



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 — 20 de junio
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" de TFG.

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, 20 de junio de 2019

Vº. Bº. del Tutor:

D. Carlos López Nozal

Resumen

Aplicación Web en Java que toma como entrada un conjunto de direcciones de repositorios públicos o privados y calcula medidas de la evolución que permiten comparar los repositorios.

Descriptores

gitlab, proyectos, repositorios, métricas, análisis, comparador, software

Abstract

Java Web Application that takes a set of addresses of public or private repositories and calculates measures of evolution that allow comparing the repositories.

Keywords

gitlab, projects, repositories, metrics, analysis, comparator, software

Índice general

Índice general	III
Índice de figuras	IV
Índice de tablas	V
Introducción	1
Objetivos del proyecto	3
2.1. Requisitos	3
Conceptos teóricos	5
3.1. Evolución de software	5
3.2. Framework de medición	6
Técnicas y herramientas	9
4.1. Herramientas utilizadas	9
Aspectos relevantes del desarrollo del proyecto	11
5.1. Definición de métricas	11
Trabajos relacionados	17
Conclusiones y Líneas de trabajo futuras	19
Bibliografía	21

Índice de figuras

3.1. Diagrama del Framework para el cálculo de métricas con perfiles.	7
-----------------------------------------------------------------------	---

Índice de tablas

Introducción

El trabajo se centra en crear una aplicación web que permita, a partir de URLs de repositorios GitLab, comparar dos o varios repositorios mediante métricas de evolución. Las métricas que se van a trabajar, en principio, son:

- Número total de issues (I1)
- Commits por issue (I2)
- Porcentaje de issues cerradas (I3)
- Media de días en cerrar una issue (TI1)
- Media de días entre commits (TC1)
- Días que han pasado entre el primer y último commit (TC2)
- Ratio de actividad de commits por mes (TC3)
- Número de commits en el mes pico (C1)

Se pretende que el proyecto pueda ser ampliado a más gestores de repositorios como GitHub o Bitbucket y que pueda calcular más métricas de las que originalmente se han trabajado, por tanto deberá presentar un diseño que facilite el mantenimiento de la misma y añadir nuevas funcionalidades.

Objetivos del proyecto

2.1. Requisitos

En este apartado se detallarán los requisitos funcionales, marcados por la aplicación que se está desarrollando y los no funcionales, planteados durante el desarrollo de la aplicación.

Requisitos funcionales

En primer lugar se detallan requisitos generales que surgen a partir del planteamiento del problema y los objetivos que se desean conseguir con este proyecto. En segundo lugar se definirán los requisitos más específicos de cada subsistema.

- Se desea obtener métricas de evolución de uno o varios repositorios especificando sus URLs, con el objetivo de compararlos.
- Las métricas que se desean calcular de un repositorio son las siguientes:
 - Número total de asuntos
 - Cambios por asunto
 - Porcentaje de asuntos cerrados
 - Media de días en cerrar un asunto
 - Media de días entre cambios
 - Días entre primer y último cambio
 - Rango de actividad de cambios por mes
 - Porcentaje de pico de cambios

Requisitos no funcionales

Conceptos teóricos

3.1. Evolución de software

Conceptos de evolución

Un sistema de Gestión de Configuración del Software (SCM) es capaz de gestionar la evolución y cambios del código fuente en el tiempo [1, 3]. En este apartado describiremos los conceptos básicos que utilizaremos para recoger, documentar, almacenar y recuperar los diferentes cambios que se produzcan en las entidades que formen parte nuestro sistema. Un proyecto software está compuesto por múltiples ficheros y directorios. Llamaremos ítem a cualquier fichero o directorio cuya evolución en el tiempo está controlada por un sistema de Gestión de Configuración del Software (SCM). Una vez que estén bajo control, será posible ver su historial de cambios o revisiones y recuperar un estado anterior. A medida que los ítems evolucionan en el tiempo, se van creando nuevas revisiones de los mismos. Como es lógico, algunos ítems sufrirán más cambios que otros a lo largo del desarrollo. El SCM almacena todas las revisiones de cada ítem, de manera que es posible volver al estado de uno de estos elementos en un momento dado. La creación de revisiones ocurre a través de las operaciones de desprotección (check-out) y protección (check-in). Cuando se va a hacer una modificación en un ítem, éste se desprotege para editarlo. Para crear una nueva revisión del ítem de forma que se pueda recuperar su estado en este punto, se protege, lo que indica al SCM que debe almacenar los nuevos contenidos del ítem. Al conjunto de revisiones de un ítem se le denomina historia y resume la evolución de ese ítem en el tiempo. Una rama es un contenedor de revisiones, capaz de almacenar la evolución de los ítems como se muestra en la Ilustración 1. Las ramas pueden contener revisiones de más de un ítem. De hecho, esta es la

situación más habitual. Ilustración 1: Concepto de rama

Una etiqueta o label es el modo de marcar revisiones para poder agruparlas según un cierto criterio, que normalmente fija el usuario. Cuando se aplica una etiqueta, se crea una instantánea de la situación de los ítems en el tiempo. Más tarde esa instantánea puede ser referenciada con facilidad para identificar ese momento específico. Una etiqueta es, en definitiva, un nombre más fácil de recordar que se asigna a un conjunto particular de revisiones. Las etiquetas se aplican siempre a las revisiones de los ítems que se encuentren actualmente en el espacio de trabajo o workspace, que es la zona donde el SCM puede mantener ítems bajo el control de versiones. A efectos prácticos, el workspace no dejará de ser un directorio en el disco. Los ítems, sus revisiones, las ramas donde se almacenan las revisiones y las etiquetas que agrupan las revisiones se almacenan en un repositorio, que será el espacio principal donde el SCM guarda todos los objetos.

Métricas de evolución

El conjunto de métricas utilizados en este proyecto proceden de la Master Tesis titulada *sPACE: Software Project Assessment in the Course of Evolution* [2].

3.2. 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 [8]. Este framework (3.1) es independiente del lenguaje y su objetivo es la reutilización en la implementación del cálculo de métricas.

Técnicas y herramientas

4.1. Herramientas utilizadas

- Eclipse Java EE IDE for Web Developers. Version: 2018-09 (4.9.0)
- Apache Maven v3.6.0
- Apache Tomcat v9.0.13
- Java SE 11 (JDK)
- gitlab4j-api v4.8.56
- JUnit 5 v5.3.1

Aspectos relevantes del desarrollo del proyecto

5.1. Definición de métricas

I1 - Número total de issues

- **Categoría:** Proceso de Orientación
- **Nombre:** NumeroTotalIssues
- **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:** Repositorio de un gestor de repositorios
- **Interpretación:** $NTI \geq 0$. Mejor valores bajos
- **Tipo de escala:** Absoluta
- **Tipo de medida:** $NTI = \text{Contador}$

I2 - Commits por issue

- **Categoría:** Proceso de Orientación
- **Nombre:** CommitsPorIssue

- **Descripción:** Número de commits por issue
- **Propósito:** ¿Cuántos commits realizados por cada issue?
- **Fórmula:** $CI = NTC/NTI$. NTI = Numero total de issues, NTC = Número total de commits
- **Fuente de medición:** Repositorio de un gestor de repositorios
- **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:** NTI, NTC = Contador

I3 - Porcentaje de issues cerradas

- **Categoría:** Proceso de Orientación
- **Nombre:** PorcentajeIssuesCerradas
- **Descripción:** Porcentaje de issues cerradas
- **Propósito:** ¿Qué porcentaje de issues definidas en el repositorio se han cerrado?
- **Fórmula:** $PIC = NTIC*100/NTI$. NTIC = Número total de issues cerradas, NTI = Numero total de issues
- **Fuente de medición:** Repositorio de un gestor de repositorios
- **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:** Proceso de Orientación
- **Nombre:** PorcentajeIssuesCerradas
- **Descripción:** Porcentaje de issues cerradas

- **Propósito:** ¿Qué porcentaje de issues definidas en el repositorio se han cerrado?
- **Fórmula:** $PIC = NTIC * 100 / NTI$. NTIC = Número total de issues cerradas, NTI = Numero total de issues
- **Fuente de medición:** Repositorio de un gestor de repositorios
- **Interpretación:** $0 \leq PIC \leq 100$. Cuanto más alto mejor.
- **Tipo de escala:** Ratio
- **Tipo de medida:** NTI, NTIC = Contador

TC1 - Media de días entre commits

- **Categoría:** Constantes de tiempo
- **Nombre:** MediaDiasEntreCommits
- **Descripción:** Media de días que pasan entre dos commits consecutivos
- **Propósito:** ¿Cuánto tiempo suele pasar desde un commit hasta el siguiente?
- **Fórmula:** $MDEC = [\text{Sumatorio de } (TC_i - TC_j) \text{ desde } i=1, j=0 \text{ hasta } i=NTC] / NTC$. NTC = Número total de commits, TC = Tiempo de Commit
- **Fuente de medición:** Repositorio de un gestor de repositorios
- **Interpretación:** $MDEC > 0$. 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
- **Nombre:** DiasEntrePrimerYUltimoCommit
- **Descripción:** Días transcurridos entre el primer y el ultimo commit

- **Propósito:** ¿Cuántos días han pasado entre el primer y el último commit?
- **Fórmula:** $DEPUC = TC2 - TC1$. $TC2$ = Tiempo de último commit, $TC1$ = Tiempo de primer commit.
- **Fuente de medición:** Repositorio de un gestor de repositorios
- **Interpretación:** $DEPUC \geq 0$
- **Tipo de escala:** Absoluta
- **Tipo de medida:** TC = Tiempo

TC3 - Ratio de actividad de commits por mes

- **Categoría:** Constantes de tiempo
- **Nombre:** RatioCommitPorMes
- **Descripción:** Muestra el número de commits relativos al número de meses
- **Propósito:** ¿Cuál es el número medio de cambios por mes?
- **Fórmula:** $RCM = NTC / 12$
- **Fuente de medición:** Repositorio de un gestor de repositorios
- **Interpretación:** $RCM > 0$. Cuanto más alto mejor
- **Tipo de escala:** Ratio
- **Tipo de medida:** NTC = Contador

C1 - Número de commits en el mes pico

- **Categoría:** Constantes de tiempo
- **Nombre:** ContadorCommitsPico
- **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:** $CCP = NCMP / NTC$. $NCMP$ = Número de commits en el mes pico, NTC = Número total de commits
- **Fuente de medición:** Repositorio de un gestor de repositorios
- **Interpretación:** $0 \leq CCP \leq 1$. Mejor valores intermedios
- **Tipo de escala:** Ratio
- **Tipo de medida:** $NCMP, NTC$ = Contador

Trabajos relacionados

- [Activiti-API](#)
- [Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones](#)
- [Software Project Assessment in the Course of Evolution - Jacek Ratzinger](#)

Conclusiones y Líneas de trabajo futuras

Bibliografía

- [1] Steve Berczuk and Brad Appleton. *Software Configuration Management Patterns: Effective Teamwork, Practical Integration*. ADDISON WESLEY LONGMAN INC DIV PEARSON SUITE 300, Boston, edición: 01 edition, November 2002.
- [2] Jacek Ratzinger. *sPACE: Software Project Assessment in the Course of Evolution*. PhD Thesis, 2007.
- [3] I. Sommerville. *Software Engineering, Global Edition*. Pearson Education Limited, 2016.