

Una aplicación práctica del método Delphi para la validación de una propuesta de Ingeniería Web

C.J. Torrecilla-Salinas¹, O. De Troyer², M.J. Escalona¹ and M. Mejías¹

¹ Grupo IWT2, Universidad de Sevilla
carlos.torrecilla@iwt2.org
{mjescalona,risoto}@us.es

²Department of Computer Science, Vrije Universiteit Brussel (VUB), Belgium
Olga.DeTroyer@vub.ac.be

Resumen. Las organizaciones que trabajan en el desarrollo de Sistemas de Información son reacias muchas veces a implantar nuevas metodologías de trabajo sin disponer previamente de ciertas garantías de éxito. Esta reacción es comprensible, ya que el éxito o el fracaso de ciertos proyectos puede suponer graves pérdidas económicas o reputacionales para las misma. En este trabajo vamos a presentar una aplicación práctica del uso de una técnica de juicio de expertos, el método Delphi, para la validación de una propuesta metodológica en el ámbito de la Ingeniería Web. El uso de estas técnicas puede suponer un buen compromiso en términos de inversión requerida y rápido retorno de la misma, obteniendo un juicio objetivo sobre una determinada propuesta sin tener que realizar una elevada inversión económica o arriesgar determinados proyectos que pueden ser clave para las organizaciones.

Palabras clave: Metodologías Ágiles, Scrum, Ingeniería Web, juicio de expertos, Delphi, Ingeniería del Software.

1 Introducción

Un elemento clave para la adopción de nuevas propuestas metodológicas en un entorno altamente competitivo como es el del desarrollo de Sistemas de Información, y en concreto el desarrollo de Sistemas Web, es el contar con una cierta seguridad de la validez de esas propuestas y de las mejoras que trae su adopción por parte de las organizaciones. De cara a obtener esta certitud, una aproximación clásica al problema ha sido la realización de diversos casos de estudio o casos prácticos. Sin embargo, la puesta en marcha de este tipo de experimentos supone muchas veces también la asunción de ciertos riesgos, ya que, para contar con resultados adecuados, los nuevos métodos deben ser aplicados a proyectos reales, con el consiguiente riesgo de fallo y pérdida económica y de reputación. Este hecho conlleva que en muchas ocasiones la adopción de nuevas propuestas metodológicas se realice de forma lenta, ralentizando la obtención de las ventajas que las mismas pueden traer consigo.

Los juicios de expertos aplicados a este problema pueden suponer un compromiso adecuado entre riesgo aceptado y beneficio obtenido, ya que serían una manera de proporcionar ciertas certitudes a las organizaciones al respecto de una propuesta concreta. Dentro de las diferentes propuestas existentes, podemos destacar el método Delphi [1], muy utilizado en el campo de las ciencias sociales y la medicina, pero no tan frecuente en los campos de la Ingeniería del Software y la Ingeniería Web.

En base a lo expuesto, el presente trabajo tiene los siguientes objetivos: Primero presentar el diseño de una aplicación concreta del método Delphi para la validación de una propuesta metodológica en el ámbito de la Ingeniería Web, así como presentar a alto nivel parte de los resultados obtenidos a lo largo del proceso de validación, y, por último, extraer una serie de conclusiones relevantes e identificar futuras líneas de investigación.

Para ello, Este trabajo está organizado en los siguientes apartados. Tras esta introducción, el Apartado 2 presenta las principales características del método Delphi. El Apartado 3 presenta el diseño del proceso de validación, el Apartado 4, los resultados del mismo. Por último, el Apartado 5 extrae las principales conclusiones y plantea líneas futuras de investigación.

2 El método Delphi

El método Delphi [1] apareció como una técnica de juicio de expertos durante los años 50 del pasado siglo y fue ganando en popularidad a lo largo del mismo [2]. Actualmente, sigue siendo aplicado ampliamente en diferentes áreas de conocimiento [3]. En el método Delphi, un grupo de expertos en una determinada área expresan sus puntos de vista a través de una serie de cuestionarios estructurados a lo largo de sucesivas rondas [1, 4]. El principal objetivo del método Delphi es la obtención del consenso entre el mencionado grupo de expertos acerca de una determinada materia, y sus características generales se pueden resumir en los siguientes 4 puntos [3, 5, 6]:

- Anonimidad, ya que los expertos no conocen al resto de miembros del panel.
- Iteratividad, ya que el proceso se lleva a cabo en sucesivas rondas.
- Realimentación controlada, ya que el método se basa en cuestionarios previamente definidos
- Análisis estadístico, ya que al final de cada ronda los resultados se comparten con el panel tras procesarlos con diversas técnicas estadísticas.

En general, el método Delphi se compone de las siguientes fases [7]:

- Selección de la materia a analizar, incluido el diseño del cuestionario y la selección del método de procesado de los datos obtenidos.
- Selección del panel de participantes, cuyo tamaño suele variar frecuentemente entre 15 y 20 miembros [8].

- Ronda inicial, que puede servir para clarificar la cuestión mediante un cuestionario abierto o para comenzar a obtener la opinión de los expertos mediante un cuestionario ya definido [7].
- Rondas sucesivas, en las que se analizan los resultados de las rondas precedentes y se comparte el feedback recibido con los miembros del panel. En teoría un proceso basado en el método Delphi podría durar indefinidamente, pero se ha comprobado que en general el panel converge hacia el consenso en un máximo de tres iteraciones [9, 10]
- Conclusiones, las cuales se realizan una vez finalizadas todas las rondas.

Hay que tener en cuenta dos elementos importantes cuando se procesan los resultados de un método Delphi [3]. El primero es el consenso, es decir la convergencia de las opiniones de los expertos hacia un valor concreto. La segunda es la estabilidad, es decir, que los valores sean consistentes a lo largo de las sucesivas rondas. En relación al método Delphi, el consenso sin estabilidad no tiene sentido, según muchos autores [11] y existen multitud de aproximaciones para medir ambas magnitudes a lo largo del mismo [3]. Estos métodos se pueden dividir en dos grupos [3]:

- Técnicas basadas en estadística descriptiva, como análisis de la media, mediana o desviación típica de las evaluaciones de los expertos.
- Técnicas basadas en estadística inferencial, usando estadísticos como la Chi cuadrado, la kappa de Cohen, la kappa de Fleiss o la W de Kendall.

3 Diseñando los detalles del método Delphi

Como se ha comentado anteriormente, el primer elemento de cualquier método Delphi es la definición del objeto de estudio. En nuestro caso, el objeto del mismo consistió en evaluar la validez de *NDT-Agile* [12], una propuesta metodológica basada en técnicas Ágiles que tiene por objetivo dar soporte a las características específicas de los sistemas Web a la vez que garantizar la consecución de todos los objetivos de todos los niveles de madurez CMMI-DEV [13].

Una vez definido el objeto de análisis, el siguiente paso es la definición del instrumento de medida: En nuestro caso, un cuestionario compuesto por 21 afirmaciones agrupadas en cuatro dimensiones diferentes, que tenían por objeto la evaluación de los diferentes aspectos de la propuesta: Agilidad (6 afirmaciones), soporte a CMMI (5 afirmaciones), soporte a especificidad Web (7 afirmaciones) y coherencia interna de la propuesta (3 afirmaciones).

Para la evaluación de cada afirmación se propuso a los expertos una escala de Likert [14] de 5 valores del 1 al 5, variando desde “Completo desacuerdo - 1” hasta “Completo acuerdo -5”. Asimismo, se propuso el uso de las siguientes técnicas:

- Estadística descriptiva: Basada en el análisis de la media, la mediana, la desviación típica y el porcentaje de acuerdo y desacuerdo sobre cada una de las afirmaciones propuestas.
- Alfa de Chronbach [15], un estadístico que permite evaluar la adecuación de un instrumento de medida al evento que se quiere medir y que ayuda a refinar la calidad del cuestionario si este varía a lo largo de las rondas. En nuestro caso, el coeficiente se calculará para el global del cuestionario tras obtener los resultados de la ronda. Según la literatura [16], valores del coeficiente alrededor de 0.8 indicarán una adecuación suficiente del instrumento de medida al evento medido.
- Análisis de Correspondencias Simple [17, 18], que permite calcular cuan homogéneas son las evaluaciones de los expertos y las notas atribuidas a las afirmaciones, así como la evolución de esta homogeneidad durante las sucesivas rondas, permitiendo evaluar el grado de consenso obtenido por los expertos. En este caso, el Análisis de Correspondencias Simple se utilizará como técnica exploratoria para representar en dos dimensiones las relaciones existentes, expresadas mediante las puntuaciones dadas, entre las 21 afirmaciones y cada uno de los expertos del panel.
- W de Kendall [19], que, aunque es un coeficiente estadístico diseñado para la medición del consenso, en nuestro caso se utilizará para evaluar la estabilidad de las opiniones de los expertos a lo largo de las sucesivas rondas. En este caso, y de cara a obtener una escala ordinal, tras cada ronda se ordenarán las afirmaciones en base a la media de las puntuaciones otorgadas por los expertos, con el objetivo de observar como varía la clasificación obtenida a lo largo de las rondas. Variaciones grandes indicarán poca estabilidad y variaciones pequeñas gran estabilidad, lo que se podrá medir a través del cálculo de la W de Kendall. Según la literatura, valores superiores a 0.7 de este coeficiente indicarán alta estabilidad [20]

Tras la identificación del sujeto de estudio, el instrumento de medida y las técnicas de evaluación, el siguiente paso es la selección del panel de expertos que participará en la evaluación. En nuestro caso, se trataba de evaluar la validez de *NDT-Agile* desde tres puntos de vista diferentes: Agilidad, CMMI y Web, con lo que idealmente, el panel debiera estar compuesto por expertos en las tres áreas. Sin embargo, encontrar ese tipo de perfiles es muy complicado, ya que, en concreto, CMMI y el mundo Ágil se han visto como opuestos durante mucho tiempo [21]. Por ello el panel se diseñó con expertos con experiencia práctica en al menos dos de los campos y teórica en el tercero. Inicialmente se invitó a 27 expertos a participar, de los cuales 20 aceptaron la invitación. La selección de los expertos se basó en general en su perfil profesional, teniendo en cuenta elementos como su experiencia en los campos de estudio. Asimismo, se intentó incorporar diversidad de perfiles y geográfica, para evitar posibles sesgos en los resultados obtenidos.

La experiencia media del panel en los diferentes campos objeto del estudio se muestra en la Tabla 1:

Tabla 1. Experiencia media del panel en las distintas áreas.

Área	Experiencia
Ágil	5.37 años
CMMI	8.75 años
Web	10.63 años

Por último, es interesante conocer la composición del panel en términos de perfil laboral, estando este compuesto por un 75% de expertos provenientes de Universidades o Centros de Investigación y por un 25% de expertos provenientes tanto del sector público como del sector privado. Asimismo, un 60% de los miembros del panel son investigadores, encontrándose también expertos que desarrollan tareas de jefatura de proyecto, así como cuadros directivos. Esto unido a la diversidad geográfica de los expertos (provenientes de 8 países diferentes, incluyendo España, Argentina, Alemania o Bélgica), nos proporcionó una diversidad y variedad de opiniones que enriquecieron las conclusiones del proceso.

4 Desarrollo del método y resultados

El método Delphi se planteó sin un número determinado de rondas predefinido, siendo el criterio para finalizar el mismo la obtención de consenso y estabilidad, tal y como se ha indicado anteriormente. Así, el proceso se desarrolló en tres rondas sucesivas, desde febrero hasta junio de 2016. Al término de la tercera se comprobó el cumplimiento de las dos condiciones mencionadas, procediéndose a dar por terminado el proceso.

Durante cada una de las tres rondas se obtuvieron las diferentes valoraciones de los expertos, las cuales se procesaron haciendo uso de las diversas técnicas estadísticas anteriormente expuestas. Es importante destacar que las evaluaciones de los expertos se llevaron a cabo a través de formularios Web, de forma remota, y que el cuestionario se modificó tras la ronda inicial para incorporar clarificaciones y sugerencias requeridas por los participantes.

Como se ha comentado, la alfa de Cronbach se calculó sobre el cuestionario global. Los cálculos se realizaron tanto al final de la primera ronda como de la segunda (ya que se realizaron cambios en el cuestionario que podrían haber afectado a su adecuación). Los resultados obtenidos en ambos casos (alrededor de 0.8 en ambos casos) nos proporcionaron una alta confianza en la adecuación del cuestionario a las diferentes medidas.

El uso combinado de la estadística descriptiva y el Análisis de Correspondencia Simple nos permitió obtener una idea clara del nivel de consenso y del progreso del mismo

a lo largo del proceso. En la figura 1 se muestra el resultado de este Análisis tras los resultados obtenidos en la segunda ronda, donde los puntos azules representan a los expertos y los rojos las preguntas, indicando la distancia entre las mismas la homogeneidad de la muestra.

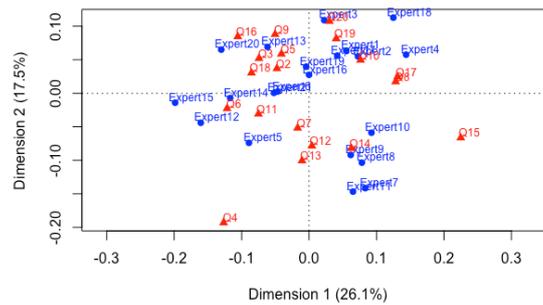


Fig. 1. Resultado del Análisis de Correspondencias Simple al final de la segunda ronda

Finalmente, la evolución de los valores de la W de Kendall (de en torno a 0.5 tras la segunda ronda hasta valores cercanos a 0.9 al final de la tercera) nos mostró como la estabilidad de las evaluaciones se alcanzó al final de la tercera ronda.

Para concluir, y de cara a interpretar los resultados finales, definimos “acuerdo fuerte” como aquel obtenido en las afirmaciones con una evaluación media de 3.7 (en una escala de 0 a 5, siendo 5 el valor máximo), una mediana de 4 (significando “Acuerdo”) y al menos el 60% de los expertos con valores 4 o 5 (“Acuerdo” o “Completo acuerdo”), con un mínimo de 12 expertos opinando. Definimos “acuerdo débil” como aquel obtenido en las afirmaciones con una evaluación media entre 3.5 y 3.7, una mediana de 3.5 y al menos el 45% de los expertos con valores 4 o 5 (“Acuerdo” o “Completo acuerdo”). La Tabla 2 muestra los resultados globales del método Delphi tras la tercera ronda:

Tabla 2. Resultados globales del método Delphi.

Nivel de acuerdo	Número de afirmaciones (%)
Acuerdo fuerte	16 (76.19%)
Acuerdo débil	5 (23.81%)
Desacuerdo	0 (0%)

5 Conclusiones y trabajos futuros

Durante el presente trabajo hemos presentado los detalles del diseño de un proceso basado en el método Delphi aplicado a la validación de una propuesta en el área de la

Ingeniería Web. Durante el mismo, hemos podido destacar sus elementos fundamentales, incluyendo la selección detallada del problema, el diseño del instrumento de medida, la selección de las herramientas estadísticas para el análisis de los datos y el establecimiento de un panel de expertos adecuado para la validación. Hemos presentado también los resultados del proceso Delphi, mostrando los valores obtenidos y los resultados generales del análisis de los mismos a alto nivel. Hemos podido ver como la aplicación del método nos permite una validación de la propuesta con un margen de confianza adecuado sin necesidad de poner en riesgo una cierta inversión por parte de organizaciones a través de la implementación de casos de estudio.

Por otro lado, y cómo posibles riesgos a tener en cuenta a la hora de diseñar un proceso de este tipo, hay que tener en cuenta que este puede extenderse a lo largo de mucho tiempo (5 meses en nuestro caso), lo que, unido a la necesidad de contar con la opinión de los expertos a lo largo de las diferentes rondas, puede generar un cierto “hartazgo” en los participantes. A la hora de diseñar el proceso estos riesgos deben ser tenidos en cuenta para minimizarlos en la medida de lo posible.

Una conclusión general del trabajo puede ser por tanto la validez de este método también para el área de la Ingeniería Web aunque, y este elemento queda como línea de trabajo futura, sería interesante definir y validar qué herramientas y reglas específicas deben adoptarse para la generalización de este tipo de validaciones a los problemas de la Ingeniería del Software y la Ingeniería Web en particular.

6 Agradecimientos

Esta investigación ha sido apoyada por el proyecto MeGUS (TIN2013-46928-C3-3-R) del Ministerio de Ciencia e Innovación y el proyecto SoftPLM (TIN2015-71938-REDT) del Ministerio de Ciencia e Innovación del Gobierno de España. Queremos agradecer también a al Dr. Pedro Antonio García, al Dr. Diego Torrecilla de Amo y al Dr. Diego Nieto Lugilde, todos ellos de la Universidad de Granada, por sus comentarios y ayuda. Por último, queremos agradecer a todos los expertos participantes en el proceso por su tiempo y ayuda.

7 Referencias

1. Dalkey, N.C.; Helmer, O. “An experimental application of the Delphi method to the use of experts”, *Manag. Sci.* 9 (1963) 458–467.
2. Rieger, W.G. 1986. “Directions in Delphi developments: dissertations and their quality”. *Technol. Forecast. Soc. Chang.* 29 (1986) 195–204.
3. Heiko, A. 2012. “Consensus Measurement in Delphi Studies: Review and Implications for Future Quality Assurance”. *Technological forecasting and social change*, 79(8), 1525-1536.
4. Linstone, H. A.; Turoff, M. 1975. “The Delphi method: Techniques and applications”. (pp. 3-12). Reading, MA: Addison-Wesley Publishing Company.

5. Rowe, G.; Wright, G. 2001. "Expert opinions in forecasting: the role of the Delphi technique". J.S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, Boston et al, 2001, pp. 125–144.
6. Dalkey, N. C. 1972. "The Delphi method: An experimental study of group opinion". In N. C. Dalkey, D. L. Rourke, R. Lewis, & D. Snyder (Eds.). *Studies in the quality of life: Delphi and decision-making* (pp. 13-54). Lexington, MA: Lexington Books.
7. Hsu, C. C.; Sandford, B. A. 2007. "The Delphi technique: making sense of consensus". *Practical assessment, research & evaluation*, 12(10), 1-8.
8. Ludwig, B. 1997. "Predicting the future: Have you considered using the Delphi methodology?". *Journal of Extension*, 35 (5), 1-4.
9. Cyphert, F. R.; Gant, W. L. 1971. "The Delphi technique: A case study". *Phi Delta Kappan*, 52, 272-273.
10. Brooks, K. W. 1979. "Delphi technique: Expanding applications". *North Central Association Quarterly*, 54 (3), 377-385.
11. Dajani J. S.; Sincoff, M.Z.; Talley, W.K. 1979. "Stability and agreement criteria for the termination of Delphi studies", *Technol. Forecast. Soc. Chang.* 13 (1979) 83–90.
12. Torrecilla Salinas, C.J.; Sedeño, J.; Escalona, M.J.; Mejías, M. 2015. "Estimating, planning and managing Agile Web development projects under a value-based perspective". *InfSoftwTechnol* 61 (2015) 124–144.
13. CMMI Product Team. 2010. "CMMI for Development, Version 1.3.". Nov. 2010. Carnegie Mellon University. Technical Report. From <http://www.sei.cmu.edu/reports/10tr033.pdf>. Accedido 02-2017.
14. Likert, R. 1932. "A technique for the measurement of attitudes". *Archives of psychology*.
15. Cronbach, L. J. 1951. "Coefficient alpha and the internal structure of tests". *Psychometrika*, 16(3), 297-334.
16. Gliem, R. R.; Gliem, J. A. 2003. "Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales". *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education*.
17. Hirschfeld, H.O. 1935. "A connection between correlation and contingency", *Proc. Cambridge Philosophical Society*, 31, 520–524
18. Benzécri; J. P. "L'Analyse des Données. Volume II, L'Analyse des Correspondances". 1973.
19. Legendre, P. 2005. "Species associations: the Kendall coefficient of concordance revisited". *Journal of agricultural, biological, and environmental statistics*, 10(2), 226-245.
20. García-Crespo, A.; Colomo-Palacios, R.; Soto-Acosta, P.; Ruano-Mayoral, M. 2010. "A qualitative study of hard decision making in managing global software development teams". *Information Systems Management*, 27(3), 247-252.
21. Staples, M.; Niazi, M.; Jeffery, R.; Abrahams, A.; Byatt, P.; Murphy R. 2007. "An exploratory study of why organizations do not adopt CMMI". *J. Syst. Softw.* 80, 6 (June 2007), 883-895.