEquals(EnumConnectionType.NOT_CONNECTED,
        repositoryDataSource.getConnectionType(),
        getErrorMsg("testConnectUserPasswordWrong",
            "Connection type must be 'NOT_CONNECTED'"));
}

```

Calidad continua

En los últimos sprints, correspondientes a la entrega de prototipos funcionales al tutor, se definía una tarea para analizar la deuda técnica que indicaba la Codacy. Se analizaban las tareas de mejoras propuesta por Codacy según su orden de prioridad y se intentaban eliminar. De esta forma además de aprender de las revisiones automáticas de Codacy se controlaba la deuda técnica, en el prototipo evolutivo. La valoración final del proyecto ha sido la máxima (A) sin ninguna revisión de calidad de alta prioridad. **Comentario de revisión...** Añadir url de codacy de tu proyecto y una captura de pantalla donde aparezca la A.

Integración y despliegue continuo CI/CD

Se han utilizado los sistemas de integración continua y despliegue continuo de GitLab para controlar el correcto funcionamiento de la aplicación después de un cambio y para mejorar la calidad de las revisiones. La integración continua se ha configurado para que cada vez que se realiza un commit en la rama principal del repositorio se compruebe que compila correctamente, que se ejecutan los test configurados en maven y se empaquete la aplicación Web en fichero war y se despliegue en Heroku. Si alguna tarea falla se indica en el repositorio con un icono redondo rojo con aspa blanca. **Comentario de revisión...**añadir una captura de pantalla con el badge Heroku del Readme de Gitlab y explicar como se actualiza cada vez que se realiza un commit. **Comentario de revisión...**añadir una captura de pantalla con los check verdes en la lista de commit del repositorio de GitLab

Trabajos relacionados

6.1. Activiti-Api

[9] Aplicación que permite realizar un estudio del estado de un proyecto en GitHub mediante distintas métricas de evolución. Permitiendo tener una visión del ritmo de trabajo del proyecto, duración, número de participantes y actividad.

Es un proyecto bastante parecido y del cual se han asentado muchas bases para este proyecto.

6.2. Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones

[9] Este trabajo ha sido la base para implementar el motor de métricas de la aplicación.

6.3. Software Project Assessment in the Course of Evolution - Jacek Ratzinger

[9] Todas las métricas utilizadas en este proyecto han surgido de las definiciones de esta tesis.

Conclusiones y Líneas de trabajo futuras

7.1. Conclusiones

En este proyecto se ha aprendido a configurar una aplicación web en Java, realizar interfaces gráficas usando solo Java, configurar un proyecto de forma que se facilite la integración y despliegue continuos y la importancia que tiene para un equipo de desarrollo software, la importancia del uso de herramientas que midan la calidad de código, la importancia que tienen las métricas de proceso a la hora de gestionar un proyecto software, la comodidad y eficiencia que tienen las metodologías ágiles.

7.2. Lineas de trabajo futuras

Se definen en lista ideas que podrían realizarse en el futuro:

- Extender la funcionalidad a nuevas métricas de evolución
- Extender las plataformas de desarrollo colaborativo a otras como GitHub, Bitbucket
- GitLab ofrece la posibilidad a los usuarios de tener su propia instancia de GitLab en un servidor propio. Por ahora solo se puede conectar al host “<https://gitlab.com/>”, se podría ampliar esta funcionalidad permitiendo realizar una conexión a servidores propios
- Realizar un histórico de mediciones y almacenarlo en una base de datos
- Hacer que la aplicación web sea adaptable (*responsive*)
- Internacionalizar la aplicación
- Los proyectos y perfiles de métricas importados y exportados se almacenan en un buffer en memoria. Mientras el proyecto sea pequeño no hay problema, pero conforme vaya creciendo habría que implementar otros sistemas de persistencia como bases de datos o ficheros.
- Se podría permitir seleccionar varios proyectos de la tabla de la página principal para poder gestionar varios proyectos a la vez. Por ejemplo: crear un perfil de métricas sólo con los proyectos seleccionados, evaluar solo los proyectos seleccionados, eliminar varios proyectos a la vez, volver a obtener métricas de varios proyectos al mismo tiempo.

- La aplicación web solo soporta una sesión, se podría preparar para poder explotarlo en un entorno multisesión.

Bibliografía

- [1] GitLab4j API (gitlab4j-api) provides a full featured Java client library for working with GitLab repositories via the GitLab REST API: gitlab4j/gitlab4j-api, July 2019. original-date: 2014-02-18T05:09:14Z.
- [2] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Patrones De Diseño: Elementos De Software Orientado a Objetos Reutilizable*. Addison-Wesley, Madrid, 1 ed. en es edition, 2002.
- [3] Diego Güemes-Peña, Carlos López-Nozal, Raúl Marticorena-Sánchez, and Jesús Maudes-Raedo. Emerging topics in mining software repositories. *Progress in Artificial Intelligence*, 7(3):237–247, September 2018.
- [4] ISO/IEC. *ISO/IEC 9126. Software engineering – Product quality*. ISO/IEC, 2001.
- [5] I. Booch Jacobson. *El proceso Unificado de Desarrollo del Software*. Addison Wesley, 2000.
- [6] Raúl Marticorena, Yania Crespo, and Carlos López. Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones. In *X Jornadas Ingeniería del Software y Bases de Datos (JISBD 2005)*, Granada, Spain ISBN: 84-9732-434-X, pages 59–66, September 2005.
- [7] Tim Olshansky. A wrapper for the Gitlab API written in Java. Contribute to timols/java-gitlab-api development by creating an account on GitHub, July 2019. original-date: 2013-05-19T02:48:09Z.
- [8] Jacek Ratzinger. *sPACE: Software Project Assessment in the Course of Evolution*. PhD Thesis, 2007.
- [9] rlp0019. Software para la autoevaluación del proceso de Trabajo de Fin de Grado (TFG) en GitHub.: rlp0019/Activiti-API, June 2019. original-date: 2019-02-17T19:30:04Z.
- [10] Iubaris Info 4 Media SL. Scrum Manager: Temario Troncal I. Versión 2.6.1:94, January 2019.
- [11] I. Sommerville. *Ingeniería de software*. Pearson Education Limited, México, 6^a edition, 2002.
- [12] TFGM_GII UBU. Plantilla Latex. Contribute to ubutfgm/plantillaLatex development by creating an account on GitHub, June 2019. original-date: 2016-10-14T10:17:24Z.