

# Mejoras semánticas para estimar la Completitud de Modelos en Lenguaje Natural

Litvak, Claudia S.<sup>1,2</sup>, Hadad Graciela D. S.<sup>1,2</sup>, Doorn Jorge H.<sup>2,3</sup>

<sup>1</sup>Universidad de Belgrano, Facultad de Ingeniería y Tecnología Informática

<sup>2</sup>Universidad Nacional de La Matanza, DIIT

<sup>3</sup>Universidad Nacional del Centro de la Provincia de Buenos Aires, Fac. Ciencias Exactas

## Abstract

*La Ingeniería de Requisitos tiene como objetivo producir requisitos de alta calidad, siendo la completitud un aspecto crítico. ¿Se ha elicitado y modelado la suficiente información para construir un sistema de software que cubra las expectativas y necesidades del cliente? Es una pregunta de difícil respuesta a pesar de algunos intentos que se están realizando en ese sentido. Un estudio de la estimación del tamaño de modelos de requisitos escritos en lenguaje natural ha dado resultados prometedores, aunque se consideraron sólo los aspectos formales y cuantitativos. Debido a la naturaleza de estos modelos, se ha introducido un análisis semántico que se realiza previo a la aplicación del método estimativo de tamaño. Todos los trabajos de estimación de tamaño, tanto con o sin análisis semántico previo, se realizaron utilizando una adaptación del método de captura-recaptura. En este artículo se presenta un nuevo estudio de estimación de tamaño sobre un modelo generado siguiendo un proceso de construcción más refinado que el utilizado en estudios anteriores. Se pudo observar que pese a las mejoras incorporadas y varios estudios realizados, persiste aún una gran incompletitud en los modelos elaborados. Se considera que es probable que sea necesario incrementar la profundidad del análisis semántico para mejorar la estimación de completitud.*

## Palabras Clave

modelado de requisitos, elicitación, completitud de requisitos, métodos de captura-recaptura, modelos en lenguaje natural

## Introducción

La incompletitud es uno de los principales problemas en el proceso de producción de requisitos [1] [2] [3]. Ningún método en Ingeniería de Software puede asegurar que se ha elicitado y modelado suficiente información para desarrollar un sistema de software que cubra las expectativas y necesidades reales de los clientes y usuarios. La incompletitud influye

negativamente en la calidad de cualquier artefacto producido, ya sea un modelo de requisitos, un modelo de diseño o un componente de software. Adicionalmente, la incompletitud en la adquisición de conocimiento puede transferirse de una actividad a lo largo del desarrollo de software sin ser notada.

Como una hipótesis a priori, los ingenieros de requisitos presuponen que al realizar la elicitación capturan toda la información pertinente al contexto donde el software se va a implantar y no habrá faltantes. Dado que en la literatura sobre requisitos de software es poco lo que se menciona acerca de la completitud, los ingenieros pueden suponer que esto no es un problema. Sin embargo, en estudios realizados se comprobó que grupos diferentes de ingenieros trabajando sobre el mismo problema encuentran en el mejor de los casos sólo el 50% de los elementos del modelo [4] [5].

El proceso de construcción de requisitos es inherentemente incompleto debido a la complejidad del mundo real [6], por lo tanto, el principio guía de la Ingeniería de Requisitos debería ser producir un modelo lo más completo posible. Sin embargo, la expresión "lo más completo posible" puede dar lugar a desestimar la incompletitud como un problema real. Esto se manifiesta por la falta de preocupación en la literatura sobre este tema.

Gran parte de las técnicas que se ocupan de la incompletitud de requisitos se enfocan parcialmente en algunos aspectos de este problema fundamental, sin una visión integral. Ejemplos de estas trabajos son el

análisis de obstáculos propuesto por Alrajeh et al. [7], el análisis de puntos de vista propuesto por Leite y Freeman [8], la definición evolutiva de requisitos de Zowghi y Gervasi [9], y la definición y verificación determinística de requisitos propuesta por Carson [10], entre otros.

Alrajeh et al. [7] proponen producir un conjunto completo de requisitos operacionales mediante la identificación y resolución de todos los obstáculos a objetivos. Plantean que la incompletitud es causada por la falta de identificación de todas las condiciones excepcionales, las cuales son obstáculos al objetivo, aunque esta identificación está sesgada por lo que se sabe sobre el contexto del problema. Leite y Freeman [8] proponen reducir la incompletitud mediante la integración de puntos de vista. Cada punto de vista está representado por un conjunto de modelos producidos por un ingeniero de requisitos, y la integración de estos puntos de vista decremента la incompletitud individual. Zowghi y Gervasi [9] proponen alcanzar la completitud de los requisitos de forma evolutiva mediante mejoras continuas de las descripciones del contexto y de requisitos, y la verificación continua de la especificación contra el modelo de dominio. La propuesta de Carlson [10] establece que la completitud de requisitos se logra cuando los ingenieros se basan en descripciones del contexto correctas y completas. En estas propuestas se sostiene la completitud en base al conocimiento del contexto del problema, sin embargo cabe preguntarse cómo se adquiere un conocimiento “completo” del contexto.

Aplicar técnicas de elicitación apropiadas permite una adquisición más completa de conocimiento [11]; algunas técnicas de modelado ponen en evidencia la falta de información [12]; las técnicas de verificación de requisitos ayudan a detectar ciertas omisiones [13], aunque las técnicas de validación dan menos soporte en ese sentido [14]; la utilización de reglas de parada facilita la toma de decisiones acerca de cuándo se asume haber completado una

actividad o producto para seguir avanzando [15] [16]. A pesar de la existencia de estas técnicas, ellas no permiten eliminar el problema ni estimar el grado de completitud. Establecer si se ha alcanzado un grado razonable de completitud es crucial para decidir avanzar en el proceso de desarrollo de software.

Se supone que si dos ingenieros aplican de forma independiente el mismo proceso de Ingeniería de Requisitos sobre el mismo contexto del problema, entonces ambos van a elicitar y modelar la misma información. Sin embargo, la evidencia obtenida no es compatible con esta hipótesis [4]. Además al estimar el tamaño del modelo construido por cada ingeniero, se observó el bajo nivel de completitud alcanzado [5] [17]. Si bien la completitud se incrementa a medida que avanza el proyecto, mejorarla en las etapas iniciales impacta positivamente en todas las etapas siguientes.

El trabajo de investigación se ha centrado inicialmente en el estudio de la completitud de un modelo denominado Léxico Extendido del Lenguaje (LEL) [20], que describe el vocabulario utilizado en el contexto de un problema. Se comenzó el estudio utilizando este modelo debido a que es el primer modelo que debería construirse en una estrategia de requisitos y, además, se buscará en un trabajo posterior establecer si el nivel de completitud de este modelo se arrastra en modelos derivados de él.

En el presente artículo, se analizan entonces los resultados obtenidos en trabajos previos sobre la estimación de tamaño del modelo LEL, utilizando nueve muestras pre-existentes del mismo, y se comparan con nuevas estimaciones basadas en tres nuevas muestras del mismo modelo.

Se supuso que estas tres nuevas muestras eran más completas que las anteriores, ya que fueron obtenidas siguiendo un proceso de construcción más refinado, incluyendo una inspección de cada muestra. Sin embargo, los resultados no confirman firmemente la hipótesis.

## **Elementos del Trabajo y Metodología**

Para estimar el tamaño de un modelo, se ha utilizado un método predictivo de captura-recaptura [18] adaptado por Wohlin et al. [19], denominado Detection Profile Method (DPM). El método requiere disponer de al menos dos muestras del mismo modelo. Este método fue aplicado inicialmente por Doorn y Ridao [5], sobre nueve muestras del modelo LEL. Este modelo contiene los términos (símbolos) utilizados en el contexto del problema, donde cada término se define con uno o varios nombres, sus connotaciones y denotaciones. Las nueve muestras del LEL se construyeron sobre un mismo contexto: el Sistema de Planes de Ahorro Previo para la Adquisición de Vehículos 0Km [21].

Doorn y Ridao [5] concluyeron que cuando el número de elicitadores independientes crece, la diferencia entre el número de elementos elicitados y el número de elementos estimados se reduce. Del análisis posterior de sus resultados, surgió que el número de elementos faltantes para cada muestra era significativamente importante. El grupo de elicitadores que descubrió el mayor número de elementos del modelo sólo alcanzó un nivel de completitud del 50% del total estimado. A primera vista, parece que el modelo construido por cada grupo elicitador fue bastante pobre.

Por lo tanto, se hizo un nuevo estudio bajo la hipótesis de que los grupos de elicitadores estaban observando contextos diferentes o tenían distintos puntos de vista al mirar el mismo contexto [4]. Para facilitar el análisis estadístico, se dividió el problema objeto de estudio en cinco subáreas y se confirmó estadísticamente que en cuatro subáreas los grupos observaban contextos con diferentes límites, mientras que en una sola subárea todos estaban observando el mismo contexto. Estas conclusiones llevan a formular una nueva hipótesis con la intención de unificar los diferentes límites del contexto mediante el estudio del contenido textual de las muestras del modelo. Es decir, se conjeturó sobre la

posibilidad de no estar identificándose apropiadamente los elementos de las muestras del modelo al aplicar el método de captura-recaptura DPM.

En un trabajo posterior [17] se determinó que para aplicar este método predictivo se requería realizar previamente un ajuste semántico sobre el conjunto de muestras dado que la estimación de tamaño se realizaba sobre modelos escritos en lenguaje natural. Esto implicaba la dificultad en establecer la igualdad de elementos entre las distintas muestras disponibles del modelo.

El análisis semántico se realizó en dos etapas: una primera etapa entre los elementos de una misma muestra (intra-análisis), estudiando pertenencia, relevancia y redundancia de cada símbolo, y una segunda etapa comparando los elementos entre las muestras (inter-análisis), estudiando sinónimos y homónimos.

La pertenencia de un símbolo en una muestra se analizó en base a la existencia del nombre del símbolo en el contexto, y a establecer si la información contenida en la denotación y connotación estaba dentro del límite del problema. La relevancia de cada símbolo implicó analizar su denotación y connotación para establecer si presentaban un exceso de nivel de detalle, innecesario para un glosario, cuyos contenidos podían ser mencionados dentro de otro símbolo de la muestra. La redundancia de un símbolo se estableció analizando si la información en su denotación y connotación ya estaba contenida en otros símbolos de la misma muestra.

El inter-análisis semántico estableció primero la correspondencia entre símbolos de distintas muestras a través de nombres idénticos o nombres con variaciones gramaticales. Luego dicha correspondencia se confirmaba por la presencia de contenidos similares en sus respectivas denotaciones y connotaciones (detección de sinónimos). En el caso de diferencias entre contenidos, se analizaba la posibilidad de homónimos. En un paso siguiente del inter-análisis, se establecían otros sinónimos por

contenidos similares de sus denotaciones y connotaciones pero sin presentar nombres idénticos.

Entonces, la corrección semántica implicó el descarte de elementos de las muestras por estar fuera de los límites del contexto (no pertenencia, no relevancia y redundancia innecesaria) y el recálculo de frecuencia de aparición de cada elemento en el conjunto muestral por detección de sinónimos y homónimos inter-muestras.

De los resultados estadísticos obtenidos [17], se estableció que realizar un análisis semántico hace más confiable la estimación del tamaño del modelo, dado que cada muestra presentó un mayor grado de completitud, aunque este permaneció bajo.

En la Tabla 1 se presentan los datos de las nueve muestras correspondientes antes del análisis semántico (PRE) y posterior a dicho análisis (POST), indicándose la cantidad de elementos elicitados (en crudo y con ajuste semántico) y el nivel de completitud alcanzado en base al tamaño estimado del modelo. La aplicación de DPM para estimar el tamaño del modelo y el detalle del análisis semántico se ha presentado en [4] y se detalla en [22]. Este método brinda un valor esperado, de modo tal que se pueda evaluar si un dato obtenido está suficientemente cerca de dicho valor y, por lo tanto, se lo puede considerar apropiado.

Se observa en la Tabla 1 que el nivel de completitud de cada grupo (excepto el grupo 6) mejoró con el ajuste semántico, donde el grupo con más alto nivel de completitud alcanzó el 58% sobre el tamaño estimado de 106 símbolos frente al nivel inicial del 50% alcanzado por dicho grupo.

Estos resultados llevaron a concluir que el problema de completitud es mucho más importante que lo que se percibe a primera vista. Es un aspecto frecuentemente subestimado. En base a esto, se decidió construir nuevas muestras del mismo modelo utilizando un proceso de construcción más refinado. La idea central fue establecer si disponer de guías más detalladas colabora en una elicitación y un

modelado independiente del grupo de ingenieros, permitiendo alcanzar un grado de completitud mayor.

Tabla 1. Cantidad de símbolos elicitados y ajustados semánticamente en un estudio previo [17]

GRUPOS INICIALES	PRE Análisis Semántico	PRE Nivel de completitud	POST Análisis Semántico	POST Nivel de completitud
Grupo 1	54	40%	46	43%
Grupo 2	29	22%	28	26%
Grupo 3	31	23%	24	23%
Grupo 4	35	26%	35	33%
Grupo 5	67	50%	62	58%
Grupo 6	27	20%	14	13%
Grupo 7	45	34%	44	42%
Grupo 8	24	18%	22	21%
Grupo 9	54	40%	52	49%
Total de símbolos elicitados	366		327	
Total de símbolos sin repetir	129		97	
Tamaño estimado	134		106	

Se produjeron, entonces, tres nuevas muestras del modelo LEL sobre el mismo contexto, cuyos datos ajustados semánticamente se presentan en la Tabla 2 (parte Nuevos Grupos). Estos grupos estaban compuestos por estudiantes de grado que realizaron un curso de Ingeniería de Requisitos y aplicaron una estrategia bien precisa para construir el LEL [23], incluyendo una inspección de la muestra [24]. Se planteó como *hipótesis de trabajo* que las nuevas muestras tendrían un mejor nivel de completitud dado que los grupos produjeron el modelo léxico con un proceso de construcción más refinado.

Para comprobar esta hipótesis se produjeron una serie de comparaciones con las muestras producidas por los nueve grupos anteriores, conformando conjuntos compuestos por tres grupos, como se indica a continuación.

#### 1) Los tres mejores grupos:

Se buscaron los mejores tres grupos de la Tabla 1, seleccionando aquellos grupos que hallaron la mayor cantidad de símbolos (post análisis semántico). Así, los grupos seleccionados fueron 1, 5 y 9. Cabe señalar que los integrantes de estos tres grupos eran expertos en la problemática de Ingeniería de Software y Requisitos, ya que eran grupos conformados por docentes universitarios especializados en Ingeniería de Software.

2) *Los tres peores grupos:*

Se buscaron los peores tres grupos de la Tabla 1, seleccionando aquellos con menor cantidad de símbolos elicitados (post análisis semántico), siendo ellos los grupos 3, 6 y 8.

3) *Los tres grupos al azar:*

Se seleccionaron tres grupos al azar a partir de los nueve grupos iniciales, ellos fueron los grupos 2, 7 y 9. Para dicha selección se colocaron todos los números de grupo - 1 al 9 - en una caja y se tomaron tres cualesquiera al azar.

En la Tabla 2 se muestra la cantidad de símbolos ajustados semánticamente y la cantidad de símbolos descartados por la corrección semántica para cada conjunto de tres grupos descrito arriba, incluyendo el conjunto de los tres nuevos grupos.

Tabla 2. Datos de las muestras de los conjuntos de grupos estudiados

<b>Nuevos grupos</b>	Cantidad elicitada ajustada	Símbolos descartados
Grupo 10	55	14
Grupo 11	53	14
Grupo 12	47	4
Total de símbolos	155	32
<b>Mejores grupos</b>	Cantidad elicitada ajustada	Símbolos descartados
Grupo 1	46	8
Grupo 5	62	5
Grupo 9	52	2
Total de símbolos	160	15
<b>Peores grupos</b>	Cantidad elicitada ajustada	Símbolos descartados
Grupo 3	24	7
Grupo 6	14	13
Grupo 8	22	2
Total de símbolos	60	22
<b>Grupos al Azar</b>	Cantidad elicitada ajustada	Símbolos descartados
Grupo 2	28	1
Grupo 7	44	1
Grupo 9	52	2
Total de símbolos	124	4

Estos cuatro conjuntos de grupos (los tres grupos nuevos, los tres mejores grupos, los tres peores grupos y los tres grupos al azar)

fueron estudiados aisladamente utilizando los datos ajustados semánticamente.

**Resultados**

Por cada conjunto, se calculó la cantidad de elementos coincidentes entre los tres grupos, la cantidad de elementos descartados a consecuencia del análisis semántico, el total de elementos del conjunto (sin repetición) y el tamaño estimado del modelo LEL según las tres muestras del conjunto aplicando DPM. Estos resultados se muestran en la Tabla 3.

Tabla 3. Resultados de los cuatro conjuntos de grupos estudiados

<b>Nuevos grupos (10, 11 y 12)</b>	
Cantidad de símbolos descartados	23
Cantidad de símbolos compartidos	26
Total de símbolos sin repetición	80
<b>Tamaño estimado</b>	<b>106</b>
<b>Mejores grupos (1, 5 y 9)</b>	
Cantidad de símbolos descartados	16
Cantidad de símbolos compartidos	25
Total de símbolos sin repetición	84
<b>Tamaño estimado</b>	<b>111</b>
<b>Peores grupos (3, 6 y 8)</b>	
Cantidad de símbolos descartados	21
Cantidad de símbolos compartidos	8
Total de símbolos sin repetición	33
<b>Tamaño estimado</b>	<b>43</b>
<b>Grupos al azar (2, 7 y 9)</b>	
Cantidad de símbolos descartados	8
Cantidad de símbolos compartidos	17
Total de símbolos sin repetición	74
<b>Tamaño estimado</b>	<b>92</b>

Como se puede observar en la Tabla 3, respecto a los conjuntos con las muestras iniciales, el conjunto de los mejores grupos (compuesto por expertos) logró elicitar la mayor cantidad de símbolos y la mayor cantidad de símbolos compartidos por los tres grupos del conjunto, y se le descartaron la menor cantidad de símbolos en el análisis semántico (excepto el conjunto de grupos al azar). Son notorios los bajos resultados alcanzados por el conjunto de los peores grupos, mientras que los grupos al azar tienen una tendencia más cercana a los mejores grupos que a los peores grupos.

Comparando el conjunto de mejores grupos contra el conjunto de los nuevos grupos, se puede observar en la Tabla 3 que tienen resultados bastante similares, lo mismo se observa gráficamente en la Figura 1. Esto podría indicar que el proceso refinado de construcción del modelo ha elevado a personas con cualidades medianas (respecto a conocimientos en Ingeniería de Requisitos) al nivel de los mejores.

La Figura 1 muestra la curva de aproximación de cada conjunto, donde se observa claramente la casi superposición de la curva que representa al conjunto de los mejores grupos con la de los nuevos grupos. Además, ambas curvas tienen el mayor coeficiente de determinación  $R^2$ .

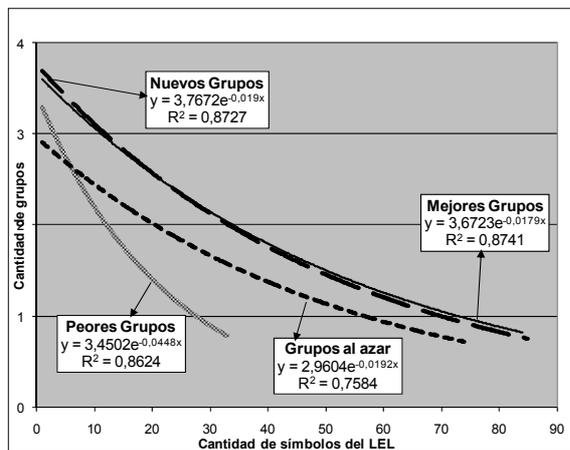


Figura 1. Aplicación de DPM post corrección semántica a los 4 conjuntos de grupos

Finalmente, se calculó el nivel de completitud de cada muestra del modelo léxico considerando ahora el conjunto de los doce grupos y el nuevo tamaño estimado aplicando DPM con el ajuste semántico (ver Tabla 4). Los nueve grupos iniciales han disminuido su nivel de completitud (comparar los resultados en la Tabla 1 vs la Tabla 4), esto está dado en parte por la cantidad de símbolos elicitados únicamente por los tres nuevos grupos, los cuales no fueron elicitados por ninguno de los nueve grupos iniciales.

Se puede entonces observar que el grupo que obtuvo el mayor nivel de completitud es el Grupo 5, aunque éste sólo alcanzó el 51%, siendo notoriamente aún más bajo el

nivel de completitud en la mayoría de los restantes grupos: 8 de 12 grupos están por debajo del 40% de completitud.

Tabla 4. Nivel de completitud en base al total de muestras estudiadas

GRUPOS	Cantidad elicitada ajustada	Nivel de completitud
Grupo 1	46	38%
Grupo 2	28	23%
Grupo 3	24	20%
Grupo 4	35	29%
Grupo 5	62	51%
Grupo 6	14	11%
Grupo 7	44	36%
Grupo 8	22	18%
Grupo 9	52	43%
Grupo 10	55	45%
Grupo 11	53	43%
Grupo 12	47	39%
Total de símbolos elicitados		482
Total de símbolos sin repetición		111
Tamaño estimado		122

Los nuevos grupos, que usaron heurísticas de construcción del LEL más precisas, sólo obtuvieron entre 39% y 45% de completitud. Sólo fueron superados o igualados por dos grupos de expertos (Grupo 5 y Grupo 9). Dados los resultados obtenidos (ver Tabla 4), no se puede señalar un nivel de completitud aceptable en ninguno de los casos. Se observa que el hecho de que expertos hayan encontrado sólo un 51% indica un exceso de omisiones.

## Discusión

Se puede concluir que los nuevos grupos, con un proceso refinado de construcción del modelo léxico (más guías de construcción y más precisas), casi alcanzaron el nivel de los expertos en cuanto a las cantidades de símbolos elicitados y símbolos comunes, aunque tuvieron más símbolos descartados en el análisis semántico. Es decir, se puede decir que el proceso refinado ayudó a los nuevos grupos conformados por novatos a suplir la experiencia de los mejores grupos. Tanto el conjunto de los mejores grupos como el conjunto de los nuevos grupos, tienen muy poca cantidad de símbolos comunes entre los grupos del conjunto (ver Tabla 3), lo cual muestra que cada grupo es

muy incompleto respecto del conjunto al que pertenece.

Analizando los mejores grupos frente a los nuevos grupos (dado que son los seis grupos que obtuvieron mayor nivel de completitud), se observa que apenas coinciden en 14 símbolos de 150 símbolos elicitados (ver Figura 2). Esto significa que el conjunto de grupos de expertos no encontró los 66 símbolos que sí detectó el otro conjunto, mientras que el conjunto de grupos novatos no descubrió los 70 símbolos encontrados por los expertos. Esto pone en evidencia, una vez más, el alto nivel de incompletitud de todos los grupos. Ni los expertos ni los que utilizaron mejores heurísticas llegaron a un nivel aceptable de completitud.

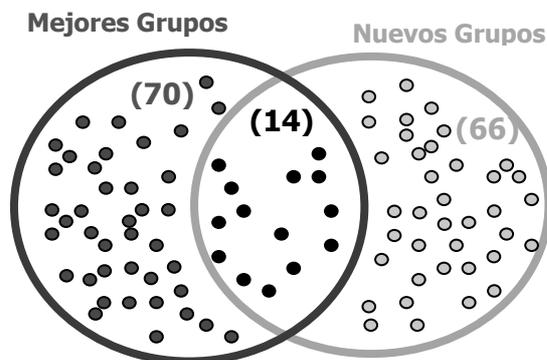


Figura 2. Estudio de muestras obtenidas por expertos y por novatos

Esta baja coincidencia de símbolos elicitados por los distintos grupos se agrava aún más si se considera el conjunto de los doce grupos (solo 8 símbolos compartidos por todos los grupos), lo que refuerza la idea de un exceso de omisiones entre muestras.

Dados los resultados que muestran la disparidad de niveles de completitud y, asimismo, el estudio previo que determinó grupos observando el contexto con límites diferentes [4], se podría suponer que cada grupo posee un punto de vista diferente en algunas subáreas del problema, lo que indicaría los bajos niveles de completitud alcanzados. Aunque estos límites diferentes del problema deberían corregirse mediante el ajuste semántico.

Si dos personas frente al mismo problema, utilizando el mismo modelo e iguales técnicas de elicitación construyen modelos poco comparables, eso es evidencia de que la libertad del modelo es excesiva o, en otras palabras, que la heurística de construcción no es suficientemente precisa. Ninguna de estas percepciones obtenidas a partir de todos los trabajos anteriores fue siquiera sugerida en ninguno de los estudios, difusiones, discusiones, o publicaciones sobre el modelo utilizado [5, 25, 22, 4, 17]. Sólo los estudios de completitud, aún posiblemente deficientes en su sustento básico, hicieron evidente lo aquí expuesto.

### Conclusión

Estrictamente el problema es doble, por un lado, las muestras estudiadas del modelo léxico son notoriamente incompletas, pero por el otro, la forma de comparar estas muestras necesita ser mejorada, por lo que los trabajos futuros deben utilizar una estrategia convergente que permita a la vez mejorar la completitud pero también mejorar las técnicas de comparación de dos modelos creados por distintas personas sobre el mismo objeto de estudio.

Si se considera que en la práctica habrá sólo un grupo de elicitadores, será mucha la incompletitud que seguramente presentará el resultado de la construcción de este modelo léxico. No se puede considerar entonces a este modelo como uno que brinde un resultado con un nivel de completitud aceptable y que, por lo tanto, responda correctamente a las necesidades de los clientes y usuarios.

Cuando se plantea un modelo es para ayudar a pensar una situación determinada, y para volcar en él la experiencia de los expertos en el tema y para facilitar la aplicación futura de los novatos en el área. Si se presenta un modelo que se considera correcto y se presupone que da buen nivel de completitud, pero los resultados experimentales muestran que no se consigue la completitud esperada, entonces se puede concluir que el modelo no es tan

bueno. Es por ello que queda mucho trabajo por desarrollar para mejorar la heurística existente de construcción del modelo léxico. Este es un defecto además extensible a todos los glosarios, no sólo aplicable al modelo LEL.

Asimismo, no es esperable que, dado dos grupos independientes de elicidores, cada uno construya el mismo modelo con tan pocas coincidencias entre ambos. Esto lleva a suponer que las comparaciones semánticas entre modelos deben profundizarse para poder alcanzar una estimación del tamaño del modelo LEL más confiable.

### Referencias

- [1] Kotonya, G., Sommerville, I.: *Requirements Engineering: Process and Techniques*. John Wiley & Sons (1998)
- [2] Loucopoulos, P., Karakostas, V.: *System Requirements Engineering*. McGraw-Hill, Londres (1995)
- [3] Firesmith, D.: Are Your Requirements Complete? *Journal of Object Technology*, Vol. 4, N° 1, pp. 27–43 (2005)
- [4] Litvak C.S., Hadad, G.D.S., Doorn, J.H.: Un abordaje al problema de completitud en requisitos de software. CACIC 2012 - XVIII Congreso Argentino de Ciencias de la Computación, pp. 827–836, Bahía Blanca (2012)
- [5] Doorn, J.H., Ridao, M.: Completitud de Glosarios: Un Estudio Experimental: WER'03 - VI Workshop on Requirements Engineering, pp. 317–328, Brasil (2003)
- [6] Leite, J.C.S.P.: 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 (2001)
- [7] Alrajeh D., Kramer J., Van Lamsweerde A., Russo A., Uchitel S.: Generating Obstacle Conditions for Requirements Completeness. ICSE 2012 - 34th International Conference on Software Engineering, Suiza, pp. 705–715 (2012)
- [8] Leite, J.C.S.P., Freeman, P.A.: Requirements Validation Through Viewpoint Resolution: *IEEE Transactions on Software Engineering*, Vol. 17, N° 12, pp. 1253–1269 (1991)
- [9] Zowghi D., Gervasi V.: The Three Cs of Requirements: Consistency, Completeness, and Correctness. REFSQ'02 - 8th International Workshop on Requirements Engineering: Foundation for Software Quality, pp. 155–164, Essen (2002)
- [10] Carson, R.S.: Requirements Completeness: A Deterministic Approach. 8th Annual International Symposium of the INCOSE, Canadá, pp. 739–746 (1998)
- [11] Goguen, J.A., Linde, Ch.: Techniques for Requirements Elicitation. RE'93 - IEEE First Intl Symposium on Requirements Engineering, IEEE Computer Society Press, EEUU, pp.152–164 (1993)
- [12] Leite J.C.S.P., Hadad G.D.S., Doorn J.H., Kaplan G.N.: A Scenario Construction Process, *Requirements Engineering Journal*, 5(1):38-61 (2000)
- [13] Leite JCSP, Doorn JH, Hadad GDS, Kaplan GN: Scenario Inspections. *Requirements Engineering Journal*, Springer-Verlag London, Gran Bretaña, 10(1):1-21 (2005)
- [14] Uchitel S, Chatley R, Kramer J, Magee J: Goal and scenario validation: a fluent combination. *Requirements Engineering Journal*, 11(2):123–137 (2006)
- [15] Pitts MG, Browne GJ: Stopping behavior of systems analysts during information requirements elicitation. *Journal of Management Information Systems*, 21(1):203- 226 (2004)
- [16] Browne GJ, Pitts MG, Wetherbe JC: Cognitive stopping rules for terminating information search in online tasks. *MIS Quarterly*, 31(1):89-104 (2007)
- [17] Litvak C.S., Hadad, G.D.S., Doorn, J.H. Correcciones semánticas en métodos de estimación de completitud de modelos en lenguaje natural. WER 2013 - XVI Workshop on Requirements Engineering, Montevideo, Uruguay, pp. 105–117 (2013)
- [18] Otis DL, Burnham KP, White GC, Anderson DR: Statistical inference from Capture on Closed Animal Populations. *Wildlife Monograph*, 62 (1978)
- [19] Wohlin C, Runeson P: Defect content estimations from Review Data. 20th International Conference on Software Engineering, pp. 400–409, Japón (1998)
- [20] Leite, J.C.S.P., Franco, A.P.M.: A Strategy for Conceptual Model Acquisition. IEEE 1st Intl Symposium on Requirements Engineering, IEEE Computer Society Press, EEUU, pp 243–246 (1993)
- [21] Mauco V, Ridao M, del Fresno M, Rivero L, Doorn JH: Ingeniería de Requisitos, Proyecto: Sistema de Planes de Ahorro. Reporte técnico, ISISTAN, UNICEN, Tandil (1997)
- [22] Hadad, G.D.S., Litvak, C.S., Doorn, J.H.: Estudio semántico de modelos construidos por elicidores independientes observando el mismo problema. Serie Documentos de Trabajo, N° 279, Departamento de Investigaciones, Universidad de Belgrano, Buenos Aires, ISSN: 1850-2512, 25 páginas (2012)

- [23] Hadad, G.D.S.: Uso de Escenarios en la Derivación de Software. Tesis Doctoral, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina (2008)
- [24] Kaplan GN, Hadad GDS, Doorn JH, Leite JCSP: Inspección del Léxico Extendido del Lenguaje. WER'00 – 3rd Workshop on Requirements Engineering, Río de Janeiro, Brasil, pp. 70-91 (2000)
- [25] Ridao M, Doorn JH: Estimación de Completitud en Modelos de Requisitos Basados en Lenguaje

Natural. WER'06 - IX Workshop on Requirements Engineering, Brasil, pp. 151–158 (2006)

**Datos de Contacto:**

*Graciela Dora Susana Hadad. Universidad de Belgrano y Universidad Nacional de La Matanza. Dirección postal: Sede FITI-UB: Federico Lacroze 1947 (C1426CPE), Ciudad de Buenos Aires. E-mail: [graciela.hadad@comunidad.ub.edu.ar](mailto:graciela.hadad@comunidad.ub.edu.ar)*