MEDICIÓN AUTOMÁTICA DE ASPECTOS HUMANOS EN EQUIPOS VIRTUALES DE DESARROLLO DE SOFTWARE

Gustavo Sevilla, Sergio Zapata, Facundo Gallardo, Estela Torres

Instituto de Informática / Facultad de Ciencias Exactas, Físicas y Naturales / Universidad Nacional de San Juan; {gsevilla,szapata,etorres,fgallardo}@iinfo.unsj.edu.ar

RESUMEN

En los últimos años la comunidad científica ha puesto mayor atención en el importante rol que juegan los aspectos humanos y sociales en el proceso de desarrollo de software. Hay un fuerte consenso que estos aspectos impactan significativamente tanto en el proceso de software como en el producto resultante.

Paralelamente, el progreso de las tecnologías de información y comunicación (TICs) y el crecimiento de prácticas de globalización han fomentado la creación de los denominados equipos virtuales de desarrollo de software. Equipos en donde los miembros interactúan en forma remota y mediante TICs. En este contexto aparece el uso extendido de las plataformas de colaboración para el desarrollo de software (software forges) como los repositorios Github, Jira, Stackoverflow, etc.

Este trabajo pretende sacar provecho de los datos que quedan registrados en estos repositorios de software y que de alguna manera reflejan las características de las interacciones entre los miembros de equipos virtuales de software. La medición y análisis de estos datos puede arrojar evidencia de la calidad de las relaciones humanas del equipo virtual.

Palabras clave: Equipos Virtuales, Software forges, Aspectos humanos y sociales en el desarrollo de software

CONTEXTO

Este trabajo está enmarcado en el proyecto "Aspectos Humanos del Desarrollo Global de Software" financiado por la Universidad Nacional de San Juan, siendo la unidad ejecutora el Instituto de Informática (IdeI) de la misma Universidad.

El proyecto es de ejecución bianual y su comienzo fue en enero de 2019. El mismo está inserto dentro del Programa de Incentivos a docentes-investigadores de la Secretaría de Políticas Universitarias del Gobierno Nacional.

1. INTRODUCCION

En los proyectos de desarrollo global o distribuido de software los integrantes de los equipos virtuales no tienen interacción personal cara a cara, excepto esporádicas y costosas reuniones localizadas. Aquí los aspectos humanos se podrían ver perjudicados por la distancia física, cultural y horaria. En tales entornos, conocer el nivel de relaciones humanas es esencial para que los líderes de proyectos adviertan posibles problemas que pudieran afectar el desarrollo de software. El desarrollador más talentoso podría tener un rendimiento inferior y, finalmente, abandonar el proyecto solo porque no está contento con su entorno o sus colegas [1].

Un equipo de desarrollo global de software es aquel que está separado geográficamente mientras colabora activamente en un proyecto de software común [2].

Los Equipos Virtuales (EV) o globales, permiten, entre otras ventajas, integrar personal calificado de distinta localización geográfica y muchas veces a un menor costo laboral. A su vez, los EV plantean problemas o desafíos tal como la gestión remota de proyectos, donde uno de los principales

factores que la afectan son las relaciones sociales entre los miembros. Las buenas relaciones sociales interpersonales en EV son importantes para la comunicación, la motivación y la creación de un buen ambiente de trabajo que contribuya a la mejora de la productividad.

La cohesión del equipo es el grado en que sus miembros se identifican entre sí y con el grupo en general [3, 4]. Los equipos globales tienden a menores niveles de cohesión que los equipos colocalizados, en gran parte debido a que colaboran remotamente a través de tecnología en lugar de hacerlo cara a cara. Por otro lado, investigaciones previas sugieren que la falta de cohesión grupal se puede superar mediante el intercambio de más comunicaciones sociales entre los miembros del grupo [5]. Las interacciones sociales permiten a los miembros del equipo crear vínculos más fuertes, lo cual permitirá aumentar los niveles de cohesión del grupo.

Por otra parte, una de las repercusiones del uso de las TIC en las actuales relaciones sociales tiene que ver con el registro digital de ellas en repositorios de datos llamadas plataformas de software forges, las cuales ofrecen una oportunidad de analizar esos registros para medir y evaluar estas relaciones. Las software forges proveen varias herramientas y facilidades a los equipos de desarrollo distribuido, tales como sistemas de control de código fuente, listas de correos electrónicos y foros de comunicación, sistemas de seguimientos de fallas, espacio para alojamiento web y similares [6].

Así, el presente trabajo ofrece un aporte en este contexto, proponiendo una aplicación que explote los datos sociales almacenados en las software forges.

Github es un repositorio, extensamente usado, para alojar proyectos de software que utiliza el sistema de control de versiones de Git. La plataforma integra una serie de características y funcionalidades que hacen única su información sobre las actividades de los desarrolladores de software a código abierto [7].

Github utiliza una forma de trabajo basada en Pull Request (PR), un nuevo método colaborativo de desarrollo distribuido de software. Un PR es un mecanismo de solicitud de cambio de código de software que facilita la colaboración entre desarrolladores, ese cambio puede ser para implementar una mejora, adicionar una nueva funcionalidad o para remover un defecto. El PR es la unidad de trabajo de GitHub y representa de alguna manera el trabajo colaborativo realizado por sus miembros. Un PR es una solicitud de cambio al software.

2. LINEAS DE INVESTIGACION Y DESARROLLO

Comprender el impacto de los aspectos humanos de los equipos virtuales de desarrollo de software en el proceso de desarrollo de software.

Objetivos específicos: a) Medir y evaluar aspectos humanos del desarrollo de software basados en las interacciones registradas en las plataformas software forges. b) Medir y evaluar indicadores de calidad y éxito de los procesos de desarrollo de software basados en las interacciones registradas en las plataformas software forges. c) Relacionar ambas mediciones realizadas.

3. RESULTADOS OBTENIDOS

Se desarrolló una aplicación web que, en base a datos extraídos automáticamente desde un repositorio de software, respecto de un proyecto de desarrollo de software, brinda información sobre los aspectos humanos del equipo de desarrollo.

Específicamente la aplicación genera información analítica, en la forma de gráficos, en base a datos extraídos de la plataforma Github con el fin de poder medir, visualizar y analizar las interacciones, y de allí relaciones humanas, que se producen entre los participantes de equipos de trabajo distribuidos.

3.1. Arquitectura de la aplicación

La aplicación se implementa como servicio web. En la Figura 1 se puede observar la arquitectura de software diseñada. Para su desarrollo se utilizó el framework Vue.js desplegado en la plataforma Netlify. Vue es un framework de javascript progresivo que permite desarrollar de manera rápida e incremental. Netlify es una plataforma que ofrece servicios de alojamiento y backend sin servidor (serverless) para sitios web estáticos.

Para hacer el relevamiento de información sobre los usuarios y sus actividades dentro del proyecto de software alojado en el repositorio de Github se utilizó la API v4 del mismo. Esta versión de la API utiliza GraphQL, el cual es un lenguaje de consulta para API y un motor en tiempo de ejecución para cubrir esas consultas con los datos existentes. Tanto las consultas como la respuesta de parte de la API usan el formato JSON.

Figura 1: Arquitectura de la aplicación

A los datos obtenidos se les aplicó un algoritmo propio de conteo que genera una matriz de interacciones entre los miembros de un PR. Se considera interacción de un miembro p1 hacia otro p2 cuando hay un comentario o reacción (emoji en respuesta a un comentario) dirigido de p1 a p2 o de p1 a todos.

3.2. Métricas para la Medición de Interacciones

Con la matriz de interacciones generada es posible aplicar técnicas de extracción de evidencia. Utilizamos las fórmulas expuestas en la Tabla 1 para el cálculo de aspectos sociales.

Tabla 1. Métricas de Aspectos Sociales

Métrica	Fórmula
Cohesión Interpersonal entre el miembro *i* del equipo y el miembro *j*. (*Mij* mensajes del miembro *i* al miembro *j*) [8].	$CINT_{ij} = 1 - \frac{\lvert M_{ij} - M_{ji} \rvert}{M_{ij} + M_{ji}}$
Cohesión Individual del miembro *i* del equipo. (*G* conjunto de miembros del grupo) [8].	$CIND_i = \frac{\sum_{j=1}^{\lvert G \rvert - 1} CINT_{ij}}{\lvert G \rvert - 1}$
Cohesión grupal del equipo *e*. (*G* conjunto de miembros del grupo) [8].	$CGR_e = \frac{\sum_{i=1}^{\lvert G \rvert} CIND_i}{\lvert G \rvert}$
Mímica o Similitud de vocabulario entre los miembros *i* y *j* del equipo. *SC* es la función coseno de similaridad, *Wi* conjunto de palabras usadas por miembro *i* y *Wj* conjunto de palabras usadas por miembro *j* [9,10].	$CMM_{ij} = SC(W_i, W_j)$
Polaridad de los comentarios entre dos miembros. Donde C_{+ij} es la cantidad de comentarios con polaridad positiva de *i* a *j*, y *Cij* es el total de comentarios de *i* a *j* (comentarios que *i* escribió en PR de *j* o comentarios donde *i* mencionó a *j*) [10].	$P_{ij} = \frac{C_{+ij}}{C_{ij}}$
Capacidad de un miembro donde *PRa(i)* es la cantidad de PRs creados por *i* que hicieron merge, y *PR(i)* es la cantidad de PRs creados por *i* [10].	$M_i = \frac{PRa(i)}{PR(i)}$
Colaboración entre los miembros donde *I(mi,mj)* es el número de interacciones entre *i* y *j*, y *I(mi)* es el total de interacciones de *i* y *j* con los miembros del PR [10].	$CL_{ij} = \frac{I(mi, mj)}{I(mi)}$
Colaboración Individual del miembro *i* del equipo. (*G* conjunto de miembros del grupo) [10].	$CLI_i = \frac{\sum_{j=1}^{\lvert G \rvert - 1} CL_{ij}}{\lvert G \rvert - 1}$
Colaboración grupal del equipo *e*. (*G* conjunto de miembros del grupo) [10].	$CLG_e = \frac{\sum_{i=1}^{\lvert G \rvert} CL_i}{\lvert G \rvert}$

3.3. Funcionalidades principales

La aplicación emplea un conjunto de cálculos que implementan las métricas mencionadas en la sección anterior, ofrece dos módulos

 RedUNCI - UNdeC. ISBN: 978-987-24611-3-3

funcionales que reportan datos estadísticos, uno a nivel de PRs y el otro a nivel de Proyecto.

3.3.1. Módulo de métricas de Pull Requests

Este módulo muestra información sobre las interacciones sociales entre los miembros de un determinado PR. Por una parte se muestra un resumen de las estadísticas grupales, ver Figura 2. En ella se informa, en una escala porcentual, el grado de cohesión, de colaboración, de mímica y de polaridad alcanzado por el grupo de miembros del PR. Estas estadísticas grupales se calculan promediando los resultados de las métricas individuales de los miembros del PR.

Figura 2: Gráficos Métricas Grupales en un PR

En el ejemplo mostrado en la Figura 2 se puede observar un bajo nivel de colaboración en el grupo lo cual puede ser una oportunidad para tomar decisiones al respecto por parte del líder del grupo.

También se informan, en una escala porcentual, métricas individuales como nivel de cohesión y colaboración de cada miembro del PR, ver Figura 3. La cohesión individual nos indica cuán cohesionado está cada miembro respecto del grupo de trabajo, mientras que la colaboración nos señala el nivel de actividad colaborativa individual.

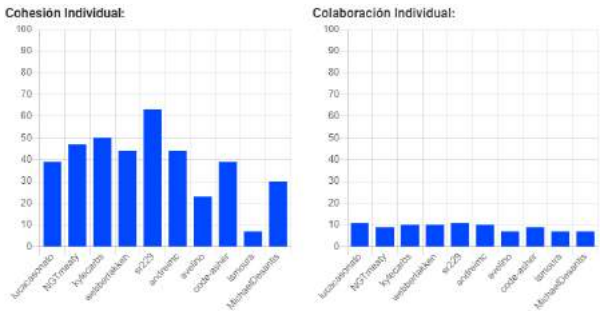

Figura 3: Gráficos de Cohesión y Colaboración Individual en un PR.

En el ejemplo presentado en la Figura 3 se puede notar que el miembro ismoura tiene bajos niveles de cohesión y colaboración, ésta es otra posible oportunidad de toma de decisiones para los administradores del proyecto.

Por último, se muestra una tabla con la cantidad de mensajes enviados, mensajes recibidos, mímica y polaridad individual de cada miembro, ver Figura 4.

Participante	Msj. Enviados	Msj. Recibidos	Mímica	Polaridad
lucacasonato	71	24	0.30	0.83
NGTmeaty	20	19	0.20	0.00
kylecarbs	22	23	0.19	0.50
webberlakken	12	22	0.26	1.00
sr229	18	21	0.12	1.00
andreimc	39	18	0.25	0.27

Figura 4: Tabla de mensajes enviados y recibidos, Mímica y Polaridad en comentarios de un PR.

3.3.2. Módulo de métricas de proyecto

Este módulo reporta datos de todos los PR de un proyecto de software determinado. En primer lugar muestra una tabla donde cada fila es un PR del proyecto, y las columnas indican los resultados de las métricas de interacción grupales (cohesión, colaboración, mímica, polaridad), junto con datos extraídos de Github sobre el PR. En este caso los líderes de proyecto pueden observar los aspectos sociales en una forma comparativa entre los distintos PRs. De esa manera podrían detectar la evolución de dichos aspectos en el tiempo.

También se puede ver una tabla donde cada fila es un miembro que participó en algún PR del proyecto y las columnas indican los resultados de las métricas de interacción individuales (cohesión individual, colaboración individual, polaridad, habilidad), junto con datos extraídos de Github sobre el miembro. En la Figura 5 se puede observar un ejemplo de los datos mencionados.

Miembro	Cohesión Individual	Colaboración Individual	Capacidad	Cant. PR Autor	Polaridad	Cant. PR Participa	Rol
Noel2	0.23	0.55	1.00	2	0.32	5	contribuyente
Victorescco	0.13	0.22	0.60	34	0.34	59	contribuyente
andrei	0.60	0.51	0.50	6	0.45	78	miembro
NGTmeaty	0.42	0.60	0.70	12	0.25	16	contribuyente
vapurr	0.30	0.28	1.00	1	0.20	1	contribuyente
CodeMar	0.66	0.37	0.88	87	0.68	134	miembro

Figura 5: Estadísticas de Miembros en un Proyecto de Software.

El módulo también muestra un resumen de las estadísticas grupales del proyecto, cada una de las métricas mostradas en el resumen se calcula promediando los valores obtenidos de la tabla anterior. En la Figura 6 se puede observar un ejemplo de los datos mencionados.

Figura 6: Gráficos Métricas Grupales en un Proyecto de Software.

4. FORMACION DE RECURSOS HUMANOS

El equipo de trabajo está formado por los autores del trabajo y profesores colaboradores de las universidades del Quindío (Colombia), ORT (Uruguay) y de la Frontera (Chile).

Con la presente línea de Investigación y Desarrollo se relacionan dos tesis de Doctorado, una en desarrollo y la otra en sus inicios, más una tesina de grado muy próxima a ser presentada.

5. BIBLIOGRAFIA

[1] Murgia, A., Tourani, P., Adams, B., & Ortu, M. (2014, May). Do developers feel emotions? an exploratory analysis of emotions in software artifacts. In Proceedings of the 11th working conference on mining software repositories (pp. 262-271). ACM

[2] Carmel, E. (1999). Global software teams: collaborating across borders and time zones. Prentice Hall PTR.

[3] Carron, A. V., Widmeyer, W. N., and Brawley, L. R. (1985). The development of an in-strument to assess cohesion in sport teams: The Group Environment Questionnaire, Journal of Sport Psychology 7 244-266.

[4] Chidambaram, L. (1996). Relational development in computer-supported groups, MIS Quar-terly 20(2) 143-165.

[5] Carless, S. A., & De Paola, C. (2000). The measurement of cohesion in work teams. Small group research, 31(1), 71-88.

[6] Squire, M., & Williams, D. (2012, January). Describing the software forge ecosystem. In 2012 45th Hawaii International Conference on System Sciences (pp. 3416-3425). IEEE.

[7] Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012, February). Social coding in GitHub: transparency and collaboration in an open software repository. In Proceedings of the ACM 2012 conference on computer supported cooperative work (pp. 1277-1286).

[8] Gallardo, F., Sevilla, G., Torres, E. y Zapata, S. (2019). Análisis de Datos de GitHub para Evaluar Cohesión en Equipos Virtuales de Desarrollo de Software. Instituto de Informática – Universidad Nacional de San Juan.

[9] Vincent (MSVCode). (2018). https://github.com/MSVCode/doc-similarity.

[10] Da Cruz, G. A. M., Moriya-Huzita, E. H., & Feltrim, V. D. (2018). ARSENAL-GSD: A framework for trust estimation in virtual teams based on sentiment analysis. Information and Software Technology, 95, 46-61.