

# PROCESO DE AGREGACIÓN PARA ESTUDIOS EXPERIMENTALES DE INFORMATICA

Enrique Fernández<sup>1,2</sup>, Patricia Pesado<sup>1,3</sup>, Ramón García-Martínez<sup>1,4</sup>, Oscar Dieste<sup>1,5</sup>

<sup>1</sup> Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. UNLP

<sup>2</sup> Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA

<sup>3</sup> Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP - CIC

<sup>4</sup> Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA

<sup>5</sup> Grupo de Ingeniería de Software Experimental. Facultad de Informática. UPM

enfernan@itba.edu.ar, ppesado@lidi.info.unlp.edu.ar, rgarciamar@fi.uba.ar

## CONTEXTO

El Proyecto articula líneas de trabajo de los siguientes Proyectos: [a] “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaria de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires; y [b] “Caracterización de Comportamientos Sociales Mediante Métodos de Explotación de Información” con financiamiento solicitado al Programa Ibero Americano de Ciencia y Tecnología para el Desarrollo. CYTED 2008-2011 / P608RT0404 a radicar en el Grupo de Ingeniería de Software Experimental de la Facultad de Informática de la Universidad Politécnica de Madrid.

## RESUMEN

El incremento en la cantidad de estudios experimentales que se desarrollan en Informática ha traído aparejado un nuevo problema, que hacer cuando dos o más estudios que analizan el mismo tema no dan el mismo resultado?, este hecho puede solucionarse mediante el desarrollo de una síntesis cuantitativa de resultados. Este proyecto propone desarrollar un procedimiento de síntesis cuantitativa acorde a las características que actualmente presenta la Informática (pocos estudios, en general, hechos con pocos sujetos experimentales), el cual dista bastante de los propuestos hasta la fecha que han sido tomados de la Medicina.

**Palabras clave:** *Síntesis Cuantitativa, Agregación, Meta-análisis*

## 1. INTRODUCCION

Desde hace varios años, la cantidad de estudios experimentales en Informática se ha incrementado significativamente. Estos experimentos abarcan los más variados temas, como ser: el desempeño de las técnicas de prueba [Sabaliauskaite, *et al*, 2004] o la educación de requisitos [Browne y Rogich, 2001]

entre otros. Si bien los experimentos aportan conocimientos interesantes en cada caso, en general son pequeños [Guerini, 2008], por ello para que la información que aportan sea valiosa los resultados deben agregarse para poder obtener conclusiones avaladas con la mayor evidencia empírica posibles.

La agregación de experimentos consiste en combinar los resultados de varios experimentos que analizan el comportamiento de un par de tratamientos específico, en un contexto determinado, para obtener un *único resultado final*. El nuevo resultado será *más general y fiable que los resultados individuales*, por que el mismo estará sustentado por un mayor nivel de evidencia empírica [Cochrane, 2008].

Es importante destacar que para que los resultados de un proceso de agregación sea realmente fiable, el mismo debe evitar caer en problemas de sesgos vinculados a la búsqueda y selección de experimentos. Para ello, en general, los procesos de agregación de resultados se incluyen dentro de una Revisión Sistemática (RS). Una SR es un procedimiento que aplica estrategias científicas para aumentar la fiabilidad del proceso de recopilación, valoración crítica y agregación de los estudios experimentales relevantes sobre un tema [Goodman, 1996].

En la Informática las RS fueron propuestas por [Basili, *et al*, 1996], desde entonteces varios autores abordaron el tema. A modo de ejemplo podemos citar al procedimiento de revisión sistemática desarrollado por [Kitchenham, 2004], el cual surge de una adaptación de los procesos de revisión sistemática desarrollados en medicina, donde se contemplan, entre otras cosas, diversos niveles de calidad de experimento acorde al contexto experimental de la Informática. En dicho procedimiento, se recomienda el uso del método estadístico Diferencia Medias Ponderadas (DMP) [Hedges y Olkin, 1985] para la agregación de los resultados de los experimentos.

Además de Basili y Kitchenham, existen otros autores que han trabajado en la aplicación de las revisiones sistemáticas en Informática, a modo de ejemplo podemos citar la revisión sistemática desarrollada recientemente por [Dyba, *et al*, 2007], donde se identifican 11 estudios experimentales

vinculados al desempeño de los programadores cuando trabajan de a pares, los cuales se Agregaron aplicando el método de Diferencias Medias Ponderadas como se sugiere en [Kitchenham, 2004]. Ahora bien, no todas las RS hechas dentro del ámbito de la Informática tuvieron éxito, a modo de ejemplo podemos citar a: [Banker y Keremer, 1989; Shull, *et al*, 2003; Hu, 1997; Wohlin, *et al*, 2003; Juristo, *et al*, 2004; Jørgensen, 2004] donde, si bien los autores pudieron desarrollar el procedimiento de búsqueda y selección de experimentos, la combinación de los mismos mediante Diferencias Medias Ponderadas resultó impracticable. Los principales motivos fueron:

- Escasez de experimentos, replicaciones y homogeneidad entre los mismos. [Davis, *et al*, 2006; Miller, 2000].
- Carencia de estándares para reportes de experimentos. Por ejemplo, [Burton, *et al*, 1990] no publican varianzas y [Crandall Klein, *et al*, 1989] ni siquiera reporta las medias de los resultados experimentales.
- Falta de estandarización de las variables respuesta. Por ejemplo, los trabajos de [Agarwal, *et al*, 1990] y [Woody, *et al*, 1996] utilizan diferentes variables respuesta para analizar un mismo aspecto, lo cual hace que estos experimentos no puedan ser agregado.

Por lo expuesto, surge la necesidad de desarrollar un procedimiento de agregación de estudios experimentales acorde a las características del actual contexto experimental de la Informática.

## 2. LINEAS DE INVESTIGACION y DESARROLLO

En el marco de este proyecto se investigará:

- [a] Las características de los estudios experimentales hechos en Informática.
- [b] Diferentes técnicas y estrategias de agregación usadas en otras áreas de la ciencia.
- [c] La fiabilidad de las distintas técnicas de agregación identificada dentro de los parámetros estándar de la Informática.

La aplicación de una estrategia de agregación para el dominio Informático que considere el uso de las técnicas de agregación de forma conjunta y combinada en lugar de aplicarlas de forma excluyente y aislada.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto tiene como objetivo general definir una estrategia de agregación de estudios experimentales en Informática que permita combinar los resultados de un conjunto estudios experimentales, para obtener piezas de conocimientos representativas a nivel general.

El proyecto se enfocará a construir una estrategia de agregación que complemente a los protocolos de Revisiones Sistemáticas para Informática (por ejemplo: [Kitchenham, *et al*, 2004]), el cual contemplará el uso combinado de varias técnicas de agregación. Esta combinación de técnicas permitirá generar conclusiones con diversos niveles de fiabilidad, los cuales estarán relacionados con la fiabilidad de la técnica y las características de los estudios combinados mediante las mismas. Se considera que el proyecto implica investigación básica en la definición del proceso de agregación para Informática.

Entre las cuestiones particulares que se abordarán en el proyecto se encuentran: la caracterización de los estudios experimentales hechos en el ámbito de la Informática, la identificación de los problema que habitualmente suceden cuando se desarrollan revisiones sistemáticas dentro del ámbito de la Informática, la identificación de técnicas de agregación aplicadas en otras áreas de la ciencia, la determinación de la fiabilidad de las distintas técnicas de agregación en el contexto de la experimentación en Informática, y el establecimiento de un proceso de agregación con varios niveles de evidencia acorde a las características de los experimentos hechos en Informática.

Se espera que este proyecto contribuya a solucionar el problema de agregación de estudios experimentales que actualmente sufren las revisiones sistemáticas en informática, permitiendo de esta forma que las piezas de conocimientos generadas empíricamente puedan ser consideradas de aplicación general como sucede en las ciencias mas maduras.

A la fecha se ha validado la necesidad de desarrollar un procedimiento [Fernández; 2007], se ha presentado una versión preliminar del procedimiento en [Fernández; *et al*; 2007], se han desarrollado trabajos vinculado al análisis de la calidad experimental de los trabajos hechos en Informática en [Guerini; *et al.*, 2006; Guerini, 2008], se está evaluando la fiabilidad de diferentes métodos de síntesis cuantitativa [Pollo Cattaneo; 2009] para determinar en que condiciones aplicar cada uno, se ha desarrollado un trabajo de validación del procedimiento de agregación [Malcrida; 2009] mediante el agregado de 25 experimentos vinculados al análisis de métodos de inspección de código. Se está trabajando en el desarrollo de un aplicativo que permita seleccionar de forma automática estudios experimentales [Caimer, 2009] el cual será validado mediante el análisis de estudios vinculados a patrones de diseño, los cuales serán agregados en [González, 2008]. Por último, se está trabajando en una tesis doctoral con el objeto de describir en detalle las pautas para el agregado de experimentos en Informática [Fernández; 2008].

#### 4. FORMACION DE RECURSOS HUMANOS

En el marco de este proyecto se esta desarrollando una tesis doctoral y cuatro tesis de maestria (una de ellas defendida y otra a la espera de defensa). Esta línea vincula al LIDI de la Facultad de Informática de la Universidad Nacional de La Plata, al Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires y al Grupo de Ingeniería de Software Experimental de la Facultad de Informática de la Universidad Politécnica de Madrid.

#### 5. BIBLIOGRAFIA

- Agarwal, R.; Tanniru, M.; 1990; *Knowledge Acquisition Using Structured Interviewing: An Empirical Investigation*; Journal of Management Information System, M.E. Sharpe; Vol. 7 N. 1
- Banker and Keremer; 1989; *Scale economies in new software development*. IEEE Transactions on Software Engineering. (15): 10, pp. 1199-1205.
- Basili, V.; Green, S.; Laitenberger, O.; Lanubile, F.; Shull, F.; Sorumgaard, S.; Zelkowitz, M.; 1996; Packaging researcher experience to assist replication of experiments ISERN Meeting
- Browne, G.; Rogich, M.; *An empirical Investigation of User Requirements Elicitation: Comparing the Effectiveness of Prompting Techniques*; Journal of management Information System; Spring 2001; Vol. 17 N. 4
- Burton, A., Shadbolt, N., Rugg, G. y Hedgecock, A.; 1990. *The Efficacy of Knowledge Elicitation Techniques: A Comparison Across Domains and Level of Expertise*. Knowledge Acquisition 2(2): 167-178.
- Caimer, P; 2009. *Uso de minería de texto para identificación de trabajos experimentales*. Proyecto de tesis de Maestría, ITBA.
- Cochrane; 2008; *Curso Avanzado de Revisiones Sistemáticas*; [www.cochrane.es/?q=es/node/198](http://www.cochrane.es/?q=es/node/198)
- Crandall Klein, B. y Asociados; 1989. *A Comparative Study Of Think-Aloud And Critical Decision Knowledge Elicitation Method*. SIGAR Newsletter, April 1989, Number 108, Knowledge Acquisition Special Issue, páginas 144-146.
- Davis, A.; Dieste o.; Hickey, A.; Juristo, N.; Moreno, A.; 2006; *Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review*; 14th IEEE International Requirements Engineering Conference (RE'06) pp. 179-188
- Dyba, T., Aricholm, E.; Sjoberg, D.; Hannay J.; Shull, F.; 2007; *Are two heads better than one? On the effectiveness of pair programming*. IEEE Software;12-15.
- Fernández, E. 2007. *Aggregation Process with Multiple Evidence Levels for Experimental Studies in Software Engineering*. Proceedings 2nd International Doctoral Symposium on Empirical Software Engineering. Pag. 75-81. ISBN 978-84-690-7340-7.
- Fernández, E., Britos, P., Dieste, O., García-Martínez, R. 2007. *Proceso de agregación con múltiples niveles de evidencia para estudios experimentales en Informática*. Proceedings IX Workshop de Investigadores en Ciencias de la Computación. Pag. 487-491. ISBN 978-950-763-075-0.
- Fernández, E. 2008. *Desarrollo de un proceso de agregación para estudios experimentales de informatica*. Proyecto de Tesis Doctoral en Ciencias Informáticas, Universidad Nacional de La Plata.
- Glass, G; 1976; *Primary, secondary, and meta-analysis of research*. Educational Researcher 5: 3-8
- González, S; 2008. *Revisión y agregación de estudios experimentales vinculados a patrones de diseño*. Proyecto de tesis de Maestría, ITBA.
- Goodman C.; 1996; *Literature Searching and Evidence Interpretation for Assessing Health Care Practices*; SBU; Stockholm.
- Guerini, M. 2008. *Revisión de Resultados Experimentales en Técnicas de Prueba y de Educción de Conocimientos*. Tesis de Magister en Ingeniería del Software. ITBA
- Guerini, M., Fernández, E., Britos, P., García-Martínez, R. 2008. *Revisión de Resultados Experimentales en Técnicas de Prueba y de Educción de Conocimientos*. Proceedings X Workshop de Investigadores en Ciencias de la Computación. Pág. 471-475. ISBN 978-950-863-101-5.
- Guerini, M., Fernández, E., Ochoa, M., Merlino, H., Diez, E., Britos, P. y García-Martínez, R. 2006. *Revisión de Resultados Experimentales sobre Performance de Técnicas Pruebas de Software*. XII Congreso Argentino de Ciencias de la Computación. San Luis. Argentina
- Hedges, L.; Olkin, I.; 1985; *Statistical methods for meta-analysis*. Academic Press.
- Hu, Q.; 1997; *Evaluating Alternative Software Production Function*. IEEE Transactions on Software Engineering. (23): 6, pp. 379-387.
- Jørgensen, M.; 2004; *A Review of Studies on Expert Estimation of Software Development Effort*. Journal of Systems and Software. (70): 1-2, pp. 37-60.
- Juristo, N.; Moreno, A. M., and Vegas, S.; 2004; *Reviewing 25 Years of Testing Technique Experiments*. Journal of Empirical Software Engineering; 9(1 - 2):7-44.
- Kitchenham, B. A.; 2004; *Procedures for performing systematic reviews*. Keele University; TR/SE-0401. Keele University Technical Report.
- Malacrida, J; 2009. *Revisión y agregación de estudios experimentales vinculados Técnicas de Inspección*. Tesis de Maestría pendiente de defensa, ITBA.

- Miller, J.; 2000; *Applying Meta-analytical Procedures to Software Engineering Experiments*. Journal of Systems and Software. (54): 1, pp. 29-39.
- Pollo Cattaneo, F; 2009. *Análisis de precisión de técnicas de agregación en contextos experimentales poco maduros*. Tesis de Maestría pendiente de defensa, ITBA.
- Sabaliauskaite, G., Kusumoto, S., Inoue, K.; 2004; *Assessing defect detection performance of interacting teams in object-oriented design inspection*; Information and Software Technology 46 (2004) 875–886; Available online at: [www.sciencedirect.com](http://www.sciencedirect.com)
- Shull, F.; Carver, J.; Travassos, G. H.; Maldonado, J. C.; Conradi, R., and Basili, V. R.; 2003; *Replicated Studies: Building a Body of Knowledge about Software Reading Techniques*. Lecture Notes on Empirical Software Engineering. Chapter 2, pp. 39-84. World Scientific.
- Thelin, T., Andersson, C., Runeson, P., Dzamashvili-Fogelström, N.; 2004; *A Replicated Experiment of Usage-Based and Checklist-Based Reading*; Proceedings of the 10th International Symposium on Software Metrics (METRICS'04)
- Wohlin, C., Petersson, H., & Aurum, A.; 2003; *Combining data from reading experiments in software inspections: a feasibility study*. (pp. 85-132). World Scientific Publishing Co., Inc.
- Woody, J.; Will, R.; Blanton, J.; 1996; *Enhancing Knowledge Elicitation using the Cognitive Interview*; Expert system with application; Vol. 10 N. 1