

Procesamiento de Lenguaje Natural para Estudiar Completitud de Requisitos

Claudia S. Litvak^{1,3}, Graciela D. S. Hadad^{2,3}, Jorge H. Doorn^{1,2}

¹Departamento de Ingeniería e Investigaciones Tecnológicas, UNLaM
Florencio Varela 1903. San Justo. Buenos Aires. 4480-8900

²Escuela de Informática, UNO

³Facultad de Ingeniería y Tecnología Informática, UB
e-mail: claudialitvak@gmail.com, gracielahadad@gmail.com, jdoorn@exa.unicen.edu.ar

Resumen

La Ingeniería de Software incluye el estudio inicial de los requisitos del sistema de software a desarrollar. A ello se dedican los Ingenieros de Requisitos, generando modelos que logren representar las necesidades de los clientes y usuarios y las soluciones a las mismas. El desarrollo de modelos de requisitos en lenguaje natural facilita la interacción entre todos los involucrados, aunque contribuye a generar ambigüedades.

Por otro lado se ha comprobado la existencia de omisiones en una cantidad muy superior a la imaginada intuitivamente. Por su propia naturaleza, las omisiones son notoriamente elusivas. El desarrollo de un mismo modelo acerca del mismo problema por personas diferentes es la forma más eficaz para vislumbrar la magnitud de las omisiones en el dominio de la Ingeniería de Requisitos y posiblemente en muchas otras áreas. Para realizar estudios sobre la completitud de los modelos desarrollados, a fin de mejorar la calidad de los mismos, se ha propuesto la generación de una herramienta de software que permita comparar modelos escritos en lenguaje natural. Se propone que esta herramienta

utilice recursos del Procesamiento del Lenguaje Natural.

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

Contexto

Esta propuesta está incluida en los proyectos “Control de Calidad de requisitos escritos en Lenguaje Natural” de la Universidad Nacional de la Matanza (UNLaM) y “Tratamiento de los factores situacionales y la completitud en la ingeniería de requisitos” de la Universidad Nacional del Oeste (UNO).

Además esta propuesta es parte del plan de trabajo de becarios alumnos de UNLaM.

Introducción

La Ingeniería de Requisitos tiene como objetivo realizar la definición de los requisitos de un sistema de software, lo más completa posible. Ello implica la mejora de la calidad de los requisitos, dado que la completitud se considera parte de la calidad de los mismos [IEEE 29148]. Diferentes autores han planteado que la incompletitud es uno de los

principales problemas que afronta la Ingeniería de Requisitos [Kotonya 98] [Loucououlos 95] [Firesmith 05].

Los requisitos se plasman mediante diferentes modelos, en particular, el modelo donde se encuadra este proyecto es el modelo presentado por Leite et al. [Leite95] [Leite97] [Leite00] [Leite01] [Leite04] donde se construyen los requisitos a partir de los modelos Léxico Extendido del Lenguaje (LEL) y Escenarios. Dichos modelos se encuentran escritos en lenguaje natural, facilitando la comprensión e interacción con todos los involucrados. En el caso del modelo LEL, este es un glosario de los términos del dominio de la aplicación, mientras que los escenarios describen situaciones de dicho dominio. En el modelo LEL, para cada término de dicho glosario se definen su nombre, conformado por una palabra o frase, con todos los sinónimos utilizados. El significado del término se lo describe a través de su noción y su impacto. La noción se refiere a la denotación del término, mientras que el impacto describe su connotación en dicho dominio.

Diferentes estudios previos han atacado el tema de la completitud en modelos de requisitos [Doorn 03] [Ridao 06] [Hadad 12] [Hadad 13] [Litvak 12] [Litvak 13a] [Litvak 13b] [Litvak 14] [Litvak 15], pero los progresos obtenidos han sido limitados. En una primera etapa la comparación de los modelos se realizó utilizando una granularidad gruesa, ya que se compararon los nombres de los símbolos. Rápidamente esto permitió observar que los resultados obtenidos habían detectado omisiones significativas, pero simultáneamente se habían creado omisiones aparentes. Esta realidad forzó a reducir la granularidad de las comparaciones a sentencias atómicas. Naturalmente este cambio a su vez desencadenó una escalada exponencial en

las dificultades lingüísticas de las comparaciones. Abordar estas dificultades requiere atender específicamente varios inconvenientes:

i) La presencia de sujetos tácitos: los sujetos tácitos deben ser completados a sujetos presentes, dependiendo del símbolo y del contexto (ver Tabla 1).

Tabla 1. Modificación de oraciones con sujeto tácito

| Definición del LEL | Ejemplo |
|---|--|
| Símbolo | Solicitud de adhesión / Solicitud de ahorro |
| Noción | Es un formulario que completa el solicitante para ser adherente de un grupo de plan de ahorro. |
| Oración obtenida sin sujeto tácito | Solicitud de adhesión es un formulario que completa el solicitante para ser adherente de un grupo de plan de ahorro. |

Tabla 2. Desdoblamiento de oraciones compuestas

| Definición del LEL | Ejemplo |
|--|--|
| Símbolo | Adherente /No adjudicatario |
| Noción | Es un solicitante al que la administradora le acepta la solicitud de adhesión. |
| Oración obtenidas sin sujeto tácito | Adherente /No adjudicatario es un solicitante al que la administradora le acepta la solicitud de adhesión. |
| Oraciones finales obtenidas | a) Adherente /No adjudicatario es un solicitante. |
| | b) Administradora acepta la solicitud de adhesión al Adherente. |

ii) La presencia de oraciones compuestas, que deben desdoblarse en varias oraciones simples. Si bien se propone al presentar la definición del LEL que se utilicen en la noción y el impacto oraciones simples, en la práctica muchas de ellas suelen ser oraciones compuestas (ver Tabla 2).

iii) La existencia de oraciones mal ubicadas, colocadas en un término al cual no pertenecen.

La propuesta es entonces la generación de una herramienta automática o semiautomática de comparación entre modelos, pensada inicialmente para el modelo LEL.

Líneas de Investigación, Desarrollo e Innovación

Al intentar generar una herramienta semiautomática de comparación de oraciones, resulta imperioso recurrir a estrategias de Procesamiento del Lenguaje Natural [Alonso 05].

Dichas herramientas realizan el reconocimiento del idioma utilizado, la segmentación de oraciones, palabras y secciones, y luego, dependiendo de la herramienta, un análisis morfológico, sintáctico y/o semántico [Alonso 05]. La herramienta a ser construida deberá ser semiautomática, dado que se espera que algunas oraciones no podrán ser apareadas automáticamente y deberán ser comparadas manualmente.

Como trabajo futuro se propone además aplicar esta herramienta para hacer estudios de completitud sobre varios casos de estudio, aplicando técnicas estadísticas a los resultados obtenidos, al comparar los LEL realizados por ingenieros de requisitos independientes. Asimismo, extender la herramienta para la comparación del modelo de Escenarios, e incluso a los requisitos mismos.

Resultados y Objetivos

Hasta la fecha se han estudiado diferentes trabajos y herramientas sobre Procesamiento de Lenguaje Natural [Alonso 05] [Vallez 07] [VilaresFerro 05] [Méndez 99].

Se planifica que la herramienta produzca un documento, con las oraciones apareadas entre ambos LEL y las oraciones de cada uno de los LEL que no pudieron aparearse, facilitando el apareamiento manual de estas últimas.

Se propone, luego de utilizar la herramienta de comparación, realizar estudios estadísticos para evaluar la completitud de los modelos [Wholin 98] [Doorn 03] [Ridao 06] [Hadam 12] [Hadam 13] [Litvak 12] [Litvak 13a] [Litvak 13b] [Litvak 14]. Se espera con ello eliminar las omisiones aparentes y comprender mejor las causas que originan las omisiones con el fin, a su vez, de mejorar las heurísticas de creación y verificación del LEL.

Formación de Recursos Humanos

El proyecto desarrollado en la Universidad Nacional de La Matanza, tiene ya más de 10 años. Está siendo desarrollado por un grupo de investigación del área de Ingeniería de Requisitos, perteneciente al Departamento de Ingeniería e Investigaciones Tecnológicas. Una de las líneas actuales de investigación de este grupo es la completitud.

El proyecto actual de la Universidad Nacional del Oeste, reside en el LITIC, creado en 2014. Cuenta con cinco investigadores, y uno de los temas abordados es la completitud. A mediados de 2015 se han incorporado dos becarios al plantel.

Además, este trabajo forma parte de la Tesis de Doctorado que la Magíster Claudia Silvia Litvak está desarrollando en UNLP, la cual trata sobre la completitud de los modelos de requisitos, denominada “Gestión de la completitud en la Ingeniería de Requisitos”.

Se incorporaron al proyecto de investigación de UNLaM durante 2015

seis alumnos de cuarto año de la carrera Ingeniería Informática de UNLaM. Los mismos son los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice, Silvina Cáceres y Belén Ortiz. Dadas estas incorporaciones, se han agregado al proyecto un par de objetivos específicos relacionados con el uso de herramientas de Procesamiento de Lenguaje Natural en el análisis y construcción de los documentos del proceso de requisitos. En particular, los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice fueron seleccionados por la Universidad para recibir Becas UNLaM, que están cumpliendo desde el 1° de septiembre de 2015. La propuesta de las becas involucra que el alumno Facundo Velazquez Santillán, tenga como objetivo la creación de una herramienta informática que facilite la creación de modelos de requisitos de software escritos en lenguaje natural, en particular los modelos LEL y Escenarios. Los alumnos Ximena Milla y Diego Lo Giudice, tienen como objetivo la construcción de una herramienta que permita la comparación de modelos de requisitos escritos en lenguaje natural, en particular para los modelos LEL y Escenarios, a fin de poder facilitar los estudios de completitud. En el estudio de las necesidades de ambas aplicaciones a desarrollar, ha surgido el primordial interés en la investigación de herramientas de Procesamiento de Lenguaje Natural, que puedan aplicarse, en particular aquellas que sean software libre. Las alumnas Silvina Cáceres y Belén Ortiz se han dedicado inicialmente al estudio de las técnicas de Procesamiento de Lenguaje Natural, y están avocadas a la búsqueda y evaluación de herramientas de software libre de Procesamiento de Lenguaje Natural, así como la factibilidad de ser aplicadas en el contexto de investigación

de los becarios, preferentemente de código abierto.

Referencias

[Alonso 05] Alonso i Alemany, L. (2005). Herramientas Libres para Procesamiento del Lenguaje Natural. 5tas Jornadas Regionales de Software Libre, Rosario, Argentina.

<http://www.cs.famaf.unc.edu.ar/~laura/freeNLP>.

[Doorn 03] 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.

[Firesmith 05] Firesmith D.G. (2005). Are Your Requirements Complete? in Journal of Object Technology, vol. 4, no. 1, January-February 2005, pp. 27-43.

http://www.jot.fm/issues/issue_2005_01/column3

[Hadad 12] Hadad G.D.S., Litvak C.S., Doorn J.H. (2012). Agregando semántica a técnicas de predicción de completitud en modelos de requisitos. WICC 2012 - XIV Workshop de Investigadores en Ciencias de la Computación, anales electrónicos, Universidad Nacional de Misiones, Posadas, ISBN: 978-950-766-082-5, pp. 392-396, <http://hdl.handle.net/10915/18863>.

[Hadad 13] Hadad, G.D.S., Litvak, C.S., Doorn, J.H., Ridao M.N. (2013). Dealing with Completeness in Requirements Engineering. Encyclopedia of Information Science and Technology, Third Edition. Editorial: IGI Global, Mehdi Khosrow-Pour (ed), Information ScienceReference, Hershey, PA, EEUU, 3° edición, cap. 279, pp. 2854-2863, Julio 2014 (10384 p.). ISBN13: 9781466658882.

[IEEE 29148] IEEE Systems and software engineering - Life cycle processes - Requirements engineering. IEEE, Nueva York (2011).

[Kotonya 98] Kotonya G., Sommerville I. (1998). Requirements engineering – processes and techniques, John Wiley & Sons UK.

[Leite 95] Leite, J.C.S.P., Oliveira, A.P.A. (1995). A Client Oriented Requirements Baseline. Second IEEE International Symposium on Requirements Engineering, IEEE Computer Society Press, pp. 108-115.

[Leite 97] Leite, J.C.S.P., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G., Hadad, G., Oliveros, A.

(1997). Enhancing a Requirements Baseline with Scenarios. *Requirements Engineering Journal*, Springer-Verlag London Ltd., Vol. 2, N° 4, pp. 184-198.

[Leite 00] Leite J.C.S.P., Hadad G.D.S., Doorn J.H., Kaplan G.N. (2000). A Scenario Construction Process. *Requirements Engineering Journal*, ISSN: 0947-3602, Vol.5, N° 1, pp. 38-61, Springer-Verlag London Ltd., Londres, Reino Unido, Julio 2000. <http://link.springer.com/article/10.1007/PL00010342>

[Leite 01] Leite, J.C.S.P. (2001). Gerenciando a Qualidade de Software com Base em Requisitos. *Qualidade de Software: Teoria e Prática*. Prentice-Hall, Rocha A, Maldonado J, Weber K (eds), Cap. 17, pp. 238-246.

[Leite 04] 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 el libro *Perspectives on Software Requirements*, Kluwer Academic Publishers, EEUU, ISBN: 1-4020-7625-8, Cap. 8, pp. 169-199.

[Litvak 12] Litvak C.S., Hadad G.D.S., Doorn J.H. (2012). Un abordaje al problema de completitud en requisitos de software. CACIC 2012 – XVIII Congreso Argentino de Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca, ISBN: 978-987-1648-34-4, pp. 827-836, Octubre 2012. <http://hdl.handle.net/10915/23715>.

[Litvak 13a] 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 Montevideo, Uruguay: ORT University, pp. 105-117.

[Litvak 13b] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2013). Mejoras semánticas para estimar la completitud de modelos en lenguaje natural. CoNaIISI 2013, 1er Congreso Nacional en Ingeniería Informática / Sistemas de Información. Córdoba, Argentina: Universidad Tecnológica Nacional.

[Litvak 14] Litvak C.S., Hadad G.D.S., Doorn J.H. (2014). Heurísticas para el modelado de requisitos escritos en lenguaje natural. CACIC 2014 - XX Congreso Argentino de Ciencias de la Computación, Universidad Nacional de La Matanza, Buenos Aires, ISBN: 978-987-3806-05-6, Octubre 2014, <http://sedici.unlp.edu.ar/handle/10915/42337>.

[Litvak 15] Litvak C.S., Hadad G.D.S, Doorn J.H. (2015). Influencia de las nominalizaciones sobre la

completitud de modelos de requisitos. WICC 2015 - XVII Workshop de Investigadores en Ciencias de la Computación, anales electrónicos, Universidad Nacional de Salta, Salta, ISBN: 978-987-633-134-0,

[Loucopoulos 95] Loucopoulos P., Karakostas V. (1995). *System requirements engineering*. McGraw-Hill, Book company Europe.

[Méndez 99] Méndez, E., Moreira González, J.A. (1999). Lenguaje natural e indización automatizada. *Ciencias de la Información*, 30(3), pp. 11-24.

[Ridao 06] Ridao, M.N., Doorn, J.H. (2006). Estimación de Completitud en Modelos de Requisitos Basados en Lenguaje Natural. 9th Workshop on Requirements Engineering. Rio de Janeiro, Brasil: Pontificia Universidade Católica do Rio de Janeiro, pp. 151-158.

[Vallez 07] Vallez M. y Pedraza-Jimenez R. (2007). El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines [en línea]. "Hipertext.net", núm. 5. <https://www.upf.edu/hipertextnet/numero-5/pln.html>.

[VilaresFerro 05] Vilares Ferro, J. (2005). Aplicaciones del procesamiento del lenguaje natural en la recuperación de información en español. <http://coleweb.dc.fi.udc.es/cole/library/ps/Vil2005a.pdf>

[Wohlin 98] Wohlin, C., Runeson, P. (1998). Defect content estimations from Review Data. 20th Intl Conference on Software Engineering, Japón, pp. 400-409.