

PROCESO DE AGREGACIÓN CON MÚLTIPLES NIVELES DE EVIDENCIA PARA ESTUDIOS EXPERIMENTALES EN INGENIERÍA DEL SOFTWARE

Enrique Fernández^{1,2}, Paola Britos^{1,2}, Ramón García-Martínez^{1,2}, Oscar Dieste³

¹Centro de Ingeniería de Software e Ingeniería del Conocimiento. Escuela de Postgrado. ITBA.
Argentina

²Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. Universidad de Buenos Aires.
Argentina

³Grupo de Investigación en Experimentación de Software. Facultad de Informática. Universidad
Politécnica de Madrid.
España
rgm@itba.edu.ar

Abstract

Current process' of aggregation of result about experimental studies, based on meta analysis from other sciences, showed that they are not suitable for software engineering. This article adds an alternative aggregation process to the standard one, based on an aggregation strategy with multiple levels of evidence; in which each level is connected with a specific aggregation technique that is assigned considering the quality and quantity of experimental studies identified.

Keywords: Aggregation with multiple levels of evidence, meta analysis

Resumen

Los actuales procedimientos de agregación de resultados de estudios experimentales basados en Meta-Análisis, provenientes de otras ramas de la ciencia, han mostrado no ser adecuados para la realidad que vive la Ingeniería del Software. El presente trabajo presenta un proceso de agregación alternativo al estándar basado en una estrategia de agregación con múltiples niveles de evidencia. Donde cada uno de los niveles de evidencia se encuentra vinculado con una técnica de agregación específica que se asigna en función de la calidad y cantidad de estudios experimentales identificados.

Palabras claves: Agregación con múltiples niveles de evidencia, meta-análisis

1. Introducción

Desde hace varios años, la cantidad de experimentos realizados dentro del ámbito de la Ingeniería del Software (SE) se ha incrementado significativamente. Estos experimentos abarcan los más variados temas, como ser: el desempeño de las técnicas de prueba, la educación de requisitos, o el desempeño de los lenguajes de programación, entre otros. Si bien los experimentos aportan conocimientos interesantes en cada caso, para que la información que aportan sea valiosa los resultados deben agregarse para poder obtener conclusiones avaladas con el mayor número de estudios posibles.

Ha habido algunos intentos de síntesis de experimentos para IS, por ejemplo [1], [2], [3], [4], [5], [6]. Pero todos estos esfuerzos no han dado los resultados esperados. También los resultados de las combinaciones informales [2], [4], [5], [6] fueron limitados. Por otra parte los intentos de combinaciones con técnicas estadísticas [1], [3] resultan impracticables debido a las diferencias en el diseño y ejecución de los experimentos realizados.

Recientemente, se ha propuesto la Revisión Sistemática (RS) [7], [8] como un método para sistematizar la agregación de estudios empíricos en IS.

Una Revisión Sistemática 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 [9]. Las RS se han comenzado a emplear recientemente en IS [10], [11]. No obstante, si bien las RS proporcionan un marco de trabajo que permite realizar la recopilación de experimentos y, en menor medida, la valoración crítica de los mismos, esta falla a la hora de agregar los resultados. El motivo de la falla reside en que la estrategia de agregación utilizada por las RS es Meta-Análisis. El Meta-Análisis es un nombre colectivo que hace referencia a un conjunto de métodos estadísticos que permiten combinar resultados experimentales, siempre y cuando se verifiquen ciertas restricciones, tales como un número mínimo de experimentos, adecuadamente recopilados y homogéneos [12]. Estas restricciones distan de cumplirse en IS. Así por ejemplo: en [11] no puede aplicarse Meta-Análisis debido a la heterogeneidad entre los experimentos y en el caso del problema fue la carencia de replicaciones.

Por consiguiente es necesario desarrollar un procedimiento de agregación específico que considere las características particulares de un entorno de investigación poco maduro como es la IS, de modo que sea posible la aplicación de dicho procedimiento: con un bajo número de replicaciones y con heterogeneidad de estudios, entre otras características. Este procedimiento permitirá mejorar el nivel de las conclusiones obtenidas sobre el desempeño de técnicas y herramientas que a diario se utilizan en IS, muchas veces adoptadas o adquiridas por la fama de quienes las patrocinan. Dando así a la IS un entorno de desarrollo propio de la Ingeniería.

La sección 2 de este artículo describe el estado del arte; la sección 3 describe la problemática identificada en esta etapa de la investigación; la sección 4 define la importancia de problema tratado; la sección 5 establece los materiales y métodos a ser utilizados para el logro de los objetivos; la sección 6 presenta una introducción a la primer versión de la propuesta de solución; la sección 7 describe cual será la estrategia de validación de los resultados. Finalmente, la sección 8 describe algunas de las conclusiones obtenidas hasta el momento.

2. Estado de la Cuestión

La agregación de estudios experimentales tiene una larga historia en disciplinas como la educación o la psicología [13], aunque últimamente su desarrollo está impulsado por las ciencias de la salud [14].

Para agregar los resultados de estudios experimentales, existen dos tipos de métodos: los interpretativos y los no interpretativos [15]. Los métodos interpretativos, tales como “Narrative Summary” [16] o “Grounded Theory” [17], se caracterizan por que las conclusiones son generadas en base a los criterios personales de quienes analizan los resultados [18]. Como estos métodos arrojan resultados poco fiables por su alta dependencia respecto del revisor, su uso ha venido decayendo a favor de los más fiables métodos no interpretativos.

Dentro de los métodos no interpretativos existen varias alternativas, tales como: “Case Surveys” [19], “Vote Counting” [12] o “Análisis Comparativo” [20]. No obstante el más sofisticado de todos los métodos es el Meta-Análisis [21]. Según [22] el Meta-Análisis es el análisis estadístico de una serie de estudios individuales, con el objeto de integrar los resultados en una medida resumen.

Al realizar un Meta-Análisis, deseamos hallar un resultado numérico que sea resumen representativo de los resultados de los estudios individuales, y por tanto que signifique una mejora sobre las estimaciones individuales. En la actualidad el meta-análisis se implementa mediante la estimación del Tamaño de Efecto a través de las Diferencias Medias Ponderadas [12], la cual es conceptualmente sencilla: el estimador de efecto global se calcula como una media ponderada de los estimadores de efecto de los estudios individuales.

Para que el Meta-Análisis arroje resultados representativos de los estudios incluidos, es necesario validar que los estudios individuales puedan resumirse y puedan combinarse. Esta propiedad se conoce con el nombre de “homogeneidad” y se determina a través de la “heterogeneidad estadística”, la cual permite identificar si la variación en los resultados de los estudios, es debido a un error aleatorio o no.

Otro aspecto a verificar cuando se desarrolla un Meta-Análisis, es determinar cuan dependiente de los estudios mas robustos es el resultado obtenido, lo que puede estimarse a través de un “análisis de sensibilidad”. Este análisis permite ver si todos los estudios fueron tenidos en cuenta en el resultado final obtenido o no.

Diversos autores [7], [8] propugnan la utilización del Meta-Análisis en IS. Sin embargo, como se mencionó mas arriba, hoy día es prácticamente imposible aplicar una estrategia de agregación mediante Meta-Análisis dentro del ámbito de la IS. Los principales motivos que impiden su aplicación son:

- Escasez de experimentos, replicaciones y homogeneidad entre los mismos. [7], [10].
- Carencia de estándares para reportes de experimentos. Por ejemplo, [23] no publican varianzas y [24] ni siquiera reporta las medias de los resultados experimentales. En estas circunstancias la aplicación del Meta-Análisis es imposible.
- Calidad interna variada. Por ejemplo, tratando un mismo tema de investigación, los trabajos de [24] y [25] muestran una gran discrepancia en la concepción y armado del estudio, esto hace que los mismos no pueden considerarse replicaciones y por tanto no puedan utilizarse para un proceso de agregación por Meta-Análisis, es mas, si se hiciese el resultado sería invalidado por el análisis de heterogeneidad.
- Falta de estandarización de las variables respuesta. Por ejemplo, los trabajos de [26] y [27] utilizan diferentes variables respuesta para analizar un mismo aspecto, lo cual hace que estos experimentos no puedan ser considerados replicaciones.

Además del Meta-Análisis existen otras técnicas alternativas menos sofisticadas y que poseen menos restricciones para su aplicación. Dentro de este grupo podemos encontrar: “Case Surveys” [19], “Vote-Counting” [12] y “Análisis Comparativo” [20]. A diferencia del Meta-Análisis, el cual

ha sido estudiado en profundidad, no se han determinado los límites de aplicabilidad de dichas técnicas, con la única excepción del “Vote Counting” [12], y su aplicación en la IS ha sido testimonial hasta la fecha [11], [28]. Por consiguiente, estas técnicas, aunque prometedoras, deben estudiarse extensivamente antes de ser utilizadas rutinariamente en la IS.

3. Descripción del Problema

Actualmente, no existe un método de agregación específicamente adaptado a las necesidades de la IS. Por su parte, los métodos más fiables, como el Meta-Análisis, poseen restricciones que limitan su aplicabilidad. Además, existe un conjunto de métodos que podrían ser aplicables, tales como “Case Survey” o “Análisis Comparativo”, pero estos no han sido nunca utilizados en IS y, por lo tanto, se desconoce su idoneidad para realizar agregaciones. En consecuencia, el presente trabajo tiene como objetivo desarrollar un método de agregación que permita obtener la mayor cantidad de piezas de conocimiento posibles, combinando la mayor cantidad de estudios, independientemente de su calidad. El mismo deberá poder trabajar dentro de las limitaciones propias de la IS, esto es: poca cantidad de estudios que traten la misma variable respuesta o tratamiento; falta de estandarización de las variables respuestas; y el reporte de pocas variables estadísticas.

4. Importancia del Problema

Desarrollar experimentos en diferentes lugares (laboratorios e industrias) genera resultados parciales respecto de las capacidades y condiciones de aplicación de las técnicas. Si por ejemplo todos los estudios realizados mostraran que la técnica “a” es mejor que la técnica “b” el proceso de agregación sería muy simple, pero en la realidad esto no es muy habitual, en general los resultados son dispares, con experimentos que muestran a la técnica “a” mejor que la técnica “b” y otros que muestran lo contrario.

El entorno en cual se desarrollaron los experimentos hace que la fiabilidad de los resultados obtenidos varíen de un estudio a otro, estos aspectos deben tenerse en cuenta a la hora de combinar los resultados para que la conclusión sea lo mas fiable posible. Por lo cual para obtener piezas de conocimiento fiables, se debería contar con una estrategia de agregación adecuada a las actuales características de la IS.

5. Materiales y método

Para producir el conocimiento necesarios seguiremos un enfoque clásico [29], [30], [31] identificando los materiales y métodos necesarios para desarrollar nuestro experimento:

Materiales:

Para el desarrollo del método de agregación utilizaremos tres tipos de materiales:

- a) Técnicas de agregación en SE: En la actualidad se utilizan en IS técnicas de Meta-Análisis [7], [8], así como técnicas mas informales [4], [5], [10], estas técnicas serán revisadas y, probablemente, utilizadas durante la presente investigación.

- b) Técnicas de Agregación en otras disciplinas: Además de las técnicas de agregación utilizadas actualmente dentro del ámbito de la IS, se estudiarán otras técnicas no utilizadas hasta el momento dentro de esta disciplina, como son el “Response Ratio” [31] o “Case Surveys” [19] por citar dos ejemplos.

- c) Estudios Experimentales: En la actualidad se cuenta con una gran cantidad de estudios experimentales de las más variadas características. Los mismos se utilizarán para definir las técnicas de agregación (estas deberán poder paliar las falencias propias de los estudios) y validar la factibilidad de uso de las mismas.

Método:

Las tareas a realizar para desarrollar el presente proceso de agregación serán las siguientes:

- Se buscará identificar técnicas de agregación alternativas al “Effect Size” a través de revisión bibliográfica y consulta a expertos.
- Se analizará las condiciones de aplicación de las técnicas de agregación encontradas, es decir, bajo que supuestos la técnica es aplicable. Por ejemplo: si la cantidad de sujetos experimentales reportados en cada estudio debe ser el mismo o puede variar de un estudio a otro.
- Se analizará la fiabilidad de la respuesta estimada por las distintas técnicas de agregación encontradas, es decir, cual es el nivel de error esperado.
- Se propondrá un método para la aplicación de las técnicas de agregación en función de su fiabilidad y restricciones de aplicación
- Se propondrá una estrategia para interpretar los resultados obtenidos por las distintas técnicas, la misma estará vinculada con la fiabilidad de la respuesta brindada por la técnica de agregación.

6. Solución Propuesta

Para solucionar el problema de agregación de estudios, se propone una estrategia de agregación múltinivel, en la cual, las técnicas de agregación no se utilicen de forma alternativa y excluyente entre si, sino, que se utilicen de forma complementaria. Así, en función de la cantidad y calidad de estudios encontrados, se aplicarán las técnicas más convenientes y los resultados se analizarán de manera conjunta. Por ejemplo, si se contase con diez experimentos, cuatro agregables por Meta-Análisis y seis que no, para los experimentos no agregables por Meta-Análisis se utilizará alguna técnica de agregación alternativa, y luego los resultados se mostrarán de forma conjunta con los resultados obtenidos por Meta-Análisis. De esta forma se obtendrá más de un nivel de evidencia, el primero vinculado al Meta-Análisis y los siguientes vinculados a las técnicas alternativas que serán menos fiables.

La estrategia de agregación propuesta se estructura en 5 pasos, tal como puede observarse en la figura 1.

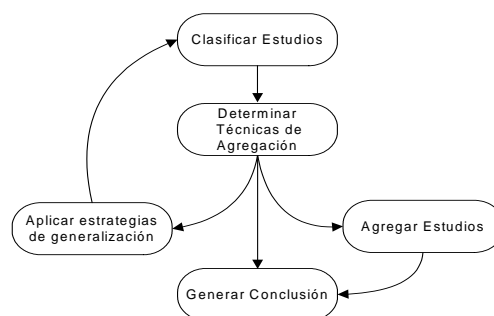


Figura 1. Ciclo de Vida del proceso de Agregación

Clasificar Estudios: La clasificación de estudios tiene como objetivo poder agrupar los distintos estudios en función de su calidad, las variables respuestas que publiquen y los tipos de tratamientos analizados.

Determinar las Técnicas de Agregación: Este paso tiene como objetivo identificar que técnicas de agregación conviene aplicar en función de la cantidad y calidad de estudios encontrados. En caso de no poder aplicar ninguna técnica de agregación, se podrá recomendar realizar algún paso alternativo: “Aplicar Estrategias de Generalización” o “Generar conclusiones”.

Aplicar Estrategias de Generalización: La estrategia de Generalización tiene como objetivo paliar los problemas vinculados al bajo número de replicaciones. Para ello se debe buscar características comunes, entre dichos estudios, que permitan agrupar tratamientos y/o variables respuesta dentro de un grupo de mayor nivel de abstracción (más general). Si bien estos grupos de mayor nivel de abstracción no son verdaderas replicaciones, su similitud permite considerarlos como replicaciones conceptuales y, por lo tanto, como estudios agregables.

Agregar Estudios: En este paso se aplicarán las distintas técnicas de agregación que permiten combinar los resultados de los estudios experimentales. Esto se hace en base a los criterios y recomendaciones hechos en el paso "Determinar Técnica de Agregación".

Generar Conclusión: El objetivo de esta paso es generar un informe (lo mas fiable posible) con estas piezas de conocimiento. Donde el análisis de resultados comenzará desde los más fiables (obtenidos mediante Meta-Análisis) a los menos fiables (obtenidos mediante técnicas alternativas). De esta forma si los resultados son compatibles (todos los niveles de evidencia afirma que un tratamiento es mejor que el otro) se habrá obtenido una conclusión mas robusta que la que se obtendría aplicando la técnicas en forma aislada. Pero si los resultados no son compatibles, se deberá intentar determinar la existencia de variables aleatorias no identificadas hasta el momento o plantear la necesidad de generar de nuevos experimentos vinculados al tema.

Las siguientes subsecciones describen en detalle cada uno de los pasos antes mencionados:

6.1. Clasificar Estudios

En esta investigación, se van a utilizar un conjunto de técnicas estadísticas que, para asegurar la fiabilidad de la respuesta que estiman, requieren, entre otras cosas, que los estudios empíricos incluidos cumplan con un conjunto de precondiciones [11]. Dichas precondiciones son las siguientes:

- **Calidad del Estudio Empírico:** grado que define en que medida el estudio ha sido bien diseñado, ejecutado y analizado. Esto permite estimar cuan fiables son los resultados expresados en el estudio. Cuando existen riesgos de que un estudio sea poco fiable, es necesario decidir si debe formar parte de la agregación, y en caso afirmativo, determinar que técnicas de agregación se van a utilizar. Si bien la determinación de la calidad de estudios empíricos es un tema resuelto en la mayoría de las ciencias, esto no esta totalmente definido dentro del ámbito de la IS, por ello para esta primer versión del proceso de agregación se utilizaran las recomendaciones hechas en [8], pretendiendo aplicar en futuras versiones las conclusiones alcanzadas en el proyecto de investigación que se esta llevando a cabo por [33].
- **Compleitud del Reporte:** Este es un segundo aspecto a ser analizado, ya que por bien construido que esté el estudio, si el reporte no refleja un conjunto mínimo de parámetros, las técnicas de agregación no podrán ser aplicadas. Los parámetros más relevantes son: Medias (M), Varianzas (V) y cantidad de sujetos experimentales (N). En caso de no publicarse las “varianzas” se debe identificar si se han publicado los estadísticos: parámetro T de Student o F de Snedecor. Asimismo, en caso de producirse la ausencia de las medias, puede ser un paliativo saber si existió o no diferencias entre las mismas.
- **Representatividad de los Tratamientos y Variables Respuesta:** como en la actualidad existen pocas replicaciones de estudios, el presente proceso de agregación propone la aplicación de una estrategia de generalización (ver paso "Aplicar Estrategia de Generalización") que permita

incluir dentro de un mismo grupo a tratamientos que no son iguales, pero sus similitudes son mayores que sus diferencias. No obstante, las diferencias entre estos estudios generalizados hacen no aplicable la estimación por “Diferencias Medias Ponderadas” o “Response Ratio”. En suma: la generalización de tratamientos limita el tipo de técnica de agregación a utilizar. Lo mismo ocurre con la Representatividad de las Variables Respuesta.

- En base al análisis de los aspectos antes mencionados, se deberá asignar una categoría a cada estudio empírico, esta categoría determina que técnicas de agregación pueden aplicarse al mismo. En la tabla 1, se describen las características principales de los estudios incluidos en cada categoría, así como la técnica que, por el momento, consideramos aplicable a la misma categoría.

La asignación de estudios a categorías puede hacer de modo completamente determinista utilizando la tabla 2. Esta tabla es una primera aproximación a la clasificación de estudios empíricos, y previsiblemente podrá ser actualizada. Sin ir mas lejos, a la tabla 2 se le deberán agregar los resultados sobre calidad de estudios que actualmente investiga [31].

Una vez categorizados los estudios, es necesario hacer una descomposición adicional en función de las variables respuesta que los estudios utilizan. Por ejemplo: Supongamos que se desea determinar cual de dos técnicas de educación llamadas “A” y “B” es mejor. Los estudios empíricos que ensayan dichas técnicas pueden utilizar diversas variables respuestas como el tiempo medio por sesión y la cantidad de conocimiento adquirido. Como estas variables no son compatibles entre si, debemos descomponer el conjunto de estudios disponibles en: “Técnica A vs. Técnica B utilizando la variable respuesta Tiempo medio de sesión” y “Técnica A vs. Técnica B utilizando la variables respuesta cantidad de información”. A esta descomposición la llamaremos par Tratamiento-Variable, y guiarán el proceso de contabilización.

Categoría	Características de los Estudios	Técnica de agregación
1	Dentro de esta categoría se aceptan estudios que no posean sesgos y sean similares en cuanto a su confección y dominio de aplicación.	“Diferencias Medias Ponderadas”
2	Dentro de esta categoría se aceptan estudios que sean similares en cuanto a su confección y dominio de aplicación, pero que en lugar de publicar las varianzas, estimen si la diferencia entre medias es significativa o no y publiquen los estadísticos t o F.	“Diferencias Medias Ponderadas” mediante formulas alternativas
3	Dentro de esta categoría se aceptan estudios con defectos leves de reporte (no publican varianzas ni análisis de diferencia entre medias).	“Vote Counting”
4	Dentro de esta categoría se aceptan estudios con defectos graves de reporte (solo expresan diferencias entre medias o dicen que un tratamiento es mejor que otro).	“Conteo de Votos Directo”

Tabla 1. Descripción de la categoría de estudios

Condiciones	R1	R2	R3	R4
Calidad del Estudio Experimental	1-2-3	1-2-3	1-2-3	1-2-3-4-5
El Reporte publica	Medias, Varianzas y Sujetos	Medias, Sujetos y Estadísticos alternativos (t o	Medias y Sujetos	que tratamiento se desempeño mejor

Condiciones	R1	R2	R3	R4
		F)		
Tratamientos y Variables Respuesta	Ninguno fue generalizado	Ninguno fue generalizado	Ninguno fue generalizado	Uno o ambos fueron Generalizados
Acciones				
Asignar Categoría	1	2	3	4

Tabla 2. Tabla de decisión para determinar la categoría de los estudios

Una vez establecidos todos los pares Tratamiento-Variable, se deberá determinar la cantidad de estudios asignable por cada par.

6.2. Determinar Técnica de Agregación

Como se mencionó anteriormente, la presente tesis propone utilizar diversas técnicas de agregación de forma conjunta. Para ello, se debe analizar la cantidad de estudios vinculados a cada par Tratamiento-Variable. Esto se debe a que la precisión de las técnicas varía en función de la cantidad de estudios disponibles [34].

Si la cantidad de estudios de categoría 1 es superior a 10 es posible aplicar “Diferencias Medias Ponderadas” [12] para estimar el “Tamaño de Efecto” y dar por terminado el proceso ya que la fiabilidad de la respuesta será muy alta [344]. En caso contrario, si existen estudios de categoría 2, se deberán acumular los estudios de categoría 1 con los de categoría 2 y se recomendará agregarlos por las técnicas de estimación de efectos alternativas [35].

En este caso parece ser suficientes unos 10 artículos para dar por finalizado el proceso. En caso contrario, como las técnicas que vamos a utilizar a partir de ahora son menos precisas, siempre que se puedan aplicar, se recomendará su uso. Cuando se analicen los estudios de categoría 3, se verificará que existen estudios para esta categoría y, en caso de existir, se los acumulará con los de categoría 1 y 2 y se recomendará aplicar la técnica de “Vote Counting” [12]. Por último, si existen estudios de categoría 4, se acumularán los estudios de las cuatro categorías y se recomendará aplicar la técnica “Conteo de Votos Directo”.

Las técnicas indicadas son las estudiadas hasta el momento, no obstante, no se descarta la utilización de otras técnicas de agregación a medida que la investigación avance.

Categoría	Técnicas de Agregación
1	▪ “Diferencias Medias Ponderadas” [12]
2	▪ “Alternatives” [34]
3	▪ “Vote Counting” [12]
4	▪ “Conteo de Votos Directo”

Tabla 3. Relación entre las categorías y las técnicas de Agregación

6.3. Aplicar Estrategias de Generalización

La generalización tiene como objetivo poner de manifiesto los aspectos comunes, de mayor nivel de abstracción de dos tratamientos o variables respuesta. Con este proceso intentamos salvar los problemas de baja cantidad de replicaciones y falta de estandarización de variables respuesta que, habitualmente, sufren los experimentos hechos en IS. Un ejemplo de generalización: Supongamos que intentando determinar si el lenguaje “C++” es mejor o no que su antecesor el lenguaje “C” se contara solo con 2 estudios que comparen estos dos lenguajes en forma directa. Esta evidencia sería muy baja para llevar adelante un proceso de agregación que genere resultado fiables. Pero, si además de estos dos estudios, se contara con otros estudios que comparen “Delphi” y “Pascal”, como “C++” y “Delphi” son Orientados a Objetos y el “C” y el “Pascal” son estructurados, podemos conjeturar que “juntando” o generalizando “C++” y “Delphi” (así como “C” y “Pascal”) podremos obtener una conclusión con mayor fiabilidad, debido al mayor número de estudios disponibles. Obviamente, estas conclusiones no responden a la pregunta “es mejor C++ que C”, pero si a otra muy similar que genera conocimiento sobre la primera.

6.4. Agregar Estudios

En este paso se aplicarán las distintas técnicas de agregación que permiten combinar los resultados de los estudios experimentales. Esto se hará en base a los criterios y recomendaciones hechas en el paso Establecer Técnica de Agregación.

6.5. Generar conclusión

Si bien este paso se encuentra todavía en estado de desarrollo, podemos dar algunos detalles acerca del mismo:

Las distintas técnicas de agregación utilizadas durante los pasos anteriores del proceso de agregación arrojan distintos resultados con diferentes niveles de fiabilidad. Se hace necesario entonces analizar si los resultados obtenidos por cada una de las técnicas son coherentes entre si, cuales son los mas fiables y cuales son las discrepancias.

Por otra parte, el informe final debe contener dos apartados: “Piezas de Conocimientos obtenidas” y “Posibles Líneas de Investigación”. Dentro del primer apartado se describirán los conocimientos obtenidos para los cuales existe evidencia firme que los sustente. Mientras que en el segundo grupo se describirán las conjeturas hechas durante la interpretación de los datos y las preguntas de investigación que no han podido ser resueltas.

7. Estrategia de validación

La fiabilidad y versatilidad del presente método serán validadas mediante comparaciones con otros métodos de agregación. Esto se hará en dos etapas, una primera etapa de laboratorio, donde a un conjunto de datos generados de forma artificial se aplicaran un conjunto de procesos de agregación, para luego comparar los resultados, y una segunda etapa en la cual se buscarán Revisiones Sistemáticas reales en las cuales se haya podido aplicar Meta-Análisis, y se compararan los resultados arrojados por estas contra los resultados estimados mediante el nuevo proceso de agregación. Esta última validación se hará con estudios hechos probablemente en otras ramas de la ciencia, debido a las limitaciones que existen para aplicar un proceso de agregación estándar con los estudios hechos en IS.

8. Conclusiones y Futuras Líneas de Investigación

El trabajo realizado en esta primera etapa de la investigación revela la complejidad de la agregación de estudios empíricos, así como la cantidad de trabajo pendiente. No obstante, a pesar de esto, el objetivo planteado, desarrollar un proceso de Agregación especialmente adaptado para IS, se demuestra factible.

Respecto de los siguientes pasos en la investigación, en este momento se estima prioritario seguir analizando otras ramas de la ciencia para tratar de encontrar más técnicas de agregación y establecer los niveles de precisión de las mismas. También se procederá en breve plazo a validar el procedimiento propuesto utilizando datos generados de forma artificial, antes de pasar a su validación utilizando datos reales.

Bibliografía

- [1] Banker and Keremer; 1989; *Scale economies in new software development*. IEEE Transactions on Software Engineering. (15): 10, pp. 1199-1205.
- [2] 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.
- [3] Hu, Q.; 1997; *Evaluating Alternative Software Production Function*. IEEE Transactions on Software Engineering. (23): 6, pp. 379-387.
- [4] 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.
- [5] 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.
- [6] Jorgensen, M.; 2004; *A Review of Studies on Expert Estimation of Software Development Effort*. Journal of Systems and Software. (70): 1-2, pp. 37-60.
- [7] Miller, J.; 2000; *Applying Meta-analytical Procedures to Software Engineering Experiments*. Journal of Systems and Software. (54): 1, pp. 29-39.
- [8] Kitchenham, B. A.; 2004; *Procedures for performing systematic reviews*. Keele University; TR/SE-0401. Keele University Technical Report.
- [9] Goodman C.; 1996; *Literature Searching and Evidence Interpretation for Assessing Health Care Practices*; SBU; Stockholm.
- [10] 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
- [11] Pickard, L. M.; Kitchenham, B. A., and Jones, P. W.; 1998; *Combining empirical results in software engineering*. *Information and Software Technology*.; 40(14):811-821.
- [12] Hedges, L.; Olkin, I.; 1985; *Statistical methods for meta-analysis*. Academic Press.
- [13] Pillemer, D. and Light, R.; 1980; *Synthesizing outcomes: How to use research evidence from many studies*. Harvard Educational Review.
- [14] Evidence-Based Medicine Working Group; 1992; Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA, 268(17), 2420-2425.
- [15] Noblit, G. W., & Hare, R. D.; 1988; *Meta-Ethnography: Synthesising Qualitative Studies*. Newbury Park, CA: Sage.
- [16] Fairbank L, O'Meara S, Renfrew MJ, Woodridge M, Sowden AJ, Lister-Sharp D.; 2000; *A systematic review to evaluate the effectiveness of interventions to promote the initiation of breastfeeding*. Health Technology Assessment; 4: 1-171
- [17] Glaser BG, Strauss AL.;1967; *The discovery of grounded theory: strategies for qualitative research*. New York: Aldine de Gruyter.

- [18] Dixon-Woods, M.; Agarwal, S.; Jones, D.; Young, B., and Sutton, A.; 2005; *Synthesising qualitative and quantitative evidence: a review of possible methods*. Journal of Health Services Research and Policy. ; 10(1):45-53B(9).
- [19] Yin, R. K. and Heald, K. A.; 1975; *Using the Case Survey Method to Analyze Policy Studies*. Administrative Science Quarterly; 20(3):371-381.
- [20] Ragin, C; 1987; *The comparative method: moving beyond qualitative and quantitative strategies*. Berkeley, California: University of California Press.
- [21] Straus, S. E. ; Richardson, W. S.; Glasziou, P., and Haynes, R. B.; 2005; *Evidence-Based Medicine*. Churchill Livingstone.
- [22] Cochrane; 2007; *Curso Avanzado de Revisiones Sistemáticas*; www.cochrane.es/?q=es/node/198
- [23] 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.
- [24] 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.
- [25] Burton, A., Shadbolt, N., Hedgecock, A. y Rugg, G.; 1988; *A Formal Evaluation of Knowledge Elicitation Techniques for Expert Systems: Domain 1*. Proceedings of Expert Systems '87 on Research and Development in Expert Systems IV. Pág. 136-145.
- [26] 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
- [27] Woody, J.; Will, R.; Blanton, J.; 1996; *Enhancing Knowledge Elicitation using the Cognitive Interview*; Expert system with application; Vol. 10 N. 1
- [28] Mohagheghi, P., & Conradi, R.; 2004; *Vote-Counting for Combining Quantitative Evidence from Empirical Studies - An Example*. Proceedings of the International Symposium on Empirical Software Engineering (ISESE'04) .
- [29] Kumar, R.; 1996; *Research Methodology: A Step-by-Step Guide for Beginners*. Adisson Wesley.
- [30] Creswell, J. 2003. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.
- [31] Marczyk, G., DeMatteo, D., Festinger, D.; 2005; *Essentials of Research Design and Methodology (Essentials of Behavioral Science)*. John Wiley & Sons.
- [32] Gurevitch, J. and Hedges, L.V.; 2001; *Meta-analysis: Combining results of independent experiments*. Design and Analysis of Ecological Experiments (eds S.M. Scheiner and J. Gurevitch), pp. 347–369. Oxford University Press, Oxford.
- [33] Grimán Padua; 2007; *Propuesta de un proceso de revisión de estudios empíricos en Ingeniería del Software*; IDoESEM.
- [34] Lajeunesse, M; Forbes, M.; 2003; *Variable reporting and quantitative reviews: a comparison of three meta-analytical techniques*. Ecology Letters, 6: 448-454.
- [35] Thalheimer W. and Cook S.; 2002; *How to calculate effect sizes from published research: A simplified methodology*. A Work-Learning Research Publication.