

Evaluación de Variantes de Inspección en un Modelo Léxico

Alberto Sebastián¹, Graciela D. S. Hadad^{1,2}, Daniela Raffo¹

¹Facultad de Ingeniería y Tecnología Informática, Universidad de Belgrano, Argentina

²Escuela de Informática, Universidad Nacional del Oeste, Argentina

{alberto.sebastian;graciela.hadad;daniela.raffo}@comunidad.ub.edu.ar

Resumen. Elaborar modelos en lenguaje natural durante el proceso de requisitos es una práctica habitual que facilita la elicitación, validación y negociación con el cliente. Es frecuente construir inicialmente un glosario con los términos relevantes utilizados por los clientes. A pesar de construirse este glosario siguiendo heurísticas concretas, suele presentar un número significativo de defectos. Por consiguiente, es importante detectarlos y corregirlos tempranamente para evitar su propagación sobre posteriores modelos. En tal sentido, la inspección es una técnica de verificación que consigue una alta eficacia en identificar defectos. Existen distintas variantes de inspección sobre modelos de requisitos, centradas en el modo de lectura del modelo para capturar defectos. En el presente trabajo, se presenta un estudio comparativo, en términos de eficacia y eficiencia, de tres variantes de inspección: lectura con checklist, lectura basada en procedimientos y lectura constructiva, aplicadas sobre un modelo léxico. Con los resultados de 23 inspecciones, se observó que la lectura constructiva logró mayor tasa de detección de defectos, focalizada principalmente en omisiones y ambigüedades, aunque insumiendo más tiempo que la lectura con checklist, mientras que la lectura basada en procedimientos fue la que insumió los mayores tiempos, capturando en promedio una menor cantidad de defectos.

Palabras Clave: Verificación, Inspecciones, Modelos en Lenguaje Natural.

1 Introducción

Es común en la Ingeniería de Requisitos (IR) escribir modelos en lenguaje natural (LN) por su cercanía con el cliente. Suele construirse inicialmente un glosario con los términos significativos del contexto de aplicación, denominado Léxico Extendido del Lenguaje (LEL) [1]. No obstante disponer de heurísticas específicas para construirlo, se capturan muchos defectos durante su verificación. Las omisiones y ambigüedades son los defectos más frecuentes en modelos en LN [2,3]. En un estudio empírico sobre casos de uso, se observó que el 50% contenía ambigüedades [4]. Trabajos realizados sobre completitud del modelo LEL arrojaron niveles de omisión superiores al 40% [5].

Una de las técnicas de verificación más estudiadas y utilizadas con gran éxito en proyectos de software es la inspección, diseñada para encontrar defectos en el código fuente [6]. La IR ha adaptado esta técnica a sus propios modelos [2,7,8].

En el presente artículo se exponen los resultados de un estudio comparativo de variantes de inspección aplicadas sobre el modelo LEL, a fin de establecer qué variantes pueden ser más eficaces para capturar determinados defectos y qué esfuerzo requieren.

La sección 2 describe la técnica de inspección. La sección 3 presenta las variantes de inspección utilizadas en el estudio comparativo. La sección 4 describe el estudio y el análisis de los resultados. La sección 5 expone conclusiones y futuros trabajos.

2 Inspección de Software

Las inspecciones de modelos u otros artefactos de software representan una herramienta confiable ampliamente utilizada en la Ingeniería de Software [9,10], que permite obtener artefactos con mucha menor cantidad de defectos.

La inspección de software implica seguir un proceso preciso para encontrar defectos, que involucra los roles de inspector, autor, moderador y escriba, desarrollándose en seis pasos [6]: 1) Planificación: se identifica el material a inspeccionar, se determinan las personas que ocuparán los roles y se establece la fecha de la reunión de inspección. 2) Apreciación Global: los autores describen a los inspectores el material a inspeccionar. 3) Preparación: los inspectores leen en forma individual el material. 4) Reunión de Inspección: los inspectores revisan el material detectando defectos y se los transmiten a los productores para su aceptación o no, asistidos por el moderador mientras el escriba registra los hechos. 5) Corrección de los defectos a cargo de los autores. 6) Seguimiento: el moderador controla el estado del material corregido. En general, en este proceso los defectos se catalogan por tipo y por severidad, y se registran los tiempos de detección, con el propósito de mejorar el proceso en futuras inspecciones.

Debido a la eficacia de la técnica, se han diseñado variantes aplicadas a los modelos de IR, según la modalidad de lectura en la Preparación [11]:

- Lectura ad-hoc: el inspector lee el material dependiendo exclusivamente de su experiencia, sin recibir apoyo para encontrar defectos.
- Lectura basada en lista de comprobación: el inspector usa una lista con controles que le sirve de guía en la detección de defectos.
- Lectura basada en procedimientos: el inspector sigue procedimientos detallados para la captura de defectos. Es una modalidad altamente estructurada.
- Lectura constructiva: el inspector produce un artefacto intermedio a partir del material, el cual luego es utilizado para detectar los defectos.

Existen trabajos que han evaluado algunas de las tres primeras variantes de inspección en IR, para establecer su productividad considerando algunos factores: modelo a verificar, tipos de defectos a detectar, severidad de los defectos, tiempo insumido o experiencia del inspector [8,11,12]. Armijos et al. [8] proponen medir la calidad de las inspecciones con una métrica propia, que califica los aciertos de los inspectores y penaliza sus desaciertos considerando el tipo y severidad de los defectos detectados en la Preparación y confirmados en la Reunión. Sin embargo, no utiliza el tiempo insumido, centrándose en evaluar la influencia de la experiencia del inspector. Paech et al. [7] comprobaron que se detecta en promedio más defectos usando la lectura basada en procedimientos que la lectura con lista de comprobación, aunque requiere más tiempo.

Es ampliamente aceptado que las inspecciones son de gran utilidad para lograr software de calidad, no obstante, algunos autores plantean su baja adopción en la industria [9]. Según Macchi y Solari [10], las principales causas son: la rigurosidad del proceso, el costo excesivo por el tiempo insumido y la cantidad de inspectores, la falta de claridad en la relación esfuerzo insumido vs calidad obtenida, y la falta de formación de inspectores y su curva de aprendizaje. Es por ello, que sería relevante poder establecer la relación entre tiempo requerido, calidad obtenida y experiencia del inspector.

3 Inspecciones aplicadas a Modelos en Lenguaje Natural

En los modelos escritos en LN, se utiliza un proceso de inspección basado en la propuesta original [6] con algunas adaptaciones para inspeccionar modelos en LN [2], tal como el LEL. Se descarta el paso *Apreciación Global*, dado que los modelos en LN deberían comprenderse por sí mismos sin necesidad de una explicación por parte de los autores. El paso *Preparación* se focaliza en la detección de defectos, en lugar de una mera lectura del material, por lo que el paso *Reunión de Inspección* se dedica a ratificar o rectificar los defectos encontrados en la *Preparación*.

Es importante en las inspecciones determinar cómo se deben percibir los defectos, es decir, cuáles son sus manifestaciones características. Esto guiará el proceso de inspección en la *Preparación*, facilitando la captura de defectos.

En este estudio, los defectos se han tipificado en: *Discrepancias o inconsistencias*, *Errores o hechos incorrectos*, *Omisiones y Ambigüedades*, aplicable a cualquier modelo en LN [2,13]. Cada defecto se califica por grado de severidad: *Alto*, *Medio* o *Bajo*.

Se utilizaron tres variantes para inspeccionar el modelo LEL, donde las guías de cada variante especifican el/los defecto/s a capturar, sus tipos y niveles de severidad.

i) *Lectura con lista de comprobación: Variante usando CheckList (VCH)*. Se usa una lista con 44 ítems que representan controles que el inspector debe aplicar al LEL.

ii) *Lectura con procedimientos: Variante usando Formularios (VFR)* [13]. El inspector completa 12 formularios siguiendo guías por formulario, que indican el objetivo del formulario, cómo llenarlo y cómo analizar los datos para identificar defectos.

iii) *Lectura constructiva: Variante construyendo Mapas Conceptuales (VMC)* [14]. Se construye un mapa conceptual desde cada término del LEL y se siguen 13 pasos para analizar los mapas y sus relaciones. Cada paso indica cómo detectar defectos en el mapa, su tipo y severidad, con recomendaciones para la posterior corrección del LEL.

4 Estudio Comparativo

Se realizó un estudio comparativo con el fin de evaluar las tres variantes de inspección, usando 6 modelos léxicos diferentes. Se realizaron un total de 23 inspecciones: 7 con VCH, 8 con VFR y 8 con VMC. En cada inspección un estudiante diferente cumplía el rol de inspector. Tanto en la *Preparación* como en la *Reunión*, se tomaron datos detallados de tiempos y cantidad de defectos por tipo y severidad, en cada ítem de variante.

Respecto a las amenazas a la validez interna del estudio, debe notarse que la selección de inspectores fue entre estudiantes de grado en el curso Ingeniería de Requisitos

en distintas universidades. A cada inspector se le asignó un LEL y una variante de inspección a aplicar sobre dicho LEL. Todos los inspectores tenían conocimiento previo sobre el modelo LEL. Los LELs utilizados fueron creados por otros estudiantes en contextos de sistemas de información para organizaciones reales, siguiendo las mismas heurísticas de construcción, en forma totalmente independiente del estudio comparativo. Los inspectores tuvieron una capacitación de 2 horas sobre la técnica de inspección y la variante a aplicar, y recibieron la documentación necesaria para seguir el proceso establecido y registrar los datos solicitados. Dado que las guías en las tres variantes estipulaban el tipo de defecto a capturar y el nivel de severidad, se evitó la interpretación del inspector y, por ende, el desvío en las métricas aplicadas.

Respecto a las amenazas a la validez externa, los sujetos del estudio son solo representativos de inspectores sin experiencia. Esto implica que los resultados obtenidos no son representativos de inspecciones en la industria, pero pueden ser válidos si solo se dispone de inspectores novatos. El material inspeccionado eran LELs elaborados por estudiantes sin experiencia en su construcción, por lo que los resultados pueden no ser representativos de la cantidad de defectos reales que pueden aparecer en este tipo de modelo construido por profesionales con experiencia en IR. En cuanto al tamaño de los LELs (22 a 43 símbolos) corresponden a proyectos pequeños a medianos, pero realizados en organizaciones reales. El proceso de inspección utilizado fue propuesto y evaluado en la literatura [2,13,12]. Los datos medidos se normalizaron dado que no se pudo utilizar el mismo modelo léxico en todas las inspecciones.

En un análisis descriptivo de los resultados de las inspecciones, se observa que la VMC, aunque requiere más tiempo por símbolo, detecta en promedio más defectos por símbolo, mientras que por el contrario la VFR lleva un esfuerzo en tiempo apenas menor que la VMC, pero detecta un número muy inferior de defectos por símbolo (ver Fig. 1-A y 1-B). Esto evidencia que en promedio el tiempo para capturar un defecto en la VFR es 2.4 veces superior que en las otras variantes (ver Fig. 1-C), aunque entre estas dos variantes VMC detecta más defectos por símbolo que VCH (ver Fig. 1-A).

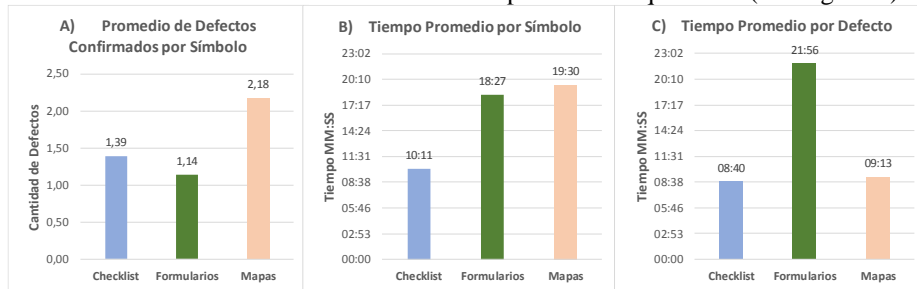


Fig. 1. Valores promedio de número de defectos y tiempo por variante de inspección

La VCH tuvo una dispersión excesivamente alta en la tasa promedio de tiempo requerido por símbolo y una dispersión relativa en la tasa de defectos por símbolo. La VFR también tuvo una alta dispersión en la tasa de defectos por símbolo. Estas dispersiones debilitan, en cierta medida, las observaciones previas. Cabe notar que, por el contrario, la VMC tuvo bajos coeficientes de variación para ambos tipos de tasa.

Todas las inspecciones (8) usando VMC lograron una tasa de detección superior al 1.50 defectos por símbolo, mientras que solo el 43% de las inspecciones con VCH (3 de 7) y el 25% con VFR (2 de 8) superaron esa tasa de detección.

Las tres variantes presentan una buena tasa promedio de defectos confirmados en la Reunión respecto a los capturados en la Preparación: para VFR y VMC más del 79% de los defectos fueron confirmados por los autores, y para VCH se confirmaron en promedio casi el 73%. Esto puede indicar guías precisas de detección para los tres casos.

La VMC presenta 4 inspecciones (de 8) que capturaron defectos de alta severidad con tasas ≥ 0.50 defectos por símbolo, frente a solo 2 inspecciones de VCH y 1 inspección de VFR. En cuanto a defectos de severidad media, las 8 inspecciones de VMC tuvieron tasas ≥ 0.50 defectos por símbolo frente a 4 inspecciones de VCH y 4 de VFR.

Ninguna inspección de VMC detectó discrepancias. Solo 2 inspecciones de VCH las detectaron y 4 inspecciones de VFR, en ambas variantes con una tasa ≤ 0.12 discrepancias por símbolo. Las tres variantes encontraron errores con tasas de detección muy bajas: 1 sola inspección de VMC (tasa ≤ 0.10), 7 inspecciones de VFR (tasa ≤ 0.34) y para VCH hubo 3 inspecciones con tasa ≤ 0.34 y 4 con tasa ≤ 0.69 . Es decir, todas las inspecciones de VCH detectaron errores, y con tasas un poco superiores a las de VFR.

Las 8 inspecciones con VMC tuvieron tasas de detección ≥ 0.70 omisiones por símbolo, frente a 3 inspecciones para VFR y 3 para VCH con esas tasas. Cabe notar que 7 inspecciones de VMC tuvieron tasas superiores a 1 omisión por símbolo. En el caso de ambigüedades, 7 inspecciones de VMC tuvieron una tasa ≥ 0.70 ambigüedades por símbolo, frente a una sola inspección de VCH con igual tasa. VCH y VFR solo tienen 2 inspecciones cada una con tasas entre 0.40 y 0.70 ambigüedades por símbolo.

5 Conclusiones

Las inspecciones demandan tiempo y esfuerzo humano, pero tienen alta efectividad en mejorar la calidad del software [9,10]. Se realizó un estudio comparativo para evaluar la efectividad de tres variantes de inspección aplicadas a diversos LELs. Se midieron tiempos insumidos, tipos de defectos y grados de severidad, observándose lo siguiente:

- La VMC lleva más tiempo de inspección que VCH, pero logra tasas de detección de defectos muy superiores. La VFR insumió mucho más tiempo capturando menos defectos por símbolo. Esto corrobora parcialmente los resultados presentados en [7,12] sobre VCH y VFR. La lista de control usada en la VCH es más extensa que la habitual en inspecciones en la industria, lo que pudo colaborar en una detección más eficaz.
- La VMC tuvo más inspecciones con tasas más altas de detección de defectos, mientras que VFR tuvo más inspecciones con bajas tasas, siendo VCH la que tuvo inspecciones con tasas muy variadas. Para la VCH es probable que la detección dependa en gran medida de la habilidad del inspector, dado que es la variante menos guiada.
- Las tres variantes tuvieron en promedio un porcentaje relativamente bajo de rechazos de defectos en la Reunión de Inspección.
- La VMC tuvo más inspecciones con las tasas más altas de detección para defectos de alta y de media severidad.
- La VMC se centra en capturar omisiones y ambigüedades, con tasas de detección muy superiores a las otras variantes. Tanto VCH como VFR detectaron discrepancias, con tasas muy bajas de detección. La VCH tuvo una mejor eficacia en capturar errores.

Dada la eficacia de la VMC en detectar omisiones y ambigüedades, se espera extenderla para mejorar la captura de discrepancias y errores.

Debe notarse que se utilizó una muestra pequeña de 23 inspecciones, realizadas por

inspectores novatos sobre LELs construidos también por novatos, por lo que los resultados no pueden ser generalizados a inspecciones en la industria, aunque pueden dar indicios de cómo se podrían detectar con mayor eficacia ciertos tipos de defectos o prever ciertos tiempos de inspección. Se espera realizar nuevas inspecciones con las tres variantes para precisar los resultados obtenidos. Sería apropiado utilizar modelos LEL de mayor tamaño e inspectores con distinto grado de experiencia. Sobre este punto ha habido varios trabajos [8], aunque no han sido conclusivos, pues no relacionan el impacto de la destreza del inspector con la tasa de detección y la variante de inspección. Por otro lado, se ha observado en la literatura una falencia en el estudio de inspecciones que aplican la lectura constructiva, como la VMC. Las variantes con checklist y con procedimientos han sido aplicadas al modelo de Escenarios [2]; se podría diseñar una variante constructiva para ese modelo, como también adaptarlas a otros modelos en LN.

Referencias

1. Leite, J.C.S.P, Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridao, M.N.: Defining System Context using Scenarios. En: Leite & Doorn (eds.) Perspectives on Software Requirements, Kluwer Academic Publishers, pp. 169-199 (2004).
2. Leite, J.C.S.P, Doorn, J.H., Hadad, G.D.S., Kaplan, G.N.: Scenario Inspections, Requirements Engineering Journal, 10(1), 1-21. Springer-Verlag, Londres (2005).
3. Berry, D., Kamsties, E.: Ambiguity in Requirements Specification. En: Perspectives on Software Requirements, pp.7-44. Kluwer Academic Publishers (2004).
4. Ben Achour, C., Rolland, C., Maiden, N.A.M., Souveyet, C.: Guiding Use Case Authoring: Results of an Empirical Study. En: International Symposium on Requirements Engineering, Limerick, Irlanda, IEEE Computer Society Press, pp. 36-43 (1999).
5. Hadad, G.D.S., Litvak, C., Doorn, J.H., Ridao, M.: Dealing with Completeness in Requirements Engineering. En: Mehdi Khosrow-Pour (ed) Encyclopedia of Information Science and Technology, 3rd ed, IGI Global, pp. 2854-2863 (2015).
6. Fagan, M.E.: Design and Code Inspections to reduce Errors in Program Development, IBM Systems Journal, 15(3), 182-211 (1976).
7. Paech, B., Denger, C., Kerkow, D., Von Knethen, A.: Requirements Engineering for Technical Products: Integrating Specification, Validation and Change Management. En: Maté & Silva (eds.) Information Science Publishing, Londres (2005).
8. Armijos, A.F., Monsalve, C., Ullón, R.H., Maya, R.D., Romero, J.A.: Calidad en la inspección de requerimientos de software: una propuesta de inspección. En: XIX Ibero-American Conference on Software Engineering, Quito, Ecuador (2016).
9. Ciolkowski C., Laitenberger, O., Biffel, S.: Software reviews: The state of the practice, IEEE software, 20(6), 46–51 (2003).
10. Macchi, D., Solari, M.: Mapeo Sistemático de la Literatura sobre la Adopción de Inspecciones de Software. En: Conferencia Latinoamericana de Informática, Medellín, pp.1-8 (2012).
11. Regnell, B., Runeson, P., Thelin, T.: Are the perspectives really different? Further experimentation on scenario-based reading of requirements. Empirical Software Engineering, 5, 331-356 (2000).
12. Porter, A.A., Votta, L.G.: Comparing Detection Methods for Software Requirements Inspections: A Replication Using Professional Subjects. Empirical Software Engineering, 3(4), 355-380 (1998).
13. Kaplan, G., Hadad, G., Doorn, J., Leite, J.: Inspección del Léxico Extendido del Lenguaje. En: 3rd Workshop on Requirements Engineering, Río de Janeiro, Brasil, pp.70-91 (2000).
14. Sebastián, A., Hadad, G., Robledo, E.: Inspección centrada en Omisiones y Ambigüedades de un Modelo Léxico. En: 20th Workshop on Requirements Engineering, Argentina (2017).