

Correcciones semánticas en métodos de estimación de completitud de modelos en lenguaje natural

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

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

²DIIT, Universidad Nacional de La Matanza, Argentina

³Fac. Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires

{claudia.litvak,graciela.hadad}@comunidad.ub.edu.ar,
jdoorn@exa.unicen.edu.ar

Abstract. La Ingeniería de Requisitos tiene como objetivo producir requisitos de alta calidad, poniendo especial interés en la completitud de los modelos elaborados. Establecer si la información elicitada y modelada es suficiente para construir un software que cubra las necesidades del cliente es una cuestión de difícil respuesta. Existen algunos intentos realizados en ese sentido, tal como una adaptación del método de captura-recaptura para estimar el tamaño de modelos de requisitos escritos en lenguaje natural. Basados en dicho trabajo y considerando la naturaleza de estos modelos, proponemos introducir un análisis semántico previo a estimar el tamaño, dado que el método predictivo solo contempla aspectos formales y cuantitativos. Dicho análisis semántico estudia la relevancia, pertenencia, sinonimia y homonimia del contenido textual del modelo. Comparando los resultados estadísticos de trabajos precedentes contra los obtenidos realizando correcciones semánticas, concluimos que estas son beneficiosas para la estimación de la completitud de modelos en lenguaje natural.

Keywords. modelado de requisitos, completitud de modelos, método de captura-recaptura, modelos en lenguaje natural.

1 Introducción

La completitud de los requisitos no solo es uno de los criterios de calidad pregonados en la literatura, sino en el ámbito práctico del desarrollo de software. El estándar IEEE 830-1998 reafirmado en 2009 [1] la considera una de las 8 propiedades que debe cumplir una buena especificación de requisitos de software.

Con lo cual, la incompletitud es uno de los principales problemas que afronta la Ingeniería de Requisitos [2-4]. La dificultad radica en poder determinar si se ha logrado elicitar y modelar toda la información necesaria para desarrollar un sistema de software que cubra las expectativas y necesidades de los clientes y usuarios. Según Leite [5], el proceso de definición de requisitos es intrínsecamente incompleto debido a la complejidad del mundo real. Esto significa que no hay ninguna posibilidad de

alcanzar la completitud de un modelo, con lo cual el principio conductor debe ser elaborar un modelo lo más completo posible.

La incompletitud es un problema común a otras actividades del proceso de desarrollo de software. En el diseño, es lógico cuestionarse si se han identificado todas las clases, todos los métodos y todos los atributos. En las inspecciones de código o de algún otro artefacto permanece siempre la duda si se han detectado todos los defectos. La misma incógnita se plantea en la etapa de testing. Es decir, pueden existir cuestiones no descubiertas en diversos artefactos o actividades a lo largo de todo el desarrollo de software y persistir estas sin ser detectadas, con la consecuente degradación en la calidad del software e inevitable insatisfacción del cliente. Aún cuando la completitud es un criterio de calidad pregonado en la literatura y exigido por muchos estándares, su cumplimiento es con mucha frecuencia parcialmente alcanzado.

Entonces, el problema de la incompletitud está directamente relacionado con un tipo específico de defecto: las omisiones. Las técnicas de verificación, en algún grado, y de validación, menos aún, ayudan a mitigar este problema, así como el uso de técnicas de elicitación apropiadas permiten lograr una adquisición de conocimiento más completa. Sin embargo, estas técnicas por sí solas no permiten eliminar el problema ni siquiera estimar el grado de completitud alcanzado. Esto significa que, si no se puede asegurar la completitud, al menos se debiera ser consciente del nivel aproximado de completitud alcanzado en los modelos para decidir cuándo avanzar hacia otras etapas del proceso de desarrollo o continuar elaborándolos.

El método de captura-recaptura [6] es un método de predicción del tamaño de poblaciones cerradas que se ha utilizado en Ingeniería de Software para estimar el número de defectos aún no descubiertos en artefactos de código [7-9] y en documentos de requisitos [10]. Posteriormente, este método se aplicó en la Ingeniería de Requisitos para predecir el número de elementos faltantes en un modelo de requisitos [11-12]. En el trabajo de Walia y Carver [10] se realizan varias inspecciones independientes sobre el mismo modelo de requisitos para estimar defectos en dicho modelo, mientras que Doorn y Ridao [11-12] realizaron la estimación de elementos no elicitados en un dominio de aplicación a partir de varios ejemplares de un modelo, construidos independientemente.

La adaptación del método de captura-recaptura realizado por Wohlin y Runeson [7], llamado Detection Profile Method (DPM), fue aplicado por Doorn y Ridao [11-12] para estimar el tamaño de dos modelos escritos en lenguaje natural: el Léxico Extendido del Lenguaje y Escenarios. El método predictivo requiere tener dos o más muestras de la misma población, por lo cual Doorn y Ridao utilizaron nueve muestras de ambos modelos. Cada muestra fue producida por diferentes grupos de elicitores. De acuerdo con los resultados que obtuvieron, el número de elementos omitidos en cada muestra era de significativa importancia. El grupo de elicitores que descubrió el mayor número de elementos sólo alcanzó un nivel de completitud del 51%, medido en referencia a la cantidad de elementos presentes en la unión de todos los valores muestrales obtenidos por todos los grupos. A primera vista, estos niveles indicarían que el modelo construido por cada grupo de elicitores es bastante pobre. Esta sola experiencia muestra claramente que el problema de la incompletitud es efectivamente mucho más serio que lo que en principio se presupone.

En base a esto, hemos conjeturado acerca de la posibilidad de que no se hayan identificado los elementos adecuados en cada una de las muestras al aplicar el método de captura-recaptura. Por lo que, hipotetizamos que se debería realizar un análisis previo de los elementos de cada muestra antes de establecer la estimación del tamaño del modelo, con el fin de mejorar la calidad de dicha estimación. Los resultados iniciales que obtuvimos confirman una mejora en la estimación del tamaño para un modelo específico escrito en lenguaje natural.

En la siguiente sección se presentan los trabajos anteriores y la motivación del presente trabajo. En la sección 3 se describe la hipótesis planteada y el trabajo desarrollado para comprobarla, detallando el análisis semántico realizado. En la sección 4 se detallan observaciones del estudio semántico-estadístico realizado y, finalmente, se exponen conclusiones y trabajos futuros.

2 Marco de Trabajo

Doorn y Ridao en [11] han estudiado experimentalmente el uso de DPM [7] para estimar el número de términos faltantes en un glosario, llamado Léxico Extendido del Lenguaje (LEL) [13]. El LEL es un glosario de términos (símbolos) que se utilizan en el universo de discurso. Cada término se describe por uno o más nombres (sinónimos), una noción (denotación), un impacto (connotación) y un tipo (clasificación del símbolo) que guía la definición del término.

En dicho trabajo [11], concluyeron que aumentando el número de elicitadores (ingenieros en requisitos), la brecha entre el número de elementos elicitados y el número de elementos estimados disminuye. Cuando se refieren al número de elicitadores, esto implica que se dispone de más muestras de la población, es decir, más modelos-muestra.

Se analizaron los resultados que obtuvieron Doorn y Ridao aplicando DPM al modelo LEL. El número estimado de elementos faltantes en el LEL fue 9 de los 118 términos elicitados en conjunto por todos los grupos. En principio, el número de elementos faltantes parece ser un número bajo, sin embargo, esto no es así cuando se estudia la muestra de cada grupo de elicitadores independientemente. Si observamos los resultados obtenidos por dos grupos cualesquiera, por ejemplo el Grupo 1 y el Grupo 5, ellos coinciden en 34 elementos, pero difieren en 50. Esta diferencia se manifiesta en que al Grupo 1 le faltó encontrar 32 elementos que sí fueron detectados por el Grupo 5, y al Grupo 5 le faltaron 18 elementos del Grupo 1. Luego, se observa que las diferencias superan ampliamente las coincidencias. Cabe destacar que estas diferencias señalan omisiones en las respectivas muestras del modelo. Esta situación se agrava si comparamos un grupo respecto a todos los otros, por ejemplo, al Grupo 1 le faltó encontrar en total 66 elementos, mientras que el Grupo 5 omitió 53 elementos. Aquí entonces, la cantidad de omisiones es más notoria aún. El hecho de que dos grupos difieran en alrededor de 50 términos requiere estudios más profundos para explicar estas diferencias.

Se debe tener en claro que la experiencia reportada en [11] y revisada en este artículo no es un trabajo profesional viable ya que se multiplica el esfuerzo necesario por

un factor de 9. Este estudio sólo es posible en el contexto de un proyecto de investigación que pretende ponderar de manera cuantitativa la magnitud de la falta de completitud. Por otro lado, ninguna de estas 9 muestras del modelo sufrió un proceso de validación. Se debe notar que en general una validación con clientes permite detectar básicamente datos erróneos y contradicciones, mientras que la identificación de omisiones tiene una baja incidencia durante una validación debido a que los clientes habitualmente se centran en analizar lo que está presente en el modelo. En otras palabras, en el mejor de los casos, las personas provenientes de la organización receptora del sistema tienen por objetivo colaborar hasta donde les sea posible con la redacción de un documento, por ejemplo un glosario, cuyo propósito no les es propio. Está fuera de su intención y habitualmente de su capacidad darse cuenta de qué palabras o frases, ellas usan frecuentemente y que son importantes para la Ingeniería de Requisitos. Por estos motivos, ellas se darán cuenta fácilmente de errores en la definición de un término y, en el mejor de los casos, de debilidades, pero seguramente estarán lejos de poder contribuir espontáneamente con nuevos términos.

Basándonos en el estudio de Doorn y Ridao del modelo LEL y considerando todas las omisiones observadas anteriormente, realizamos un nuevo estudio bajo la hipótesis de que los grupos de elicidores estaban observando diferentes universos de discurso [14-15]. En dicho trabajo previo, dividimos el universo bajo estudio en cinco sub-áreas para facilitar la realización del análisis estadístico. Confirmamos estadísticamente, que en cuatro sub-áreas los grupos habían observado universos con límites diferentes, mientras que en sólo una sub-área, todos estaban observando el mismo universo. La Fig. 1 presenta las tres visiones diferentes de la sub-área Administración de Bien Tipo. Esto significa que 4 grupos observaron el mismo universo en esta sub-área, 3 grupos observaron otro universo y otros 2 grupos tuvieron otra visión del universo. Es de destacar que estos tres universos fueron detectados y confirmados utilizando técnicas estadísticas, habiéndose logrado un grado de confianza muy alto en el resultado.

Estas conclusiones nos llevaron a formular una nueva hipótesis bajo la cual debemos estudiar el contenido textual de las muestras internamente (intra-análisis semántico) y entre muestras (inter-análisis semántico) para unificar los diferentes límites observados del mismo universo y poder estimar adecuadamente el tamaño del modelo.

Es de destacar que tanto los estudios originales [11-12] como esta revisión, se basan en datos no influenciados por el trabajo de investigación realizado, ya que en el momento de construcción de los 9 modelos léxicos no existía ninguna intención de realizar un estudio comparativo de los mismos, siendo que este estudio fue realizado varios años después, dado lo cual no hubo contaminación por parte de los investigadores. Por otro lado, debemos mencionar que estos modelos se basaron en un caso real detallado en forma escrita, evitando la contaminación cruzada entre grupos de elicidores. Debe notarse, además, que el número 9 de muestras es un valor importante para aplicar métodos formales de estimación.

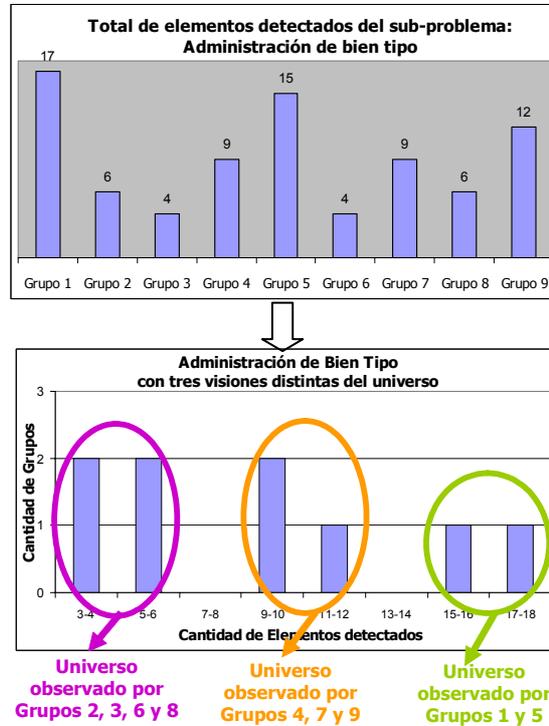


Fig. 1. Visiones de los elicidores en una sub-área: Administración de Bien Tipo [14]

3 Desarrollo del trabajo

3.1 Hipótesis de Trabajo

En base al análisis de los resultados presentados en un artículo preliminar [15], planteamos la siguiente hipótesis de trabajo: “Se requiere un análisis semántico sobre los elementos del modelo en lenguaje natural antes de aplicar métodos formales de estimación de tamaño del modelo”.

Para comprobarlo, partimos de las mismas muestras utilizadas por Doorn y Ridao en [11], es decir, dispusimos de 9 muestras del modelo LEL generadas por 9 grupos de ingenieros (alumnos de posgrado y profesores) que estudiaron el mismo universo: el sistema de Planes de Ahorro Previo para la Adquisición de Vehículos Okm.

Comenzamos realizando un estudio estadístico para estimar el tamaño del modelo LEL sin ningún ajuste en los elementos de las muestras (símbolos del léxico generado por cada grupo de elicidores). Luego, analizamos semánticamente cada elemento de la muestra, para finalmente volver a realizar la estimación del tamaño del modelo con las correcciones semánticas necesarias. En las siguientes sub-secciones detallamos estas actividades.

3.2 Trabajo Estadístico Inicial

Para realizar el estudio estadístico aplicando DPM sobre las 9 muestras, se requirió primero identificar los elementos comunes (símbolos) elicitados por todos los grupos, luego identificar aquellos elicitados por solo 8 grupos, y así sucesivamente hasta obtener los elementos elicitados por un solo grupo. Es decir, se determinó la unión de las 9 muestras, registrando la frecuencia de aparición de cada elemento (en cuantos grupos apareció el mismo elemento, ver Tabla 1).

La identificación de elementos comunes entre distintas muestras se realizó por simple sinonimia de nombre. Esto significa que dos símbolos de distinto grupo representan un único elemento del universo si ambos tienen el mismo nombre, o coinciden en al menos un nombre (caso que el símbolo tenga sinónimos). Cabe aclarar que la comparación de nombre consideró variantes gramaticales, tales como flexiones nominales, flexiones verbales y formas sustantivas de verbos. Este fue un criterio similar al aplicado en [11] con menos refinamiento, para identificar el conjunto de símbolos distintos detectados por todos los grupos.

Tabla 1. Frecuencia de aparición de cada símbolo en el conjunto de muestras

SÍMBOLOS	G1	G2	G3	G4	G5	G6	G7	G8	G9	Frecuencia
Solicitud de adhesión / Solicitud / Pl	1	1	1	1	1	1	1	1	1	9
Solicitante	1	1	1	1	1	1	1	1	1	9
Adherente / Miembro de grupo / Titul	1	1	1	1	1	1	1	1	1	9
Administradora / Sociedad Administra	1	1	1	1	1	1	1	1	1	9
Sorteo / Sorteo mensual / Sortear / Ad	1	1	1	1	1	1	1	1	1	9
Bien tipo / Automóvil / Bien / Vehículo	1	1	1	1	1	1	1	1	1	9
Fabricante / Importador / Proveedor	1	1	1	1	1	1	1	1	1	9
Plan de ahorro / Plan / Sistema de Ah	1	1	1	1	1	1	1	1	1	9
Grupo / Grupo de Adherentes	1	1	1	1	1	1	1	1	1	9
Comunicación fehaciente / Notificació	1	1		1	1	1	1	1	1	8
Adjudicatario / Adjudicado	1	1	1	1	1		1	1	1	8
Licitación / Adjudicar por licitación / A	1	1		1	1		1	1	1	7
Adjudicación / Acto de adjudicación /	1	1	1		1	1	1		1	7
Cuota pura	1	1	1	1	1			1	1	7
.....	
Cesionario / Adherente Cesionario					1				1	2
Incumplimiento imputable al grupo					1				1	2
Notificación de rechazo	1									1
Aceptación / Aceptado			1							1
Desistir de Solicitud					1					1
Rechazo de solicitud					1					1
Solicitante aceptado							1			1
Cantidad de símbolos: 132	54	29	31	35	67	27	46	24	55	368

El método DPM [7] aplicado se basa en ajustar la curva que representa los datos referidos a la cantidad de grupos que elicitaron cada símbolo (obtenidos de la planilla confeccionada cuyo extracto se muestra en la Tabla 1), utilizando una curva teórica simple. En este caso se usó una función exponencial decreciente, ecuación (1), cuyos parámetros incluyen las variaciones de probabilidad de detección de un símbolo y de la habilidad de elicitación del grupo.

Entonces, la ecuación que representa la curva de ajuste exponencial es:

$$m_x = A \times e^{-bx} \quad (1)$$

siendo

m_x : cantidad total de grupos que detectaron el símbolo x

b : decrecimiento de la función exponencial

A : constante

Utilizando técnicas de regresión se estimaron los parámetros A y b, obteniendo $A = 7,0514$ y $b = 0,0191$, con un coeficiente de determinación $R^2 = 0,8529$. Este coeficiente de determinación es un parámetro para evaluar la distancia entre la estimación y los datos experimentales ($0 \leq R^2 \leq 1$). La cantidad total de símbolos estimados se calcula por el mayor valor de x para el cual la ecuación (1) produce un valor mayor o igual que 0,5, es decir, haciendo $m_x = y$ para $y = 0,5$. Esto da como resultado el valor estimado 138. La Fig. 2 representa esta curva de aproximación.

Entonces, entre todos los grupos encontraron 132 símbolos (sin repetición) de los 138 símbolos estimados aplicando DPM, es decir, les ha faltado, en conjunto, elicitar 6 símbolos.

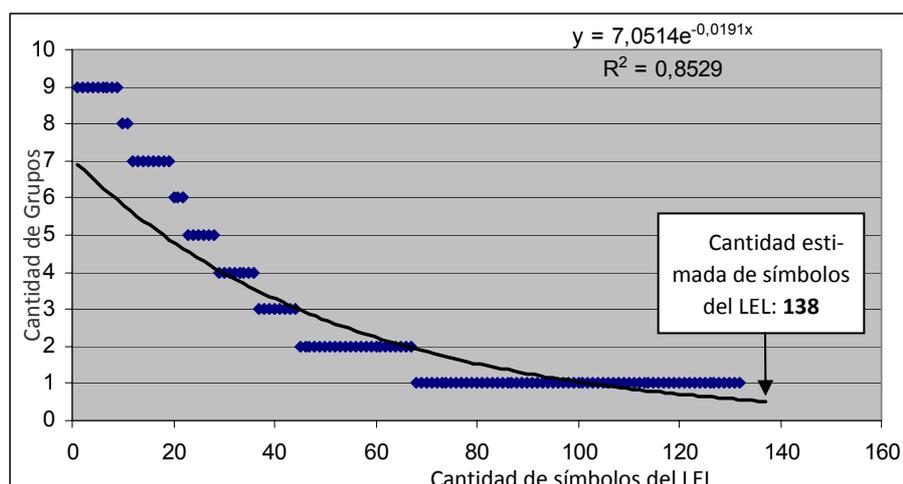


Fig. 2. Aplicación de DPM antes de la corrección semántica

3.3 Análisis Semántico

Se realizó un estudio de la semántica de todos los términos del LEL definidos por los 9 grupos de elicidores. Ahora, además de considerar los nombres de los símbolos, se analizó la denotación y connotación de los mismos.

Primero, se hizo un análisis semántico dentro de cada muestra, independiente de las restantes, donde se estableció la pertenencia y la relevancia del símbolo en el universo de discurso. Posteriormente, se realizó un análisis semántico entre símbolos de distintas muestras para establecer sinónimos y homónimos. Tanto en el intra-análisis semántico como en el inter-análisis, el estudio comenzó por aquellos símbolos de

menor frecuencia de aparición en el conjunto de todas las muestras. Una baja frecuencia implica que un único grupo o unos pocos grupos elicitaron el símbolo.

Para establecer la pertenencia, primero se validó la existencia del nombre del símbolo en dicho universo. Luego, se analizó su contenido (denotación y connotación) para establecer si el símbolo estaba fuera de los límites del universo bajo estudio o no. En el caso de que el contenido del símbolo tuviese información del universo pero el nombre fuese inválido (creado artificialmente por los elicitadores), podía existir otro nombre en el universo no detectado por el grupo elicitor. En este punto, el símbolo quedó marcado para ser tenido en cuenta en el inter-análisis semántico para establecer sinónimos con símbolos de otros grupos.

La relevancia se analizó en base a la denotación y connotación del símbolo. La relevancia de un símbolo se determinó por el aporte de conocimiento al modelo LEL. Esto es, se consideraron no relevantes aquellos símbolos con exceso en el nivel de detalle, innecesario para un glosario, o con información parcial ya contenida en otro símbolo.

El inter-análisis semántico comenzó por confirmar los sinónimos ya establecidos para el estudio estadístico inicial (presentado en la anterior sub-sección). Ahora, ya no alcanzó tener nombres de símbolos idénticos o nombres con variaciones gramaticales, se debían tener contenidos similares. Por lo tanto, se confirmaron sinónimos estudiando la denotación y connotación entre cada par de símbolos. En el caso de diferencias, se descartó la relación semántica de sinonimia, y se analizó la posibilidad de homónimos (homógrafos). Luego, este análisis semántico continuó estableciendo nuevos sinónimos entre símbolos de distintas muestras cuyos nombres no eran idénticos (ni las flexiones gramaticales). Para ello, se compararon las denotaciones y connotaciones de cada símbolo.

En resumen, como consecuencia del intra-análisis semántico, hubo símbolos que se descartaron, fueron aquellos que cumplían con algunas de las siguientes características: a) con información fuera del límite del universo de discurso; b) con información con un exceso de nivel de detalle, innecesaria para un glosario, que podía ser mencionada dentro de otro símbolo; y c) con información que ya estaba contenida en otros símbolos.

Como consecuencia del inter-análisis semántico, se ajustó la cantidad de símbolos sin repetición del conjunto de muestras, en base a los siguientes aspectos: a) detección de sinónimos entre muestras, y b) detección de homónimos entre muestras.

La Fig. 3 muestra los 9 grupos con los símbolos del LEL detectados por cada uno para la sub-área Adhesión. En esta figura se identifican los símbolos descartados: 8 símbolos de un total de 21 símbolos sin repetición, siendo el conjunto muestral de 60 símbolos. Las intersecciones representan los sinónimos. Se puede observar visualmente que hubo 4 símbolos compartidos por todos los grupos para esta sub-área, y que a su vez coincidieron con el conjunto muestral de los Grupos 2 y 8.

Los resultados del ajuste semántico se resumen en la Tabla 2. Antes del estudio semántico se habían detectado un total de 132 símbolos sin repetición, elicitados en conjunto por los 9 grupos, mientras que concluido el estudio semántico se descartaron 35 símbolos (esta diferencia está dada por no existentes, no relevantes, no identifica-

dos como sinónimos y desglose de homónimos inicialmente considerados como sinónimos), quedando entonces un conjunto de 97 símbolos.

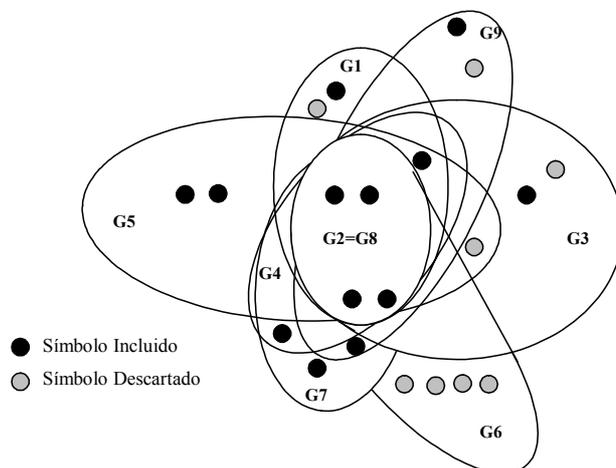


Fig. 3. Distribución visual de símbolos por grupos para sub-área Adhesión

Tabla 2. Resultados del análisis semántico

	PRE Análisis Semántico	POST Análisis Semántico
Grupo 1	54	46
Grupo 2	29	28
Grupo 3	31	24
Grupo 4	35	35
Grupo 5	67	62
Grupo 6	27	14
Grupo 7	46	44
Grupo 8	24	22
Grupo 9	55	51
Total de símbolos elicitados	368	326
Total de símbolos sin repetición	132	97
Símbolos Descartados		
Nombre no existe y no es relevante		9
Nombre existe pero símbolo no es relevante		23
Sinónimo no detectado en el grupo		3
Homónimo no detectado		(3)
Total de Símbolos Sin Repetición Descartados		35

3.4 Estimación Estadística con Ajuste Semántico

Una vez realizado el ajuste semántico se tuvo una nueva frecuencia de aparición de cada símbolo en el conjunto de las muestras. Se aplicó nuevamente DPM a las muestras corregidas y se obtuvo la curva de ajuste presentada en la Fig. 4, donde se estima-

ron los nuevos valores de A y b, obteniéndose $A = 9,561$ y $b = 0,0279$, con un coeficiente de determinación $R^2 = 0,9394$. Ahora, la cantidad de símbolos estimados fue de 105, es decir, han faltado encontrar 8 símbolos del conjunto corregido semánticamente.

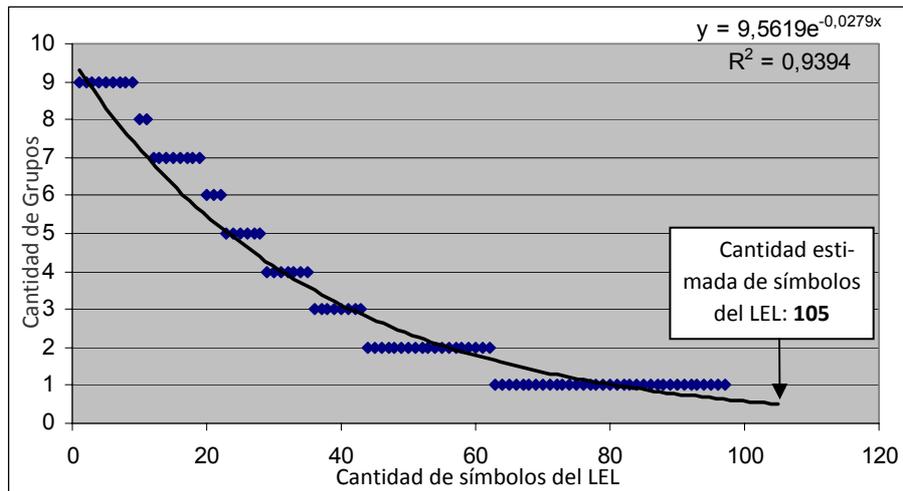


Fig. 4. Aplicación de DPM después de la corrección semántica

Los coeficientes de determinación permiten comparar ambos gráficos, Fig. 2 y 4. En el primer caso, sin las consideraciones semánticas (Fig. 2), el coeficiente de determinación obtenido fue $R^2 = 0,8529$. Al aplicar la corrección semántica (Fig. 4), el valor fue $R^2 = 0,9394$. Se observa entonces que este último valor está mucho más cerca de 1, lo que implica que la curva tiene estadísticamente una mejor calidad que la inicial.

Tabla 3. Comparación de algunos Resultados

		PRE Análisis Semántico	POST Análisis Semántico
Total de Grupos	Cantidad Elicitada	132	97
	Tamaño Estimado	138	105
	Omisiones Estimadas	6	8
Grupo 5	Cantidad Elicitada	67	62
	Nivel de Completitud	49%	59%

Si analizamos los valores obtenidos en la segunda estimación respecto de la primera, aún cuando ha crecido levemente la cantidad de omisiones, el nivel de completitud alcanzado individualmente por cada grupo ha mejorado. Como se muestra en la Tabla 3, el Grupo 5 que había elicitado la mayor cantidad de símbolos solo alcanzaba un

49% de completitud, y una vez realizadas las correcciones semánticas ha mejorado al 59%.

4 Observaciones sobre el estudio semántico-estadístico

En base al análisis semántico y posterior estudio estadístico, se puede puntualizar lo siguiente:

- Desde una perspectiva semántica, se puede afirmar que el Grupo 6 observó débilmente al universo, pues del total de 27 símbolos que elicitó, solo 14 símbolos fueron considerados relevantes. Además, identificó 7 símbolos cuyos nombres no existían en el universo de discurso. Es decir, en general tuvo una visión diferente del universo dada una captura de símbolos muy distorsionada frente al resto de los grupos.
- En el otro extremo, el Grupo 5 fue el que identificó más símbolos relevantes. Del total de 67 símbolos identificados por este grupo, solo 5 fueron descartados. Considerando el ajuste semántico, este grupo encontró el 64% de símbolos del total elicitados entre todos los grupos (97), aunque el 59% del total estimado (105).
- Considerando el conjunto de símbolos ajustados semánticamente (97), hubo 9 símbolos elicitados por todos los grupos y 10 símbolos detectados en común por 7 u 8 grupos diferentes. Es decir, hubo casi un 20% de símbolos comunes elicitados entre 7 y 9 grupos, valor que puede considerarse relativamente bajo.
- Se comprobó semánticamente la existencia de 16 símbolos cuyos nombres no pertenecían al universo de discurso, aunque el significado (denotación y connotación) de 7 de ellos era relevante, es decir, dicha información no debería omitirse. Efectivamente, esa información estaba contenida en otros símbolos de otros grupos, por lo tanto ellos no fueron desestimados. Además, debe notarse que había más nombres inexistentes pero tenían sinónimos dentro del grupo, por lo tanto, el nombre inexistente se descartó, sin alterar las estadísticas.
- En resumen, se identificaron en total 35 símbolos no necesarios de un total de 132 símbolos distintos identificados entre todos los grupos. Es decir, se descartó un 26% de los símbolos hallados por los grupos para realizar el estudio estadístico.

Todo esto confirma lo expuesto en la sección 2, en el sentido de que el problema de completitud es mucho más importante que lo que se percibe a primera vista.

5 Conclusiones

En base a trabajos anteriores realizados sobre la estimación del tamaño de modelos escritos en lenguaje natural aplicando un método de captura-recaptura, aún cuando parecía encontrarse un camino promisorio para encarar la incompletitud, el análisis de los resultados estadísticos no parecían ajustarse a la realidad. Partiendo de este análisis previo, se decidió mejorar las estimaciones ajustando los valores experimentales de cada muestra. Es decir, se resolvió determinar correctamente la cantidad de símbolos léxicos elicitados por cada grupo respecto de los otros, pues al manejar elementos

descriptos en lenguaje natural no suele ser tan fácil establecer si dos elementos de dos muestras distintas se refieren al mismo elemento o no.

De los resultados estadísticos presentados en este trabajo, se puede establecer que realizar un análisis semántico hace más confiable las estimaciones de completitud del modelo. Esta mejora en las estimaciones hacen más sólidas las conclusiones del trabajo original, en el sentido que los problemas de completitud son usualmente muy subestimados.

Las 9 muestras del modelo LEL fueron construidas en 1997 cuando la creación del glosario tenía guías básicas a seguir [16]. Dados los resultados, estas muestras no lograron asegurar que los elicitadores tuviesen una visión unificada del universo de discurso. El próximo paso será elaborar más muestras del modelo LEL para el mismo universo de discurso basados en un proceso de creación del LEL afianzado [17]. Esto permitirá establecer si disponer de guías más detalladas colabora en una actividad de elicitación cuyo resultado sea independiente del elicitor, permitiendo alcanzar un grado de completitud mayor. Por otro lado, el análisis semántico realizado en el presente trabajo será estudiado para elaborar nuevas guías en la creación de glosarios o refinar algunas presentes en [17]. Consideramos que, a partir de este trabajo, es necesario modificar el perfil de la heurística de construcción del LEL, la que hasta este momento es notoriamente sintáctica, debiéndosele incorporar más aspectos semánticos.

Una cuestión pendiente es la transferencia de esta experiencia a otros modelos que utilicen el lenguaje natural con estructuras que presenten distintos grados de formalidad.

Adicionalmente, se profundizará el análisis de la influencia del nivel de completitud de un modelo sobre otro modelo derivado de él. Este es el caso de estudiar el nivel de completitud de un conjunto de Escenarios Actuales construidos a partir de información derivada del modelo léxico. También se podría estudiar la completitud del conjunto de Escenarios Futuros partiendo de información proveniente de Escenarios Actuales.

Referencias

1. IEEE Std 830-1998 (R2009), IEEE Recommended Practice for Software Requirements Specifications (ANSI), IEEE, Nueva York, (1998) (R2009)
2. Kotonya, G., Sommerville, I.: Requirements Engineering: Process and Techniques. John Wiley & Sons (1998)
3. Loucopoulos, P., Karakostas, V.: System Requirements Engineering. McGraw-Hill, Londres (1995)
4. Firesmith, D.: Are Your Requirements Complete? Journal of Object Technology, vol. 4, n° 1, pp. 27- 43 (2005)
5. 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)
6. Otis DL, Burnham KP, White GC, Anderson DR: Statistical inference from Capture on Closed Animal Populations. Wildlife Monograph, 62 (1978)

7. Wohlin C, Runeson P: Defect content estimations from Review Data. En: 20th International Conference on Software Engineering, pp. 400-409, Japón (1998)
8. Briand, L., El Emam, K., Freimut, B., Laitenberger, O.: A Comprehensive Evaluation of Capture-Recapture Models for Estimating software Defects Contents. IEEE TSE, Vol 26, N° 6, pp. 518-540 (2000)
9. Petersson, H., Thelin, T., Runeson, P., Wohlin, C.: Capture-Recapture in Software Inspections after 10 Years Research - Theory, Evaluation and Application. The Journal of Software and Systems, vol. 72, pp. 249-264 (2003)
10. Walia, G.S., Carver, J.C.: Evaluation of Capture-Recapture Models for Estimating the Abundance of Naturally Occurring Defects. En: 2nd ACM-IEEE Intl Symposium of Empirical Software Engineering and Measurement, ISBN: 978-1-59593-971-5, pp.158-167, Alemania (2008)
11. Doorn, J.H., Ridaio, M.: Completitud de Glosarios: Un Estudio Experimental. En: VI Workshop on Requirements Engineering, pp. 317-328, Brasil (2003)
12. Ridaio M, Doorn JH: Estimación de Completitud en Modelos de Requisitos Basados en Lenguaje Natural. En: IX Workshop on Requirements Engineering, pp. 151-158. ISSN: 1413-9014, Brasil (2006)
13. Leite, J.C.S.P., Franco, A.P.M.: A Strategy for Conceptual Model Acquisition. En: IEEE 1st Intl Symposium on Requirements Engineering, IEEE Computer Society Press, pp 243-246, EEUU (1993)
14. Hadad, G.D.S., Litvak, C.S., Doorn, J.H.: Estudio semántico de modelos construidos por elicitadores independientes observando el mismo problema, Proyecto: Completitud de Modelos de Requisitos, Serie Documentos de Trabajo, N° 279, Departamento de Investigaciones, Universidad de Belgrano, Buenos Aires, ISSN: 1850-2512, 25 páginas (2012)
15. Litvak C.S., Hadad, G.D.S., Doorn, J.H.: Un abordaje al problema de completitud en requisitos de software XVIII Congreso Argentino de Ciencias de la Computación, ISBN: 978-987-1648-34-4, pp. 827-836, Bahía Blanca, Argentina (2012)
16. Hadad, G.D.S, Kaplan, G.N., Oliveros, A., Leite, J.C.S.P.: Integración de escenarios con el léxico extendido del lenguaje en la elicitación de requerimientos: aplicación a un caso real, Revista de Informática Teórica y Aplicada (RITA), ISSN 0103-4308, Vol.6, N°1, pp.77-103, Porto Alegre, Brasil (1999)
17. Hadad, G.D.S., Doorn, J.H., Kaplan, G.N.: Creating Software System Context Glossaries, en Encyclopedia of Information Science and Technology. Editorial: IGI Global, Mehdi Khosrow-Pour (ed), Information Science Reference, ISBN: 978-1-60566-026-4, 2° edición, Vol. II, pp. 789-794, Hershey, EEUU (2008)