

Clasificación de Replicaciones para la Síntesis de Experimentos en Ingeniería del Software

Omar Salvador Gómez Gómez

Facultad de Informática,
Universidad Politécnica de Madrid,
Boadilla del Monte 28660, Madrid, España
ogomez@ieee.org

Resumen. La Ingeniería del Software Experimental (ESE en Inglés) traslada a la Ingeniería del Software (IS) el paradigma experimental que se ha aplicado con éxito en diversas disciplinas científicas. El objetivo de la ESE es hacer del desarrollo de software una actividad predecible gracias al conocimiento de las relaciones entre los procesos de producción de software y los productos que se obtienen. Para avanzar en el paradigma experimental en IS no es suficiente con aplicar las técnicas de diseño experimental y el análisis estadístico de datos, sino que es necesario construir una metodología (bien desde cero o adaptada de otras disciplinas) basada en los principios generales del experimentalismo. La motivación principal de esta investigación es trabajar en la adaptación de un aspecto particular del paradigma experimental a la experimentación en IS: La replicación, en el contexto de una serie de experimentos. El objetivo de la investigación consiste en identificar y definir una clasificación de replicaciones para experimentos en IS. Esta clasificación permitirá prescribir los distintos niveles de síntesis de resultados que se recomiendan para los diferentes tipos de replicaciones identificadas.

Palabras clave: Clasificación de Replicaciones, Replicaciones de Experimentos, Síntesis de Replicaciones, Paradigma Experimental.

1 Introducción

En la actualidad, los Ingenieros de Software cuentan con una variedad de herramientas, lenguajes de programación, tecnologías, metodologías y procesos que pueden utilizar en las distintas etapas del desarrollo de software. A menudo, los Ingenieros de Software tienen la necesidad de evaluar y seleccionar en base a un criterio las herramientas, lenguajes, metodologías o procesos más adecuados para el desarrollo o mantenimiento de un producto software. El realizar una evaluación errónea de estos elementos puede ocasionar retrasos en la entrega del producto, incrementos en los costes, o defectos en el producto resultante.

Actualmente, la comunidad científica de IS está de acuerdo en la carencia de una base teórica sólida para la disciplina [1–5]. Esto significa que no existe información suficiente para una adecuada evaluación y selección de las herramientas,

lenguajes, metodologías o procesos a usar en IS. A través de la experimentación es posible evaluar y seleccionar de forma adecuada y objetiva los diferentes elementos involucrados en el desarrollo de software. Por ejemplo, si a un Ingeniero le interesa evaluar de manera objetiva la efectividad de dos técnicas de detección de faltas puede hacerlo a través de un experimento.

La experimentación comenzó a utilizarse en disciplinas científicas como la Física y la Química, desde entonces ésta ha sido adaptada y utilizada en otras disciplinas. Por ejemplo, Campbell y Stanley [6], Cook y Campbell [7], Cohen [8], entre otros más, han contribuido en la adaptación del paradigma experimental en disciplinas de las Ciencias Sociales como son la Psicología, Sociología y Antropología.

En la Ingeniería del Software (IS) se han realizado algunas adaptaciones del paradigma experimental. Por ejemplo, Singer [9] propone adaptar las guías de estilo de la APA (American Psychological Association) para reportar experimentos en IS. Kitchenham et. al. [2], así como Jedlitschka y Pfahl [10] han desarrollado dos tipos de guías para reportar experimentos controlados, mientras que Basili et al. [11], Wohlin et al. [12], así como Juristo y Moreno [13] han trabajado en la adaptación de procesos para la realización de experimentos.

La motivación principal de la presente investigación es trabajar en la adaptación de un aspecto particular del paradigma experimental a la experimentación en IS. Este aspecto es la repetición o replicación de experimentos de IS en el contexto de una serie de experimentos afines.

El objetivo de esta investigación consiste en identificar y definir los posibles tipos de repeticiones que pueden existir en experimentos en IS. Los diferentes tipos de repeticiones implicarán diferentes niveles de síntesis de los resultados experimentales. Utilizando este conocimiento, se generará un proceso que permita llevar a cabo la síntesis de conjuntos de repeticiones. El resto del artículo está organizado de la siguiente manera: La sección 2 presenta el estado del arte donde se describe de manera específica la problemática a tratar. En la sección 3 se describe la aproximación a la resolución de la problemática. En la sección 4 se presentan algunos resultados preliminares. Por último, en la sección 5 se describe la contribución de la tesis.

2 Estado del Arte

En sus inicios, la Ingeniería del Software Experimental (ESE) se centró en estudiar la aplicación a la IS de los principios del laboratorio y el experimento. Tras un par de décadas, la realización de experimentos en ESE se ha convertido en una tarea bastante bien comprendida. Pero la realización de experimentos aislados es sólo un primer paso en la secuencia de actividades del paradigma experimental. Queda pendiente el análisis y adaptación de otros principios del experimentalismo. Actualmente, la investigación en ESE se centra, no ya en la realización de experimentos de laboratorio a pequeña escala (aislados) sino pasar a la experimentación a gran escala (concatenando conjuntos de experimentos)

[14]. Tal migración dejará atrás los resultados poco generalizables para producir evidencias sobre la construcción de software.

Para llegar a la generalización de resultados en la ESE, es necesario repetir un experimento en distintos contextos; es decir, contar con varias repeticiones. Una vez que se tienen varias repeticiones, debe llevarse a cabo la síntesis de resultados para obtener alguna evidencia concluyente del fenómeno estudiado.

La realización de repeticiones en ESE ha sido complicada debido al estado de madurez (fruto de su juventud) en que se encuentra la disciplina, cualquier cambio en la estructura de una replicación por mínimo que sea puede ocasionar una alteración en los resultados. En ESE, se han realizado algunos intentos de sintetizar resultados de conjuntos de repeticiones [15–18] a través del meta-análisis, no obstante los resultados no han sido satisfactorios, ya que en las repeticiones se han introducido tantos cambios que no se tiene el conocimiento necesario para interpretar los resultados así como para compararlos.

Muchas de las veces los experimentadores de manera consciente o inconsciente modifican de forma no sistemática los elementos del experimento a replicar. Estas modificaciones ocasionan una variabilidad importante en los resultados de la replicación. Debido a estas variaciones es difícil corroborar los hallazgos del experimento previo. Ante esta situación, se ha comenzado a sugerir el uso de otras formas de validar los resultados tales como la triangulación [20].

Los resultados obtenidos de las repeticiones hechas en ESE han originado cierta controversia respecto a la forma de verificar los hallazgos previamente observados. Por una parte están quienes recomiendan realizar repeticiones reutilizando parte de los materiales del experimento base [21, 22], y por otra parte están quienes recomiendan efectuar las repeticiones utilizando métodos y materiales distintos a los del experimento base [23, 19].

Esta discusión puede deberse a que los autores parecen estar hablando de diferentes tipos de replicación. Aunque la función principal de la replicación es verificar algún hallazgo observado, la replicación tiene varias funciones particulares, cada una con un propósito específico. El contar con una clasificación de repeticiones para la experimentación en IS ayudará a comprender de una mejor manera las diferentes funciones que desempeña cada tipo de replicación y los cambios que implican cada tipo. En esta investigación se está desarrollando una clasificación de repeticiones que facilite la síntesis de conjuntos de repeticiones.

3 Aproximación a la Resolución

La problemática que se plantea en esta investigación se abordará de la siguiente manera. La resolución partirá del estudio del concepto de replicación en diferentes disciplinas científicas, en estas disciplinas se examinarán los diferentes tipos y formas de replicación comúnmente aceptados. Con los resultados del estudio, se identificarán los elementos que conforman un experimento o situación experimental. Una vez identificados dichos elementos, se analizarán los posibles cambios que se puedan realizar en la situación experimental, estos cambios per-

mitirán identificar los distintos tipos de replicación que se pueden realizar en ESE.

Una vez identificados los tipos de replicación, se llevará a cabo un estudio sobre los diferentes métodos que existen para sintetizar resultados experimentales. Este estudio servirá para prescribir el tipo de síntesis que mejor se adecua a cada uno de los tipos de replicación.

La resolución concluye con el desarrollo de un proceso que permita caracterizar un conjunto de replications así como prescribir los tipos de síntesis que se puedan llevar a cabo en este conjunto.

Para la validación de la resolución se utilizará un conjunto de experimentos replicados referentes al área de pruebas del software. Estas replications tienen diferentes niveles de similitud respecto al experimento base. Algunas de las replications se han realizado en diferentes entornos con diferentes experimentadores, en otras replications se han variado los materiales y artefactos utilizados, es decir, todas las replications de este conjunto contienen diferentes variaciones en su estructura.

4 Resultados Preliminares

Por el momento se cuenta con una versión de la clasificación de replications así como una versión preliminar del proceso caracterizador. Respecto a la clasificación de replications se han identificado seis aspectos que caracterizan una replicación, estos son:

1. Por sitio donde se efectúa la replicación.
2. Por tipo de experimentadores que participan en la replicación.
3. Por el tipo de similitud de la replicación.
4. Por tipo de validez.
5. Por función.
6. Por tipo de cambio.

En cada aspecto se ha realizado una clasificación de replications. A continuación se describe brevemente estos aspectos así como los tipos de replicación identificados.

Por sitio donde se efectúa la replicación. Una replicación es *interna* si se realiza en el mismo sitio donde se efectuó el experimento base, y es *externa* si ésta se realiza en otro sitio distinto al del experimento base.

Por tipo de experimentadores. En este aspecto se tienen tres tipos de replications: *dependiente*, *semi-independiente* o *independiente*. La replicación es *dependiente* si se realiza por los mismos experimentadores que llevaron a cabo el experimento base, es *semi-independiente* si parte de los experimentadores del experimento base participan en la replicación realizada por otros experimentadores, o es *independiente* si ésta se realiza por experimentadores distintos a los del experimento base.

Por tipo de similitud de la replicación. En este aspecto se identificaron cuatro tipos de replicación:

Exacta. En esta replicación se pretende realizar una copia idéntica de un experimento base, sin embargo esto no es posible debido a que siempre hay condiciones que varían entre la replicación y el experimento base. En algunas áreas de las Ciencias Naturales es posible llegar a una aproximación bastante cercana de este tipo de replicación ya que hay un mayor control sobre las condiciones experimentales.

Literal. Es equivalente a la replicación exacta en el sentido de que el experimentador utiliza los mismos materiales y el mismo método que en el experimento base. En este tipo de replicación se asume que existen condiciones imposibles de controlar.

Similar. El experimentador varía de forma sistemática algún elemento del experimento a replicar. Algunos de estos elementos implican la modificación de otros elementos que son dependientes del elemento a modificar.

Alejada. En este tipo de replicación se varía de forma no sistemática más de un elemento a la vez. A diferencia de los anteriores tipos, esta replicación conlleva un mayor riesgo. Por ejemplo, si los resultados difieren del experimento base, es difícil saber cual de los elementos modificados en la replicación influyó en los resultados.

Por tipo de validez. Una replicación puede llevarse a cabo considerando alguno de los tipos de validez existentes como son: validez de conclusión estadística, validez interna, validez de constructo y validez externa [7].

Por función. Se identificaron 14 categorías de funciones que una replicación puede tener. Algunas de estas funciones son: control del error muestral, control de sitio, control de materiales e instrumentos, entre otras más.

Por tipo de cambio. Se han identificado cinco elementos primarios que conforman la estructura de un experimento. Una replicación puede realizarse variando alguno de estos elementos. Los elementos que pueden variar son: operacionalizaciones de los constructos, poblaciones, entorno, experimentadores y protocolo.

5 Contribución de la Tesis

Desde una perspectiva general, la contribución de la tesis se centra en adaptar un aspecto específico del paradigma experimental a la experimentación en IS. Desde una perspectiva particular, la resolución propuesta pretende ayudar a clasificar conjuntos de replications con el fin de realizar diferentes tipos de síntesis. Con las síntesis resultantes se busca obtener diferentes piezas de conocimiento que puedan utilizarse en la práctica del desarrollo de software.

Referencias

1. Tichy, W.: Should Computer Scientists Experiment More?. IEEE Computer, vol. 31, pp. 32–40. IEEE Computer Society Press (1998)
2. Kitchenham, B., Pfleeger, S., Pickard, L., Jones, P., Hoaglin, D., El Emam, K., Rosenberg, J.: Preliminary Guidelines for Empirical Research in Software Engineering. IEEE Trans. on SE, vol. 28, pp. 721–734. IEEE Computer Society Press (2002)

3. Hannay, J.: A Systematic Review of Theory Use in Software Engineering Experiments. *IEEE Trans. on SE*, vol. 33, pp. 87–107. IEEE Computer Society Press (2007)
4. Sjøberg, D.I.K., Dybå, T., Anda, B.C.D., Hannay, J.E.: Building Theories in Software Engineering. In: *Guide to Advanced Empirical SE*, pp 312–336. Springer (2008)
5. Shull, F., Feldmann, R.L.: Building Theories from Multiple Evidence Sources. In: *Guide to Advanced Empirical Software Engineering*, pp. 337–364. Springer (2008)
6. Campbell, D.T., Stanley, J.C.: *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin Company (1963)
7. Cook, T., Campbell, D.: *The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings*. Rand McNally (1976)
8. Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates (1988)
9. Singer, J.: Using the American Psychological Association (APA) Style Guidelines to Report Experimental Results. In: *Proc. of the Fifth IEEE Workshop on Empirical Studies of Software Maintenance (WESS '99)*, pp. 71–75. Oxford (1999)
10. Jedlitschka, A., Pfahl, D.: Reporting Guidelines for Controlled Experiments in Software Engineering. In: *ACM/IEEE Int. Symp. on Empirical SE* (2005)
11. Basili, V., Selby, R., Hutchens, D.: Experimentation in Software Engineering. *IEEE Trans. on SE*. vol. 12, pp. 733–743. IEEE Computer Society Press (1986)
12. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: *Experimentation in Software Engineering: An Introduction*. Kluwer Acad. Publishers (1999)
13. Juristo, N., Moreno, A.M.: *Basics of Software Engineering Experimentation*. Kluwer Academic Publishers (2001)
14. Pfleeger, S.L.: Soup or Art? The Role of Evidential Force in Empirical Software Engineering. *IEEE Software*, vol. 22, pp. 66–73. IEEE Computer Society Press (2005)
15. Miller, J.: Can Results from Software Engineering Experiments be Safely Combined?. In: *Proc. of the 6th Int. Symp. on Software Metrics (METRICS '99)*, pp. 152. IEEE Computer Society Press (1999)
16. Miller, J.: Applying Meta-Analytical Procedures to Software Engineering Experiments. *J. of Systems and Software*, vol. 54, pp. 29–39. Elsevier Science Inc. (2000)
17. Hayes, W.: Research Synthesis in Software Engineering: A Case for Meta-Analysis. In: *Proc. of the 6th Int. Symp. on Software Metrics (METRICS '99)*, pp. 143. IEEE Computer Society Press (1999)
18. Hannay, J., Dybå, T., Arisholm, E., Sjøberg, D.: The Effectiveness of Pair Programming: A Meta-Analysis. *Information and Software Technology, Special Section: Software Engineering for Secure Systems*, vol. 51, pp. 1110–1122. Elsevier (2009)
19. Miller, J.: Replicating Software Engineering Experiments: A Poisoned Chalice or the Holy Grail. *Information and Software Technology*, vol. 47, pp. 233–244 (2005)
20. Miller, J.: Triangulation as a Basis for Knowledge Discovery in Software Engineering. *J. of Empirical SE*, vol. 13, pp. 223–228. Kluwer Academic Publishers (2008)
21. Basili, V., Shull, F., Lanubile, F.: Building Knowledge Through Families of Experiments. *IEEE Trans. on SE*, vol. 25, pp. 456–473. IEEE Computer Society Press (1999)
22. Shull, F., Carver, J., Vegas, S., Juristo, N.: The Role of Replications in Empirical Software Engineering. *J. of Empirical SE* vol. 13, pp. 211–218. Kluwer Academic Publishers (2008)
23. Kitchenham, B.: The Role of Replications in Empirical Software Engineering a word of warning. *J. of Empirical SE* vol. 13, pp. 219–221. Kluwer Academic Publishers (2008)