

Reducción de la Subjetividad en los procesos de requisitos

Gladys Noemí Kaplan¹, Jorge Horacio Doorn^{2,3}

¹Departamento de Ingeniería e Investigaciones Tecnológicas. Universidad Nacional de la Matanza. Argentina

²Escuela de Informática, Universidad Nacional del Oeste

³Departamento de Ingeniería, Universidad Nacional de Tres de Febrero
gkaplan@unlam.edu.ar, jdoorn@exa.unicen.edu.ar

RESUMEN

Los procesos de requisitos son esencialmente cognitivos, ya que consisten en generar un nuevo sistema de software a partir de la información que se adquiere, del conocimiento existente y de la valoración y elaboración de estos elementos. A medida que se avanza en el proceso, los modelos ya construidos pasan a ser una fuente de información del próximo modelo. De esta manera se avanza más rápidamente y se ayuda al ingeniero de requisitos a evitar el conocido “bloqueo del escritor” que es un tipo de parálisis cognitiva que se genera al comenzar un nuevo modelo. El proceso de requisitos utilizado en el presente artículo, construye como primera actividad, un glosario denominado LEL, desde el cual se genera una primera versión de escenarios actuales. Esta derivación, que toma de manera textual la información del glosario, no ha valorado correctamente que el LEL es esencialmente declarativo y que los escenarios son esencialmente procedurales. El presente proyecto propone generar un nuevo mecanismo de derivación que aproveche la información disponible en el LEL ampliando el conocimiento con toda la documentación disponible, revisando la misma desde una mirada de procesos. De esta manera se espera obtener escenarios mucho más completos y reales.

Palabras clave: Proceso de Requisitos, Lenguaje Natural, Escenarios, Derivación.

CONTEXTO

La línea de investigación que se presenta es parte de los proyectos de investigación “Aspectos no funcionales en los procesos de

requisitos” de la Universidad Nacional de La Matanza (UNLaM).

1. INTRODUCCIÓN

Este proyecto está enmarcado en una estrategia de Ingeniería de Requisitos [1] [2] [3] orientada al cliente, ya que se basa en la construcción de modelos en lenguaje natural (LN) [4]. Su objetivo es alcanzar una profunda comprensión del dominio antes de definir el sistema de software. A tal efecto, requiere de dos grandes etapas bien distinguibles: una de aprendizaje y la otra de definición. De existir conocimiento previo, la primera etapa, se convierte en una actividad confirmatoria. Los modelos que se utilizan en esta estrategia son:

1) un modelo léxico, LEL (Léxico Extendido del Lenguaje) [5] [6], el cual describe el vocabulario utilizado en el dominio de la aplicación.

2) un modelo organizacional que describe los procesos actuales del negocio, denominado Escenarios Actuales [7].

3) un modelo organizacional que describe los procesos del negocio proyectados incluyendo el sistema de software a desarrollar [8] [9], denominado Escenarios Futuros.

Los escenarios futuros tienen empotrados los requisitos del software. El documento de especificación de requisitos de software (ERS), se obtiene extrayendo los requisitos del software desde estos escenarios [10].

Debido a la importancia que tiene la comprensión del contexto en el presente proceso de requisitos, es que para definir una propuesta adecuada a las necesidades de cada

cliente, es necesario comprender lo mejor posible el contexto y sus particularidades. La heurística de derivación [11], no contribuye a mejorar esa comprensión ya que toma de manera casi directa la información desde el LEL, con un mínimo de procesamiento. Luego, con este conjunto de escenarios candidatos, se retorna al UdeD para completarlos, rectificando o ratificando la información del LEL.

La actividad Derivar se realiza teniendo en cuenta el conocimiento del LEL, el modelo de escenarios y la heurística de derivación. Consta de tres sub actividades: Identificar los actores del UdeD, Identificar los Escenarios y por último Crearlos usando el LEL. De esta manera se crea una lista inicial de escenarios candidatos con los que se comienza a trabajar. Los pasos son:

a) IDENTIFICAR ACTORES

Los símbolos que representan actores en el UdeD son identificados en el LEL. Ellos corresponden a los símbolos clasificados con el tipo Sujeto.

b) IDENTIFICAR ESCENARIOS

Cada impacto de los símbolos Sujeto representan un posible Escenario candidato y son incorporados a la lista de Escenarios Candidatos. El título del escenario está compuesto por la acción (verbo) incluido en el impacto pero presentado en infinitivo.

Cuando diferentes actores ejecutan la misma acción, puede producir que dos o más escenarios de la lista puedan compartir el título; en este caso es recomendable no eliminar ninguno hasta asegurarse que realmente son iguales.

c) CREAR

Aquí se intenta construir los Escenarios con la mayor cantidad posible de información extraída del LEL, aplicando la heurística de creación de Escenarios. El producto son los Escenarios candidatos derivados.

El contenido de cada impacto es analizado para encontrar en ellos, símbolos del LEL del tipo Verbo.

Si existe un símbolo Verbo en el impacto:

El objetivo es definido en base al título del Escenario y la noción del símbolo Verbo. Los actores y recursos son identificados en la información del símbolo Verbo, mirando los símbolos Sujeto y Objeto existentes en él. Los episodios son derivados de cada impacto del símbolo Verbo. La precondition del Contexto es derivada de las situaciones que impiden que la acción se realice del símbolo Verbo.

Si no existe un símbolo Verbo en el impacto: Se verifican los símbolos del impacto en cuestión y se toman como posible fuente de información. El objetivo se define en base al título del Escenario. Posibles actores y recursos son seleccionados de los símbolos en cuestión leyendo toda su descripción. Los actores son derivados de los símbolos Sujetos que le dieron origen y los recursos de los símbolos Objetos. No se derivan episodios desde el LEL, ellos serán detectados en los pasos siguientes.

En ambos casos, la ubicación geográfica y la ubicación temporal del contexto puede ser extraída de los impactos del símbolo que originó el Escenario (símbolo Sujeto). También se puede obtener información relevante de los símbolos relacionados con el símbolo Sujeto.

Al completar estos escenarios a partir de la nueva información obtenida desde el UdeD, algunos escenarios se transforman en episodios, otros se dividen en varios escenarios o se unifican en uno solo. Cabe destacar que estos escenarios candidatos actúan como un corsé en la búsqueda de nuevo conocimiento, porque el ingeniero de requisitos ya tiene identificadas las situaciones que debe analizar (lista de escenarios candidatos). El paso siguiente en el proceso, describir los escenarios, se reduce prácticamente a elicitar conocimiento para completar estos escenarios y eliminar la etiqueta de candidatos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La derivación existente tiene varias debilidades que paradójicamente, en vez de ayudar al ingeniero de requisitos a avanzar eficientemente, genera un conjunto de escenarios con problemas estructurales y potencialmente incompletos. Este último punto es particularmente relevante porque para construir el LEL se cuenta con mucho conocimiento del dominio, que no fue utilizado en la construcción del glosario. Este conocimiento existente del dominio se encuentra en documentación del contexto (manuales, protocolos, etc.) y principalmente en las transcripciones de las entrevistas realizadas para elicitación del léxico y luego, para validar. La derivación desde el LEL, solo toma el conocimiento existente en el mismo glosario, ignorando el resto del conocimiento con el que ya se cuenta. Por lo tanto, el retorno al UdeD es, en muchas ocasiones, elicitación de conocimiento que ya se posee.

Puntualmente, las debilidades de la derivación existente son:

- 1) La generación de una lista inicial de escenarios candidatos
- 2) El pasaje de la información casi directa desde el LEL, sin una revisión de la misma con una visión procedural.
- 3) No se utiliza todo el conocimiento existente en el LEL para construir los escenarios, es decir utiliza solo algunos símbolos.
- 4) La heurística existente ignora los símbolos de tipo Estado, dejando un aspecto relevante del dominio sin analizar.
- 5) La heurística existente no contempla las jerarquías conceptuales [12] y los puntos de vista del contexto, los cuales contienen aspectos específicos del dominio.

Estas situaciones obtenidas desde el LEL son efectivamente reales del contexto, pero al ser analizadas desde una mirada procedural, pueden tener una granularidad diferente a la existente en el dominio. Esto se debe a que hay un salto conceptual entre la perspectiva propuesta en el LEL y la de los escenarios.

Una de ellas observa el UdeD buscando léxico y la otra se ocupa de comprender los procesos del negocio. En los escenarios, el LEL es utilizado para reducir la ambigüedad y mejorar la precisión de las descripciones. Puede observarse que el objetivo del LEL es transversal a los escenarios, siendo difícil que se pueda obtener una mirada de los procesos con solo trasladar conocimiento de un modelo a otro, como propone la derivación actual. Debe observarse que las acciones relevantes del dominio no suscitan un tratamiento profundo. Al crear los escenarios, estas situaciones tienen la misma granularidad del LEL, heredando la dispersión de conocimiento adecuado para describir los términos de un glosario, pero no cuando se desea visualizar los procesos. En este contexto, la aparición de nuevas situaciones queda casi a merced de que el cliente-usuario las mencione o de la habilidad del ingeniero de requisitos en reconocerlas. Si bien este corsé es difícil de aflojar, no es imposible. Se puede analizar la literatura del dominio para reestructurar el orden propuesto por el LEL y comprender los procesos del negocio.

También hay que tener en cuenta que la lista de escenarios candidata es dependiente de la calidad del LEL [13] y esto puede perjudicar o beneficiar el procedimiento de derivación. Un LEL más completo permitirá identificar más situaciones del dominio, aunque no necesariamente la estructura correcta.

La presente línea de investigación propone generar un nuevo mecanismo de derivación que contemple las características propias de los modelos intervinientes, utilizando todo el conocimiento ya existente del dominio, especialmente las transcripciones de las entrevistas, eliminando toda la subjetividad que tiene la derivación actual. Este nuevo mecanismo se concentrará en incluir a todos los símbolos del LEL, sin excepciones, con un adecuado tratamiento de los estados, las jerarquías conceptuales y los puntos de vista del dominio. De esta manera el conjunto de escenarios derivados serán esencialmente más

representativos de la realidad y su completitud aumentará significativamente.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con los trabajos preliminares que sirven de antecedentes al presente proyecto se han podido detectar algunas regularidades que se han transformado en hipótesis que se esperan corroborar.

La presencia de un problema en la derivación se constató empíricamente al revisar los escenarios obtenidos en varios casos reales. Esta comparación consistió en controlar el pasaje de los escenarios candidatos a los definitivos. Durante muchos años se consideró exitosa una derivación cuando existía un alto porcentaje de coincidencia. En la Tabla 1 se puede observar que los casos donde la lista final es menor a la de los candidatos, fila 1, 3 y 5, corresponden a LELs donde los impactos de los Sujetos representan acciones simples en vez de actividades. En todos los casos, sea que los escenarios se separen o unan, la coincidencia semántica entre ambas listas es muy alta. Es de destacar que los casos presentados en la Tabla 1 fueron seleccionados al azar de una lista mucho más numerosa.

Casos de Estudio	Cant. Esc. Derivados	Cant. Esc. Finales
1	16	10
2	23	25
3	21	11
4	17	18
5	28	15

Tabla 1- Diferencias entre la lista de escenarios candidatos y la lista final

Otro aspecto a observar es que la granularidad del LEL repercute en la granularidad de los escenarios e induce a un retrabajo significativo, siendo necesario juntar o separar los escenarios una vez derivados.

Una vez confirmada esta debilidad inicial de la derivación con respecto a generar una lista inicial, se pasó a analizar la semántica de los escenarios derivados con mayor profundidad, y se observó que solo se incorporaba detalle a las mismas situaciones originales. Por lo tanto, esta similitud resultó no siendo una buena indicación de la calidad del proceso. Esto fue lo que alertó sobre la posibilidad de que el problema se concentrara en la misma actividad de derivación. Se concluyó que el problema comienza con la construcción de una lista inicial de situaciones, ya que esta lista limita la elicitación de conocimiento, debido a que el ingeniero de requisitos va en busca de conocimiento con el objetivo de completar esas situaciones ya determinadas. Luego, retorna al UdeD con escenarios conformados para completarlo, y solo aparecieron nuevas situaciones en algunos casos donde existió un conocimiento previo del dominio o donde hubo un cliente-usuario en alerta.

Para asegurar que esta visión acerca de que la derivación resulta poco apropiada, se analizaron los escenarios derivados pasándolos a diferentes tipos de gráficos. En primer lugar, se realizó una red semántica donde se comprobó que los escenarios se podían mejorar incluyendo información de otros símbolos del LEL. Estos escenarios mejorados eran semánticamente similares a los de la derivación original. Esto se debe a que la red semántica utilizaba todos los símbolos del LEL pero sigue sin incorporar nueva información.

Con el uso de las redes semánticas se identificó que el problema central de los escenarios derivados es que no reflejan procesos, o lo hacen de una manera inadecuada. Esto llevó a estudiar los escenarios con representaciones gráficas de procesos, como un diagrama de actividad, BPMN y otros. Se realizó la representación de los escenarios con la derivación original. Luego, a este gráfico se le fue incorporando información del contexto, releendo las transcripciones de las entrevistas y notando

una mejora sustancial en la completitud de los escenarios construidos y en la comprensión del problema. Se pudo observar en la comparación de los escenarios derivados originalmente con los mejorados, que los escenarios derivados tenían algunos problemas de secuencia y muchas omisiones de información. Como ya se mencionó, esta falta de información en el LEL es aceptable para comprender el vocabulario del dominio, pero no lo es para los escenarios.

En esta revisión de la derivación, se pudo detectar algunas omisiones y errores en el LEL. Por lo tanto, aporta una validación diferente al glosario construido, retroalimentando conocimiento.

Estos resultados preliminares han confirmado la necesidad de modificar la heurística de derivación, que tome inicialmente el conocimiento del LEL pero que luego se nutra con toda la información que se posee, principalmente de las transcripciones de las entrevistas, para construir la primera versión de los escenarios. Este mecanismo de derivación debe estar orientado a los procesos y contemplar todas las particularidades del contexto expresadas en el LEL, asegurando un conjunto de escenarios lo más completos posible que reflejen desde el comienzo, los procesos del negocio reales.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación presentada es parte directa de la tesis doctoral de la Mg. Gladys Kaplan, colabora en la tesis de maestría de la Lic. Renata Guatelli y en las tesis doctorales de la Ing. Andrea Vera y de la Lic. María Pepe.

5. BIBLIOGRAFÍA

[1] Davis, A. M (1993), *Software Requirements: Objects, Functions and States*, Prentice Hall, Englewood Cliffs, NJ.
 [2] Elizabeth Hull, Ken Jackson, Jeremy Dick (2010), *“Requirements Engineering”*, Springer, Third Edition, ISBN-10: 1849964041, ISBN-13: 978-1849964043.

[3] Pohl, K. (2010), *“Requirements Engineering: fundamentals, principles, and techniques”*, Springer Publishing Company, Incorporated.,
 [4] Leite, J., C., S., P., Doorn, J. H., Kaplan, G., Hadad, G., D., S., Ridao, M., N., (2004) *“Defining System Context using Scenarios”*, in *Perspectives on Software Requirements*, Kluwer Academic Publishers, cap.8, pp. 169-199.
 [5] Leite J.C.S.P., Franco, A.P.M., (1990) *“O Uso de Hipertexto na Elicitação de Linguagens da Aplicação”*, Anais de IV Simpósio Brasileiro de Engenharia de Software, SBC, pp. 134-149.
 [6] Hadad, G.D.S., Doorn J.H., Kaplan G.N. (2008), *“Creating Software System Context Glossaries”*, in *Encyclopedia of Information Science and Technology*, Second edition, Mehdi Khosrow-Pour (ed.), Idea Publishing.
 [7] Leite, J. C. S., Hadad, G. D. S., Doorn, J. H., & Kaplan, G. N. (2000), *“A scenario construction process”*, *Requirements Engineering Journal*, Springer-Verlag London Ltd., Vol.5, Nº1, pp. 38-61.
 [8] Doorn J.H., Hadad G.D.S., Kaplan G.N. (2002), *“Comprendiendo el Universo de Discurso Futuro”*, WER’02: Workshop de Engenharia de Requisitos, Valencia-España.
 [9] Kaplan, G.N., Doorn, J.H., Gigante, N.C. (2013), *“Evolución Semántica de Glosarios en los Procesos de Requisitos”*, XIX Congreso Argentino de Ciencias de la Computación, CACIC 13, Mar del Plata.
 [10] Hadad, G., Doorn J., Kaplan G. (2009) *“Explicitar Requisitos de Software usando Escenarios”*, WER’09: Workshop de Engenharia de Requisitos, Medellín, Colombia.
 [11] Souza, R.A.C., Cysneiros Filho, G.A.A., Batista, G.H.C. (2015), *“A Heuristic Approach for Supporting Innovation in Requirements Engineering”*, WER’15: XIX Workshop de Engenharia de Requisitos, Perú.
 [12] Kaplan G. y Doorn J. (2017), *“Jerarquías Naturales en el Contexto del Proceso de Requisitos”*, WER’17: IXX Workshop de Engenharia de Requisitos, UCA, Bs As.
 [13] Ridao, M., Doorn, J.H. (2006), *“Estimación de Completitud en Modelos de Requisitos Basados en Lenguaje Natural”*, WER’06: IX Workshop on Requirements Engineering, Brasil, pp. 151-158.