

Mapas Conceptuales para tratar Nominalizaciones en Modelos de Requisitos

Graciela Hadad^{1,2}, Alberto Sebastián¹, Claudia Litvak^{1,2}, Jorge H. Doorn²

¹Facultad de Ingeniería y Tecnología Informática, Universidad de Belgrano
Federico Lacroze 1947 (C1426CPE), Ciudad de Buenos Aires, (011) 4511-4716

²Escuela de Informática, Universidad Nacional del Oeste
graciela.hadad@comunidad.ub.edu.ar, asebastian@outlook.com.ar,
claudia.litvak@comunidad.ub.edu.ar, jdoorn@exa.unicen.edu.ar

Resumen

La Ingeniería de Software ha experimentado una evolución en los procesos de desarrollo, focalizándose en campos como la Ingeniería de Requisitos. Uno de los pilares de este campo es la comunicación, por lo que se incentiva el uso de modelos escritos en lenguaje natural durante la definición de requisitos. Es conveniente entonces escribir estos modelos usando el propio lenguaje del universo de discurso bajo estudio. Si bien se logra así una mejora sustancial en la comunicación, surgen problemas inherentes al lenguaje natural, tales como las ambigüedades. Paralelamente, se ha comprobado, mediante estudios estadísticos de completitud, la ocurrencia de omisiones significativas en modelos del proceso de requisitos. Las actividades de verificación y validación suelen tener dificultades en la detección de estos problemas. Es por ello que se propone extender una técnica de inspección de modelos en lenguaje natural, basada en la construcción de mapas conceptuales, potencialmente útil para la detección de ciertas omisiones y ambigüedades provenientes de la nominalización de verbos, que se ha observado con

frecuencia en la comunicación y se traslada esencialmente a los modelos escritos en lenguaje natural.

Palabras clave: Ingeniería de Requisitos, Completitud de Modelos, Lenguaje Natural, Ambigüedad, Nominalizaciones.

Contexto

La propuesta que se presenta es parte de los proyectos de investigación “Estrategia para mejorar la completitud de modelos orientados al cliente en la ingeniería de requisitos” de la Universidad de Belgrano y “Tratamiento de los factores situacionales y la completitud en la ingeniería de requisitos” de la Universidad Nacional del Oeste.

Introducción

El proceso de definición de requisitos es considerado uno de los más complejos en el proceso de software [1], y es donde se presentan la mayor cantidad de defectos [2]. Este proceso se enfoca inicialmente en adquirir todo el conocimiento relevante del universo de discurso donde estará inmerso el

software, por lo que la comunicación es una actividad intensiva. Este se vale de diversas técnicas y modelos, dentro de los cuales, el Léxico Extendido del Lenguaje (LEL) permite mejorar la comunicación entre clientes e ingenieros de software [1]. Esto se debe a que se construye para representar, por un lado, y comprender, por otro, el vocabulario utilizado en el universo de discurso, y sirve de ancla para toda comunicación escrita (modelos e informes) y oral (entrevistas y exposiciones). Este modelo se suele construir al inicio del proceso de requisitos y, por lo tanto, su calidad tiene un impacto directo en los subsiguientes modelos que se construyan, como así también a lo largo de todo el proceso de desarrollo. Uno de los aspectos negativos en el proceso de requisitos son las omisiones y ambigüedades que se reflejan en los modelos [3, 4, 5] y que se manifiestan con mayor intensidad en aquellos modelos escritos en lenguaje natural [6].

En un proyecto de investigación de fines de los '90 se desarrolló un proceso de inspección para detectar defectos en el modelo LEL basado en formularios y guías de detección [7], el que se ha utilizado en cientos de casos con una tasa aceptable de detección de defectos. Sin embargo, en trabajos posteriores se puso en evidencia el bajo nivel de completitud que persistía en los LELs construidos [4, 5, 8, 9, 10, 11, 12], donde una buena parte de las omisiones correspondían a la categoría de ambigüedades. Debido a esto, se ha elaborado un proceso de inspección que se aboca principalmente a la detección de estos defectos, el cual está aún en etapa de prueba. No obstante lo cual, se ha observado a partir de resultados preliminares, que se requiere un estudio más profundo respecto a ciertos tipos de ambigüedades que suelen presentarse en estos modelos,

provenientes del uso de verbos y sus formas nominales (sustantivos) [13].

Líneas de Investigación, Desarrollo e Innovación

En el LEL se representan los términos relevantes utilizados en el universo de discurso, donde cada término se describe a través de los componentes: *nombre* formado por una palabra o frase (puede ser más de un nombre en caso de sinónimos), *noción* que describe el significado del término e *impacto* que describe cómo repercute en el universo de discurso. Estos términos se clasifican en 4 *tipos*: sujetos, objetos, verbos y estados.

A través de los trabajos realizados sobre completitud del modelo LEL mencionados anteriormente, se han observado algunas dificultades en el tratamiento de los verbos, ya que algunos ingenieros de software describen las actividades que se desarrollan en el universo de discurso utilizando la forma verbal y otros la forma nominal, lo que se denomina en lingüística *nominalización* del verbo [13]. Esta nominalización no depende estrictamente de aspectos cognitivos del ingeniero sino también del propio uso del término en el universo de discurso; cualquiera sea el caso la nominalización o no, puede provocar ambigüedades. La nominalización consiste en construir sustantivos a partir de verbos o de adjetivos [14, 15, 16, 17, 18]. Las diferencias de uso de estas derivaciones en los verbos pueden ocultar la existencia de resultados o efectos provenientes de dicha actividad, y/o pueden producir una mala interpretación en el uso nominal del verbo. En otras palabras, la nominalización de un verbo puede describir una *acción* o un *efecto* de dicha acción, y en ciertos casos esa nominalización puede involucrar más de

un significado: la *acción* y el *efecto*, es decir, se está en presencia de homónimos. Un ejemplo de nominalización del verbo “Facturar”, extraído de un LEL desarrollado para un instituto de salud, es el término “Facturación Mensual”, que podría interpretarse como la actividad de generar facturas o bien como el producto de esa actividad. Si este fenómeno es ignorado puede ocurrir y ocurre frecuentemente que quien construya el LEL sólo perciba uno de los significados posibles y en consecuencia omita impactos que luego redundarán en omisiones en otros modelos derivados de él y, finalmente, en omisiones en los requisitos del sistema. En la Figura 1, se presenta la definición de dicho término, donde en su contexto es utilizado como el efecto/resultado de la acción y, por ello, es clasificado como *objeto*. Asimismo, un verbo puede desencadenar en más de una nominalización. Por ejemplo: i) Envolver: Envoltura–Envoltorio; ii) Agrupar: Agrupación–Agrupamiento. Se debería evaluar qué nominalizaciones son usadas en un universo de discurso específico, a fin de evitar omisiones de términos en el LEL y el uso ambiguo de términos nominalizados.

<p>Nombre: FACTURACIÓN MENSUAL Tipo: Objeto Noción:</p> <ul style="list-style-type: none"> • Son los servicios prestados por el Instituto al prestador, en valores monetarios <p>Impacto:</p> <ul style="list-style-type: none"> • El Instituto la genera como resultado del proceso de facturar en el período de facturación correspondiente • Contiene el detalle de cada factura emitida
--

Figura 1. Término del LEL nominalizado

Por otro lado, en un proyecto de investigación en curso se ha propuesto un proceso de inspección del LEL basado en

la construcción de mapas conceptuales, que está fuertemente centrado en la detección de *omisiones* y *ambigüedades* [19]. Básicamente, consiste en construir un mapa conceptual [20, 21] de cada término definido en el LEL, donde las oraciones en cada definición son proposiciones en el mapa (ver ejemplo de la Figura 2 que representa el término descrito en la Figura 1). Luego, se sigue una secuencia de pasos para estudiar cada mapa y el conjunto de ellos, mediante un análisis visual, estadístico y semántico de conceptos que no son términos del LEL, del uso de conceptos que sí son términos del LEL y de las relaciones entre conceptos, a fin de identificar conceptos candidatos a términos del LEL, sinónimos no detectados, proposiciones ambiguas, omisiones menores de información y omisiones de términos candidatos del tipo verbo. Este proceso de inspección está en una etapa de prueba, llevándose a cabo un experimento controlado. No se ha incluido específicamente el tema de nominalizaciones en la inspección, por lo que se considera que sería conveniente estudiar la posibilidad de definir algunas guías de detección a través de los mapas conceptuales.

Resultados y Objetivos

Las consecuencias del uso no controlado de las nominalizaciones pueden producir un deterioro del proceso de requisitos, a través de una comunicación deficiente entre los clientes y los ingenieros.

El estudio de las nominalizaciones de verbos para ser representadas en el LEL ha llevado a establecer los siguientes casos:

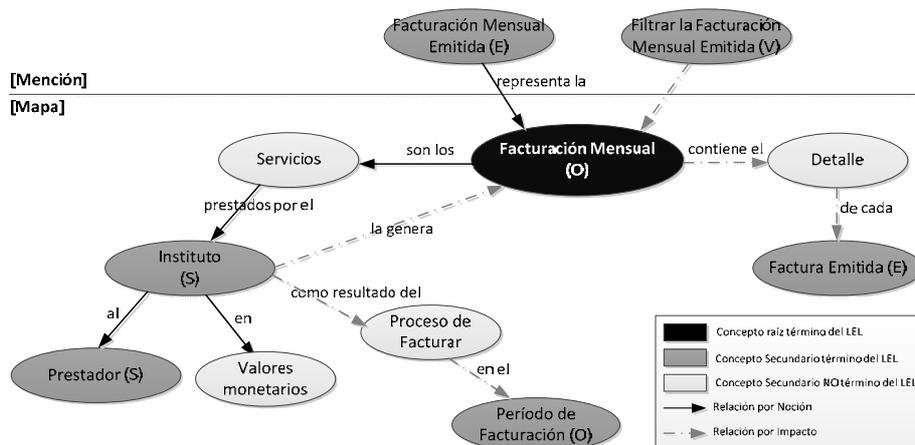


Figura 2. Mapa conceptual del término “Facturación Mensual”

- Se usa solo el verbo en el contexto: registrar el término tipo verbo en el LEL.
- Se usa solo la nominalización, siendo las posibilidades: a) representa solo la *acción*: registrar el término tipo verbo, b) representa solo el *efecto*: registrar el término tipo objeto o estado, o c) representa *acción y efecto*: registrar término tipo verbo y término tipo objeto o estado, siendo éstos homónimos.
- Se usa la forma verbal y la nominalización, siendo las posibilidades: a) representan solo la *acción*: registrar el término tipo verbo con el nombre de ambas formas (sinónimos), o b) representan *acción y efecto*: registrar ambas formas como término verbo y como término objeto o estado, considerando que la nominalización puede ser además un sinónimo de la forma verbal, produciendo entonces entradas al LEL homónimas.

Al elaborar el proceso de inspección del LEL usando mapas conceptuales se observó cierta dificultad en el tratamiento de conceptos y relaciones cuando el concepto era en sí mismo un verbo nominalizado. Esto ha llevado a la idea de realizar un estudio de las

nominalizaciones mediante mapas conceptuales, y cómo estos podrían ayudar a descubrir ambigüedades u omisiones de algunos términos relevantes en el universo de discurso debido a una interpretación deficiente de los verbos nominalizados o una visión parcial de los mismos. Las relaciones entre conceptos en un mapa conceptual aparentan ser más ricas de lo que a simple vista se aprecia [21] cuando los mapas se construyen a partir de un LEL. Esta riqueza podría inducir a descubrir principalmente verbos ignorados y, por otro lado, a cuestionar las nominalizaciones en los conceptos y su semántica en las proposiciones, de manera de llevar a capturar la existencia en el universo de discurso de su forma verbal o de otra semántica oculta.

Se considera que la detección de omisiones y ambigüedades a través de mapas conceptuales y la identificación de sus causas podrían derivar en nuevas heurísticas que colaboren en mejorar el nivel de completitud del modelo LEL.

Formación de Recursos Humanos

En el proyecto de la Universidad de Belgrano participan tres investigadores, uno de ellos en formación, mientras que el proyecto de la Universidad Nacional

del Oeste está integrado por cinco investigadores, uno de ellos en formación, y dos alumnos becarios.

La propuesta presentada es parte del Trabajo Final de Posgrado “Verificación de un modelo Léxico Extendido del Lenguaje con Mapas Conceptuales”, del Lic. Alberto Sebastián que está finalizando en la UCA, y de la Tesis de Doctorado “Gestión de la completitud en la Ingeniería de Requisitos”, de la Mg. Claudia Litvak que está desarrollando en la UNLP.

Referencias

- [1] Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridao, M.N. (2004). Defining System Context using Scenarios. En *Perspectives on Software Requirements*, Springer US, pp. 169-199.
- [2] Westland, J.C. (2002). The cost of errors in software development: evidence from industry. *Journal of Systems and Software*, vol.62, pp.1-9.
- [3] Firesmith, D. (2005). Are Your Requirements Complete? *Journal of Object Technology*, vol.4, nº1, pp.27-43.
- [4] Doorn, J.H., Ridao, M. (2003). Completitud de Glosarios: Un Estudio Experimental. 6th Workshop on Requirements Engineering Paracicaba, Brasil: Universidade Metodista de Piracicaba, pp. 317-328.
- [5] Hadad, G.D., Litvak, C., Doorn, J.H., Ridao, M. (2015). Dealing with Completeness in Requirements Engineering. *Encyclopedia of Information Science and Technology*, Third Edition. IGI Global, Mehdi Khosrow-Pour (ed), Information Science Reference, Hershey, pp.2854-2863.
- [6] Berry, D., Kamsties, E.(2004). Ambiguity in requirements specification. *Perspectives on software requirements*, Springer US, pp. 7-44.
- [7] Kaplan, G.N, Hadad, G., Doorn, J., Leite, J. (2000). Inspección del Léxico Extendido del Lenguaje. 3rd Workshop on Requirements Engineering, Río de Janeiro, pp.70-91.
- [8] Doorn, J.H., Ridao, M.N. (2009). Completeness Concerns in Requirement Engineering. *Encyclopedia of Information Science and Technology*, Second Edition. IGI Global, Mehdi Khosrow-Pour (ed.). Information Science Reference, pp.619-624.
- [9] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2012). Un abordaje al problema de completitud en requisitos de software. XVIII Congreso Argentino de Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca, pp. 827-836.
- [10] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2013). Correcciones semánticas en métodos de estimación de completitud de modelos en lenguaje natural. 16th Workshop on Requirements Engineering. Universidad ORT, Montevideo, pp. 105-117.
- [11] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2013). Mejoras semánticas para estimar la completitud de modelos en lenguaje natural. 1er CONAIIISI. Universidad Tecnológica Nacional, Córdoba.
- [12] Hadad, G.D.S., Litvak, C.S., Doorn, J.H. (2014). Problemas y Soluciones en la Completitud de Modelos en Lenguaje Natural. II Congreso Argentino de Ingeniería, Universidad Nacional de Tucumán, San Miguel de Tucumán, T366.
- [13] Real Academia Española y Asociación de Academias de la Lengua Española (2009). *Nueva gramática de la lengua española*, Madrid, Espasa Libros.
- [14] Rozwadowska, B. (1997). Towards a unified theory of nominalizations: External and internal eventualities (No. 2031). *Wdawn. Uniwersytetu Wrocławskiego*.
- [15] Alexiadou, A. (2001). *Functional Structure in Nominals. Nominalization and ergativity*. Amsterdam. John Benjamins Publishers.
- [16] Azpiazu, S. (2002). Las estrategias de nominalización y el adverbio: estudio contrastivo del caso español. Ediciones Universidad de Salamanca, España.
- [17] Bisetto A., Melloni C. (2005). Result Nominals: a Lexical-Semantic Investigation. Geert Booij et al. (eds.). *Fifth Mediterranean Morphology Meeting*, Fréjus, University of Bologna, pp. 393-412.
- [18] Rothmayr, A. (2009). *The structure of stative verbs (Vol.143)*. John Benjamins Publishing.
- [19] Sebastián, A., Hadad, G.D.S. (2015). Mejoras a un Modelo Léxico mediante Mapas Conceptuales. XXI CACIC, Universidad Nacional del Noroeste, Junín.
- [20] Novak, J., Gowin, D.B. (1988). *Aprendiendo a aprender*. Edic. Martínez Roca, Barcelona.
- [21] Novak, J., Cañas, A. (2008). *The Theory Underlying Concept Maps and How to Construct and Use Them*. Technical Report, Florida Institute for Human and Machine Cognition.