



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance y final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	2.1
Vigencia	13/10/2015

Unidad Ejecutora:
Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT)

Título del proyecto de investigación:
Control de Calidad de requisitos escritos en Lenguaje Natural

Código del proyecto:
C175

Programa de acreditación:
PROINCE

Director del proyecto:
Doorn, Jorge Horacio

Co-Director del proyecto:
Litvak, Claudia Silvia

Integrantes del equipo:
Gigante, Nora; Guatelli, Renata; Kaplan, Gladys Noemí; Vera, Andrea Fabiana.
Alumnos becarios: Velazquez Santillán, Facundo; Lo Giudice, Diego; Milla,
Ximena;
Alumnos: Cáceres, María Silvina; Ortiz, Belén

Fecha de inicio: 2015/01/01

Fecha de finalización: 2016/12/31

Informe final

Sumario:

1. Resumen.....	Página 2
2. Memoria descriptiva.....	Página 2
3. Informe final.....	Página 5
4. Anexos.....	Página 16

Control de Calidad de requisitos escritos en Lenguaje Natural

1. Resumen

Como objetivo se había planteado la revisión crítica de los supuestos esenciales del proceso de requisitos en uso, a fin de determinar su vigencia y validez dado los resultados no satisfactorios obtenidos en casos de estudio anteriormente analizados. En esos momentos sólo se había obtenido una completitud del 51 %. Se planteó entonces la creación de una herramienta que facilite la generación de los modelos de requisitos bajo estudio. Durante el desarrollo de la herramienta se incorporaron los conceptos correspondientes al Procesamiento de Lenguaje Natural (PNL), y se buscaron aplicaciones de PNL que pudieran ser aplicados en la herramienta que estaba siendo desarrollada. Luego de un estudio comparativo, se seleccionó un framework que fue incorporado a la herramienta. En la etapa de creación del modelo inicial de requisitos, denominado Léxico Extendido del Lenguaje, se buscó interrelacionar entrevistas o documentos escritos con la generación de dicho modelo, aplicando PNL. Se realizaron estudios preliminares, sobre casos de estudio de dimensión mediana, obteniendo una importante mejora en la calidad de los modelos, logrando un 39% más de componentes de los mismos. Actualmente se están realizando estudios sobre casos reales de gran envergadura. Los resultados iniciales son muy promisorios.

Palabras claves: Ingeniería de Requisitos, Lenguaje Natural, Completitud de Modelos, Calidad.

Objetivo General:

Revisar críticamente los supuestos esenciales del proceso de requisitos en uso, procurando determinar su vigencia y validez a la luz de los resultados obtenidos en los casos de estudio ya realizados.

Objetivos específicos:

- a) Agregar de extensiones ocultas a los modelos construidos en lenguaje natural para facilitar las actividades de verificación y validación.
- b) Mejorar y extender el uso de las técnicas de estimación de completitud.
- c) Reducir el esfuerzo de creación de trazas incorporando esta actividad en el proceso de requisitos.
- d) Ponderar la importancia del modelado del proceso del negocio futuro en el proceso de requisitos.

2. Memoria descriptiva

Todas las actividades planificadas fueron realizadas a lo largo del proyecto, y todos los objetivos fueron alcanzados, si bien se enfatizó sobre el área de la estimación y mejora de la completitud, donde se lograron resultados conceptuales y prácticos de gran importancia.

Se incorporaron durante este proyecto nuevos elementos a los estudios planificados, el concepto de las Nominalizaciones y el Procesamiento de Lenguaje Natural. Esto implicó una gran mejora en los resultados obtenidos.

Desde junio de 2015 se han incorporado al proyecto 5 alumnos de cuarto año de la carrera Ingeniería Informática. Los mismos son Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice, Silvina Cáceres y Belén Ortiz. En particular los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice fueron seleccionados por la Universidad para

recibir Becas UNLaM, que cumplieron desde el 1º de septiembre de 2015 hasta el 30 de agosto de 2016, continuando luego en el desarrollo del proyecto hasta la finalización del mismo.

Se desarrollaron nuevas dos herramientas de software una para la comparación de dos modelos Léxico Extendido del Lenguaje (LEL), de modo semiautomático, y otra para la generación de todos los modelos bajo estudio, con la particularidad de ayudar al Ingeniero de Requisitos en la creación del modelo LEL, mediante la aplicación de técnicas y software específico de PNL. Ambas herramientas surgieron al incorporarse los becarios y alumnos al proyecto y no habían sido originalmente planeadas.

A partir de enero de 2016 se incorporó a la Investigadora Gladys Noemí Kaplan, que posee vasta experiencia en el área de investigación de requisitos.

Durante 2015 y 2016 se mantuvo la vinculación con el Profesor Julio Leite, del Grupo ER, de la PUC-Rio, Brasil. También se continuó la colaboración con la Mg. Marcela Ridaó, del INTIA, en UNICEN y con el Profesor Alejandro Oliveros, en UB y en UAI. Se continuó reprocesando los resultados de los trabajos de Alejandro Oliveros, de manera de poder compararlos con trabajos propios. Estos resultados aún no han sido publicados, están en etapa de preparación.

Durante 2015 hubo un total de 200 alumnos en el curso de grado “Ingeniería de Requerimientos” de UNLAM, y en 2016 hubo un total de 180 alumnos en el mismo curso de grado, realizándose una interesante transferencia de conocimientos, donde los alumnos aplicaron la estrategia utilizada por nuestro grupo de investigación para el desarrollo de modelos en lenguaje natural a casos de estudio específicos surgidos de empresas existentes de mediano tamaño, entrevistando clientes en varias ocasiones y obteniendo finalmente la Especificación de Requisitos correspondiente a cada negocio planteado, en base a los estándares de [IEEE 29148-2011]. Además se utilizaron resultados de las entrevistas planteadas por los alumnos en casos de estudio reales y se comparó lo obtenido por lo alumnos durante el curso, con lo obtenido mediante la herramienta generada con PNL. Es allí donde se logró elevar la completitud del 51 % al 89 %, en un caso de estudio real, a cuya empresa asistieron los alumnos de la materia a realizar entrevistas durante el primer cuatrimestre de 2016. Dichas entrevistas fueron transcritas y se les solicitaron los audios de las mismas para validar el contenido. Ese material transcrito por los alumnos fue la entrada a la nueva herramienta generada aplicando PNL.

También se estableció difusión en 2015 y 2016 a través de los cursos de grado “Ingeniería de Software V: Ingeniería de Requerimientos” de UB y en los cursos de posgrado “Ingeniería de Requerimientos” de UCA e “Ingeniería de Requisitos avanzados” de la Universidad Nacional de Luján.

2.1. Producción científico-tecnológica

Se presentan las publicaciones producidas en el marco de desarrollo de este proyecto de investigación, a saber:

2.1.2 Publicaciones

a) Artículos

Publicaciones en Actas de Congresos Nacionales con Referato

- **“Aporte del Procesamiento de Lenguaje Natural a la Ingeniería de Requisitos”**. Claudia S. Litvak, Jorge Doorn, Facundo Velazquez Santillán, Diego Lo Giudice, Ximena Milla, Silvina Cáceres. CONAISI 2016. 4to. Congreso Nacional de Ingeniería Informática / Sistemas de Información. Realizado en la Universidad Católica de Salta. ISBN: en trámite.
- **“Procesamiento de Lenguaje Natural para Estudiar Completitud de Requisitos”**. Actas del XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016), realizado en la Universidad Nacional de Entre Ríos, el 14 y 15 de abril de 2016. ISBN: 978-950-698-377-2. Extensión: 1094 p. Autores Claudia S. Litvak, Graciela D. S. Hadad, Jorge H. Doorn.
- **“Mapas Conceptuales para tratar Nominalizaciones en Modelos de Requisitos”**. Actas del XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016), realizado en la Universidad Nacional de Entre Ríos, el 14 y 15 de abril de 2016. ISBN: 978-950-698-377-2. Extensión: 1094 p. Autores Graciela D. S. Hadad, Alberto Sebastián, Claudia S. Litvak, Jorge H. Doorn.
- **“Influencia de las nominalizaciones sobre la completitud de modelos de requisitos”**. Claudia S. Litvak, Graciela D. S. Hadad, Jorge H. Doorn. WICC 2015: XVII Workshop de Investigadores en Ciencias de la Computación. Universidad Nacional de Salta. Salta. ISBN: 978-987-633-134-0.
- **“Interacción de atributos al priorizar requisitos a través de objetivos”**. Graciela D. S. Hadad, Jorge H. Doorn, Claudia S. Litvak, Carlos Imparato. WICC 2015: XVII Workshop de Investigadores en Ciencias de la Computación. Universidad Nacional de Salta. Salta. ISBN: 978-987-633-134-0.
- **“Agrupamientos en Glosarios del Universo de Discurso”**, Tecnología y Ciencia, Vol. 27, página 5 y siguientes, Revista de la Universidad Tecnológica Nacional, Mg. Marcela N. Ridaio, Ing. Jorge H. Doorn, ISSN: 1666-6917, 2015.

b) Capítulos de libro con Referato

- **“Nominalizations in Requirements Engineering Natural Language Models”**, en Encyclopedia of Information Science and Technology. Editorial: IGI Global, ed. Mehdi Khosrow-Pour, Information Science Reference, Hershey, PA, EEUU, 4º edición, a ser publicado en 2017, Mg. Litvak C.S., Dra. Hadad G.D.S., Ing. Doorn J.H.
- **“Dynamic Situational Adaptation of a Requirements Engineering Process”**, en “Encyclopedia of Information Science and Technology, Fourth Edition”, ed. Mehdi Khosrow-Pour, IGI Global, a ser publicado en 2017, Dra. Graciela D. S. Hadad, Ing. Jorge H. Doorn, Ing. Viviana A. Ledesma.
- **“Displaying Hidden Information in Glossaries”**, en “Encyclopedia of Information Science and Technology, Fourth Edition”, ed. Mehdi Khosrow-Pour, IGI Global, a ser publicado en 2017, Mg. Marcela N. Ridaio, Ing. Jorge H. Doorn.

c) Libros

- **“Ingeniería de Requisitos”**, en redacción, Dr. Julio Cesar Sampaio do Prado Leite, Dra. Graciela D. S. Hadad, Ing. Jorge H. Doorn, publicación estimada 2018.

d) Asistencia a Congresos

- **CONAISI 2016: 4to Congreso Nacional de Ingeniería Informática / Sistemas de Información.** 17 y 18 de noviembre de 2016. Autora y Expositora del paper “Aporte del Procesamiento de Lenguaje Natural a la Ingeniería de Requisitos”. Realizado en la Universidad Católica de Salta. ISSN: EN TRÁMITE.
- **WICC 2016: XVIII Workshop de Investigadores en Ciencias de la Computación.** 14 y 15 de abril de 2016. Autora y Expositora de los papers “Procesamiento de Lenguaje Natural para Estudiar Completitud de Requisitos” y “Mapas Conceptuales para tratar Nominalizaciones en Modelos de Requisitos”. Actas del XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016), realizado en la Universidad Nacional de Entre Ríos, el 14 y 15 de abril de 2016. ISBN: 978-950-698-377-2. Extensión: 1094 p.
- **WICC 2015: XVII Workshop de Investigadores en Ciencias de la Computación.** 16 y 17 de abril de 2015. Autora y Expositora de los papers “INFLUENCIA DE LAS NOMINALIZACIONES SOBRE LA COMPLETITUD DE MODELOS DE REQUISITOS” e “INTERACCIÓN DE ATRIBUTOS AL PRIORIZAR REQUISITOS A TRAVÉS DE OBJETIVOS”), realizado en la Universidad Nacional de Entre Ríos, el 14 y 15 de abril de 2016. ISBN correspondiente a los anales del WICC 2015 en formato e-book es 978-987-633-134-0.

e) Asistencia a Jornadas

- **1er Taller dirigido a Directores y Codirectores de proyectos de Investigación de UNLaM “¿Para qué investigamos en la UNLaM?** Secretaría de Ciencia y Tecnología. UNLaM. 19 de mayo de 2016.

3. Informe Final

El plan original del proyecto tenía como objetivo facilitar las actividades de verificación y validación, mejorar y extender el uso de las técnicas de estimación de completitud, reducir el esfuerzo de creación de trazas incorporando esta actividad en el proceso de requisitos, y ponderar la importancia del modelado del proceso del negocio futuro en el proceso de requisitos.

A lo largo de todo el proyecto se trabajó en todos los objetivos específicos con diferente grado de éxito, ya que en el dominio de la estimación y mejora de la completitud se lograron resultados conceptuales y prácticos de gran importancia.

En este sentido se incorporó un nuevo concepto nunca antes considerado en el estudio de la completitud de modelos, consistente en el estudio de las Nominalizaciones. La Nominalización consiste en la construcción de sustantivos, a partir de verbos o adjetivos [Alexiadou01] [Azpiazu02] [Bisetto05] [NGRAE09]. Se consideraron entonces las ventajas e inconvenientes que acarrea el uso de las Nominalizaciones de los verbos en los modelos construidos, ya que se supone que un uso poco cuidado de las mismas puede desencadenar en la generación de omisiones significativas, es decir pérdida de información.

En el área de creación de trazas en forma automática o semiautomática se continuó desarrollando la aplicación reflexiva de la metodología a la metodología misma, ampliando la versión original creada en la Universidad Nacional del Centro de la Provincia de Buenos Aires hace ya más de una década. Es decir se ha trabajado en la actualización del Léxico Extendido del Lenguaje (LEL) del LEL, del LEL de los Escenarios Actuales, del LEL de los Escenarios Futuros, del documento de especificación de requisitos; junto con los Escenarios Actuales de la construcción de esos mismos modelos. En este punto se planifica una bifurcación de estas actividades donde uno de los caminos avanzaría hacia el diseño de una herramienta y el otro hacia la identificación de puntos de extracción de trazas.

A partir de junio de 2015 se han incorporado al proyecto 5 alumnos de cuarto año de la carrera Ingeniería Informática. Los mismos son los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice, Silvina Cáceres y Belén Ortiz. Como consecuencia de esta incorporación y de algunos resultados intermedios en el área de completitud, 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. Desde esa fecha asistieron una vez por semana al Pramin I. En particular los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice fueron seleccionados por la Universidad para recibir Becas UNLaM, que cumplieron desde el 1º de septiembre de 2015. La propuesta de las becas involucra que el alumno Facundo Velazquez Santillán, haya tenido 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, tuvieron 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 el estudio de herramientas de Procesamiento de Lenguaje Natural (PNL), que puedan aplicarse, en particular aquellos que sean software libre. Las alumnas Silvina Cáceres y Belén Ortiz se han dedicado inicialmente al estudio de las técnicas de PNL, y estuvieron avocadas a la búsqueda y evaluación de herramientas de software libre de PNL, así como la factibilidad de ser aplicadas en el contexto del proyecto. Esto presenta también una nueva visión, nunca antes considerada, al incorporar las estrategias de PNL.

Se puede concluir que el trabajo realizado durante 2015 fue fructífero, centrándose en la profundización de estudios de completitud, al incorporar el estudio de las nominalizaciones y del PNL.

Como resultado del estudio inicial de las Nominalizaciones se ha producido un artículo para el congreso WICC 2015 - Workshop de Investigadores en Ciencias de la Computación, realizado en Salta en abril de 2015, denominado "Influencia de las nominalizaciones sobre la completitud de modelos de requisitos" por la Mg. Claudia S. Litvak, la Dra. Graciela D. S. Hadad, y el Ing. Jorge H. Doorn, presentado en dicho congreso por la Mg. Claudia S. Litvak.

Posteriormente se profundizó el estudio de las Nominalizaciones, analizando la influencia de la nominalización en la calidad de los requisitos producidos. Un ingeniero de requisitos debe ser consciente de las diferencias sustanciales de significado, que a veces surgen al utilizar el modo nominal de un verbo o su modo verbal, ya que la acción y el efecto de una nominalización del verbo pueden producir sinónimos o incluso homónimos. Cuando el significado del verbo nominalizado se refiere a la *acción* entonces se estará en presencia de un sinónimo del verbo, pero cuando el significado sea el del *efecto* entonces se estará hablando de un estado o de un nuevo objeto, el que se constituirá en homónimo de la acción con alto riesgo de ocultar información relevante. Estos sinónimos u homónimos deben

considerarse en todas las actividades especialmente en aquellas relacionadas con el LEL. En particular en lo que se relaciona con comparaciones de modelos tiene intrínseca relación con los estudios realizados anteriormente por este grupo de investigación sobre la completitud del modelo LEL. Se muestra en la Tabla 1 la guía generada para manejar verbos y sus nominalizaciones.

Tabla 1. Guía para manejar verbos y nominalizaciones

Forma sintáctica	Significado	Acción
Usado como verbo	Acción de	Registrarlo como un símbolo de tipo verbo
	Efecto de	N/A
	Ambos	N/A
Usado como una nominalización	Acción de	Registrarlo como un símbolo de tipo verbo
	Efecto de	Registrarlo como un símbolo de tipo objeto o estado
	Ambos	Registrar ambos, como un símbolo de tipo verbo y como un símbolo de tipo objeto o de tipo estado. Etiquetar ambos como homónimos.
Usado como una nominalización con más de un nominalizado	Acción de	Registrar cada uno de ellos como símbolos de tipo verbo. Considerar que eventualmente una nominalización puede ser sinónimo de otra.
	Efecto de	Registrar cada uno de ellos como símbolos de tipo objeto o de tipo estado. Considerar que eventualmente una nominalización puede ser sinónimo de otra.
	Ambos	Registrar cada uno de ellos como símbolos de tipo verbo y como símbolos de tipo estado. Considerar que eventualmente una nominalización puede ser sinónimo de otra.
Usado en ambas formas	Acción de	Registrar ambas formas como símbolos de tipo verbo, elegir la más usada como el nombre clave de entrada y el menos usado como el sinónimo.
	Efecto de	N/A
	Ambos	Registrar ambas formas como símbolos de tipo verbo y como como un símbolo de tipo objeto o estado. Considerar que la nominalización puede ser también un sinónimo de la forma de tipo verbo, entonces etiquetar ambas entradas como homónimos.
Usado en ambas formas con más de una nominalización	Acción de	Registrar todas las formas como símbolos de tipo verbo. Considerar que toda nominalización puede ser un sinónimo de la forma de tipo verbo; seleccionar la forma más usada como el nombre clave de entrada y la otra como sinónimo. Etiquetar homónimos si es aplicable.
	Efecto de	N/A
	Ambos	Registrarlos como un símbolo de tipo verbo y como un símbolo de tipo objeto o de tipo estado. Considerar que toda nominalización puede ser un sinónimo de la forma de tipo verbo; seleccionar la forma más usada como el nombre clave de entrada y la otra como sinónimo. Considerar que eventualmente toda nominalización puede ser el sinónimo de otra. Etiquetar homónimos si es aplicable.

Como resultado de la profundización del estudio de Nominalizaciones se ha escrito el capítulo para la Encyclopedia of Information Science and Technology, de Editorial IGI Global, denominado “Nominalizations in Requirements Engineering Natural Language Models” cuyos autores son la Mg. Claudia S. Litvak, Dra. Graciela D. S. Hadad y el Ing. Jorge Doorn. El capítulo propuesto ha sido recientemente aceptado y será publicado en la cuarta edición de dicha enciclopedia en el año 2017.

Durante 2015 se han obtenido los siguientes resultados:

- Herramienta de generación de modelos en lenguaje natural (LEL y Escenarios) en proceso de desarrollo, ya realizada para el LEL.
- Incorporación a dicha herramienta de PNL, para obtener los símbolos del LEL a partir de un documento, en etapa inicial de desarrollo.
- Herramienta de colaboración para la comparación de dos modelos escritos en lenguaje natural, aplicando Procesamiento de Lenguaje Natural (PNL), dado que se ha observado que el trabajo manual de comparación de estos modelos es inabordable, en etapa inicial de desarrollo.
- Estudio comparativo de herramientas de Procesamiento de Lenguaje Natural (PNL) de código abierto, ya finalizado.
- Estudio de la interacción entre las herramientas que se están desarrollando y las herramientas de PNL, ya finalizado.
- Estudio estadístico de completitud de Escenarios Actuales, con resultados de completitud similares.
- Estudio estadístico de resultados obtenidos por investigaciones de UAI, de Alejandro Oliveros, utilizando los resultados obtenidos de generación de requisitos construidos en base encuestas realizadas por un grupo de alumnos a través de una cámara Gessell. El nivel de completitud obtenido fue un poco menor a estudios de completitud anteriores.

Respecto al presupuesto asignado fue utilizado para la presentación del artículo para el congreso WICC 2015 - Workshop de Investigadores en Ciencias de la Computación, realizado en Salta en abril de 2015, denominado “Influencia de las nominalizaciones sobre la completitud de modelos de requisitos”, presentado por la Mg. Claudia S. Litvak. Además se realizó la compra de una computadora que permitiera el desarrollo del nuevo sistema, incorporando PNL, ya que dicha incorporación implicó la necesidad de una computadora de gran capacidad y memoria.

En 2015 no ha habido dificultades con ningún miembro del equipo de trabajo.

Durante el primer año de trabajo se continuó la vinculación con el Profesor Julio Leite, del Grupo ER, de la PUC-Rio, Brasil. Además en Argentina se trabajó en colaboración con la Mg. Marcela Ridao, del INTIA, en UNICEN y con el Profesor Alejandro Oliveros, en UB y en UAI. En particular de los trabajos de Alejandro Oliveros se han utilizado los resultados obtenidos de generación de requisitos construidos en base encuestas realizadas por un grupo de alumnos a través de una cámara Gessell. Con estos datos se aplicaron sobre ellos estudios estadísticos de completitud.

Durante 2015 hubo un total de 200 alumnos en el curso de grado “Ingeniería de Requerimientos” de UNLAM, logrando una interesante transferencia de conocimientos, dado que todos ellos aplicaron la estrategia utilizada por nuestro grupo de investigación para el desarrollo de modelos en lenguaje natural a casos de estudio específicos surgidos de empresas existentes de mediano tamaño, entrevistando clientes en varias ocasiones y obteniendo finalmente la Especificación de Requisitos correspondiente a cada negocio planteado, en base a los estándares de [IEEE 29148-2011].

También se estableció en 2015 la difusión a través de los cursos de grado “Ingeniería de Software V: Ingeniería de Requerimientos” de UB y el cursos de posgrado “Ingeniería de Requerimientos” de UCA.

En 2016 se continuó trabajando con los 5 alumnos de la carrera Ingeniería Informática. Los mismos son los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice, Silvina Cáceres y Belén Ortiz. Siguieron trabajando con la beca los alumnos Facundo Velazquez Santillán, Ximena Milla, Diego Lo Giudice. Continuaron asistiendo una vez por semana al Pramin I, Silvina Cáceres y Belén Ortiz. Desde fines de agosto, los becarios también siguieron participando del proyecto en forma ad honores. Facundo Velazquez Santillán, finalizó la creación de la herramienta informática para facilitar 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, finalizaron la construcción de una herramienta que permite la comparación de modelos de requisitos escritos en lenguaje natural, en particular para los modelos LEL y Escenarios, para facilitar los estudios de Completitud. Se realizó un intenso estudio comparativo de herramientas de Procesamiento de Lenguaje Natural (PNL). Las alumnas Silvina Cáceres y Belén Ortiz se han dedicado al estudio de las técnicas de PNL, a la búsqueda y evaluación de herramientas de software libre de PNL, así como la factibilidad de ser aplicadas en el contexto del proyecto. También se desarrollaron los manuales de las herramientas definidas anteriormente.

Posteriormente se inició el desarrollo de un caso de estudio de gran envergadura, realizado separadamente por Facundo Velázquez Santillán, Silvina Cáceres y Ximena Milla. Esto se encuentra aún en desarrollo, dado el extenso trabajo que implica.

Además, en la búsqueda de mejora de la calidad de los modelos se comenzó a pensar en las características del ser humano que realiza los modelos, por lo que la Mg. Claudia S. Litvak realizó reuniones en enero y marzo de 2016 con la Directora de Psicopedagogía del Instituto de Neurociencia de la Fundación Favaloro, la Lic. Laura Steinberg, perteneciente al equipo del Dr. Facundo Manes. Se evaluaron diferentes estudios cognitivos, se procedió a estudiar teorías de SKINNER (Aprendizaje) y de Chomsky (INNATO). Chomsky plantea que el ser humano tiene el conocimiento como algo innato, mientras que Skinner plantea que es un proceso de aprendizaje continuo, dependiendo del contexto donde la persona se forme y de los estímulos que recibió a lo largo de su vida. Se continuó esta nueva línea hasta arribar al concepto de Ergonomía Cognitiva, como visión innovadora.

La Ergonomía Cognitiva pertenece dentro de la psicología, a la Psicología Experimental. Cuando se evalúa el concepto de Ergonomía, se la considera como una disciplina científica que estudia el diseño de las herramientas y los sistemas con los que las personas realizan su trabajo. Etimológicamente el término Ergonomía deriva de las palabras griegas “ergon”, que significa trabajo, y “nomos”, que significa ley. Inicialmente en Estados Unidos se la denominaba como “Factores humanos” pero hoy en día ambos términos son sinónimos. Dentro de la Ergonomía se plantea la existencia de tres dominios: i) La Ergonomía Física, que se refiere a las condiciones físicas que deben tener los elementos con los que trabaja una persona, ii) la Ergonomía Cognitiva, conocida como Ergonomía Psicológica y iii) la Ergonomía Organizacional, que se refiere a la optimización de los sistemas socio-técnicos, incluyendo sus estructuras organizativas, políticas y procesos. Si bien se ha iniciado esta línea de trabajo, queda aún mucho por desarrollar.

Durante 2016 se han obtenido los siguientes resultados:

- Finalización de la herramienta de generación de modelos en lenguaje natural (LEL y Escenarios).
- Finalización de la Incorporación a dicha herramienta de PNL, para obtener los símbolos del LEL a partir de un documento.

- Finalización de la herramienta de colaboración para la comparación de dos modelos escritos en lenguaje natural, aplicando Procesamiento de Lenguaje Natural (PNL), dado que se ha observado que el trabajo manual de comparación de estos modelos es inabordable, en etapa inicial de desarrollo.
- Estudio de casos de aplicación de mediana envergadura aplicando la nueva herramienta.
- Estudios comparativos de la calidad de los modelos obtenidos mediante la aplicación de la herramienta con PNL y sin ella, obteniendo una mejora importante respecto a los casos en que no se aplicaba PNL.
- Inicio de la investigación del concepto de Ergonomía Cognitiva.

En 2016 no ha habido dificultades con ningún miembro del equipo de trabajo.

Se mantuvo durante 2016 la vinculación con el Profesor Julio Leite, del GrupoER, de la PUC-Rio, Brasil. También se continuó la colaboración con la Mg. Marcela Ridaio, del INTIA, en UNICEN y con el Profesor Alejandro Oliveros, en UB y en UAI. Se continuó evaluando los trabajos de Alejandro Oliveros, realizando algunos estudios estadísticos de mayor profundidad. Estos resultados aún no han sido publicados, están en etapa de preparación.

Durante 2016 hubo un total de 180 alumnos en el curso de grado “Ingeniería de Requerimientos” de UNLAM, realizándose nuevamente una interesante transferencia de conocimientos, donde los alumnos aplicaron la estrategia utilizada por nuestro grupo de investigación para el desarrollo de modelos en lenguaje natural a casos de estudio específicos surgidos de empresas existentes de mediano tamaño, entrevistando clientes en varias ocasiones y obteniendo finalmente la Especificación de Requisitos correspondiente a cada negocio planteado, en base a los estándares de [IEEE 29148-2011]. Además se utilizaron resultados de las entrevistas planteadas por los alumnos en casos de estudio reales y se comparó lo obtenido por los alumnos durante el curso, con lo obtenido mediante la herramienta generada con PNL. Es allí donde se logró elevar la completitud del 50% al 89%, en un caso de estudio real, a cuya empresa asistieron los alumnos de la materia a realizar entrevistas durante el primer cuatrimestre de 2016. Dichas entrevistas fueron transcritas y se les solicitaron los audios de las mismas para validar el contenido. Ese material transcrito por los alumnos fue la entrada a la nueva herramienta generada aplicando PNL.

También se estableció difusión en 2016 a través de los cursos de grado “Ingeniería de Software V: Ingeniería de Requerimientos” de UB y en los cursos de posgrado “Ingeniería de Requerimientos” de UCA e “Ingeniería de Requisitos avanzados” de la Universidad Nacional de Luján.

Respecto al presupuesto asignado fue utilizado para la presentación del artículo para el congreso WICC 2016 - Workshop de Investigadores en Ciencias de la Computación, realizado en Entre Ríos en abril de 2016, denominado **“Procesamiento de Lenguaje Natural para Estudiar Completitud de Requisitos”**, presentado por la Mg. Claudia S. Litvak.

En noviembre de 2016 se presentó el artículo denominado **“Aporte del Procesamiento de Lenguaje Natural a la Ingeniería de Requisitos”**, en el CONAISI 2016: 4to Congreso Nacional de Ingeniería Informática / Sistemas de Información, realizado en la Universidad Católica de Salta. La expositora del artículo fue la Mg. Claudia S. Litvak. La recepción general de los investigadores de Ingeniería de Software respecto a este artículo en el congreso fue de gran interés general y una amplia aceptación del mismo.

En 2016 no ha habido dificultades con ningún miembro del equipo de trabajo.

Conclusiones

Se puede concluir que los objetivos del proyecto se han cumplido, aunque con un cierto grado de asimetría, ya que los objetivos específicos relacionados con la completitud han superado las previsiones iniciales, mientras que aquellos relacionados con las trazas han sido algo escuetos. La disponibilidad de la herramienta facilitó la realización de comparaciones entre los resultados obtenidos por Ingenieros de Requisitos que apliquen PNL y aquellos Ingenieros de Requisitos que no lo apliquen.

Si en instancias anteriores se había conseguido una completitud individual del orden del 50%, el haber logrado una mejora del 39%, muestra la importancia de los logros obtenidos.

Queda pendiente la aplicación a más casos de estudio y a otros de mayor envergadura.

Se proponen realizar nuevos contactos con la Directora de Psicopedagogía del Instituto de Neurociencia de la Fundación Favaloro, la Lic. Laura Steinberg, perteneciente al equipo del Dr. Facundo Manes. Así como también ampliar la investigación del concepto de Ergonomía Cognitiva, y su aplicabilidad en la Ingeniería de Requisitos.

Bibliografía

a) Referencias

[Leite00] 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, Nº 1, pp. 38-61, Springer-Verlag London Ltd., Londres, Reino Unido, Julio 2000.

[Leite04] 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.

[Hadad13] 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, 3º edición, cap. 279, pp. 2854-2863, Julio 2014 (10384 p.). ISBN13: 9781466658882.

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

[ISO 25010-2011] ISO/IEC 25010:2011 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models, International Organization for Standardization, International Electrotechnical Commission, 2011.

[Alexiadou01] Alexiadou, A. (2001). Functional Structure in Nominals. Nominalization and ergativity. Amsterdam, The Netherlands. John Benjamins Publishers, ISBN 9789027227638.

[Azpiazu02] Azpiazu S. (2002). Las estrategias de nominalización y el adverbio: estudio contrastivo del caso español. Salamanca. Ediciones Universidad de Salamanca, ISBN 8478007873.

[Nettle00] Nettle, D., Romaine, S. (2000). *Vanishing Voices*. Oxford: Oxford University Press. ISBN10: 0195152468.

[Bisetto05] Bisetto A., Melloni C. (2005). *Result Nominals: a Lexical-Semantic Investigation*. Geert Booij et al. (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*, Fréjus 15-18 September 2005, University of Bologna, pp. 393-412.

[NGRAE09] REAL ACADEMIA ESPAÑOLA y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009). *Nueva gramática de la lengua española*, Madrid, Espasa Libros.

b) Bibliografía

- Alexander I., Maiden N. (eds.), “Scenarios, Stories, Use Cases. Through the Systems Development Life-Cycle”, 1º edición, John Wiley & Sons, Septiembre 2004.
- Alexander I., Robertson S., “Understanding Project Sociology by Modeling Stakeholders”, *IEEE Software*, Vol. 21, N°1, Enero/Febrero, 2004, pp.23-27.
- Antonelli L, Rossi G, Leite JCSP, Oliveros A, “Deriving requirements specifications from the application domain language captured by Language Extended Lexicon”, *WER 2012 – XV Workshop on Requirements Engineering*, ISBN: 978-987-1635-46-7, Buenos Aires, 2012.
- Antonelli, R.L., Rossi, G., Leite, J.C.S., Oliveros, A.: *Buenas prácticas en la especificación del dominio de una aplicación*. 16th Workshop on Requirements Engineering, Uruguay (2013)
- Castañeda, V.: *Un Framework basado en las Tecnologías de la Web Semántica para dar soporte a la Generación de Especificaciones de Requerimientos de Software*. Tesis de Maestría, Universidad Tecnológica Nacional, Facultad Regional Santa Fe (2012)
- Beltrán Cano JD, Calderón JE, Cabrera RJ, Moreno Vega JM (2003) Reglas de parada dependientes e independientes del problema. Estudio comparativo para el Strip Packing Problem. X Conferencia de la Asociación Española para la Inteligencia Artificial - CAEPIA 2003, Vol. II, pp 1-10, ISBN: 84-8373-564-4, San Sebastián, España
- Boender CGE, Rinnooy Kan AHG, Vercellis C (1986) Stochastic Optimization. *Methods. Stochastics in Combinatorial Optimization*, pp 94-112
- Briand L, El Emam K, Freimut B, Laitenberger O (2000) A Comprehensive Evaluation of Capture-Recapture Models for Estimating software Defects Contents. *IEEE TSE*, 26(6):518-540
- Brinkkemper S.: *Method Engineering: Engineering of Information Systems Development Methods and Tools*. *Information and Software Technology* 38(4), 1996, 275-280.
- Bucher T., Klesse M., Kurpjuweit S., Winter R. *Situational Method Engineering. On the Differentiation of “Context” and “Project Type”*, *IFIP International Federation for Information Processing*, Vol. 244/2007, 33-48, DOI: 10.1007/978-0-387-73947-2_5, 2007.
- Coulin C.R. (2007) *A situational approach and intelligent tool for collaborative requirements elicitation*, Doctoral Thesis, University of Technology, Sydney.
- De Almeida Ferreira David, Rodrigues Da Silva Alberto. “RSL-PL: A linguistic pattern language for documenting software requirements”. 3rd International Workshop on Requirements Patterns, RePa 2013 - Proceedings (2013)
- Espada Patrícia , Goulão Miguel, Araújo João. “A framework to evaluate complexity and completeness of KAOS goal models”. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2013)

- Gotel O.C.Z., Finkelstein A.C.W., "An analysis of the requirements traceability problem", ICRE'94, First IEEE International Conference on Requirements Engineering, IEEE Computer Society Press, Colorado Springs, Abril 1994, pp.94-101.
- Hadad G.D.S., Doorn J.H., Kaplan G.N., "Creating Software System Context Glossaries", Encyclopedia of Information Science and Technology, editor Mehdi Khosrow-Pour, D.B.A., Information Science Reference, 2º edición, 2008.
- Hadad G.D.S., Doorn J.H., "Introducing Variability in a Client-Oriented Requirements Engineering Process", ER@BR2013 - Requirements Engineering @ Brazil, Pontifícia Universidade Católica do Rio de Janeiro, ISSN: 1613-0073, Río de Janeiro, Brasil, pp.8-13, 2013.
- Hadad GDS, Litvak CS, Doorn JH, Ridao Marcela (2014) Dealing with Completeness in Requirements Engineering, (2014) igi-global.com p. 2854-2863.
- Hadad G.D.S., Ledesma V., Doorn J.H., "Proceso de Requisitos Adaptable a Factores Situacionales", WICC 2014 – XVI Workshop de Investigadores en Ciencias de la Computación, ISBN: 978-950-34-1084-4, Universidad Nacional de Tierra del Fuego, Antártida e Islas del Atlántico Sur, Ushuaia, Tierra del Fuego, 2014.
- Hadad G.D.S., Riera G.A., Doorn J.H., "Priorizar Requisitos: un Estudio sobre sus Propósitos", CONAISI 2014 – 2º Congreso Nacional de Ingeniería Informática / Sistemas de Información, Universidad Nacional de San Luis, San Luis, 2014.
- IEEE 29148-2011: IEEE Systems and software engineering - Life cycle processes - Requirements engineering. IEEE, Nueva York (2011)
- Kaplan G.N., Doorn J.H., Hadad G.D.S.: Handling Extemporaneous Information in Requirements Engineering. Encyclopedia of Information Science and Technology, editor Mehdi Khosrow-Pour, D.B.A., Information Science Reference, 2º edición, pp.1718—1722, 2008.
- Kaplan, G.N., Doorn, J.H., Gigante, N., "Elicitación de Conocimiento Guiada por Modelos", CACIC 2014 - XX Congreso Argentino de Ciencias de la Computación, San Justo, 2014.
- Leite JCSP, Moraes EA, Castro CEPS, "A Strategy for Information Sources Identification", X Workshop on Requirements Engineering (WER'07), Toronto, Canadá, Mayo 2007, pp.25-34.
- Litvak, C.S., Hadad, G.D.S., Doorn, J.H.: Un abordaje al problema de completitud en requisitos de software. En: XVIII Congreso Argentino de Ciencias de la Computación, pp. 827--836. Bahía Blanca (2012)
- Leonardi, M.C., Ridao, M., Mauco, M.V., Felice, L., Montejano, G., Riesco, D., Debnath, N.: An ATL Transformation from Natural Language Requirements Models to Business Models of a MDA Project. En: 11th International Conference on Telecommunications for Intelligent Transport Systems. Rusia (2011)
- Maté, J.L., Silva, A., "Requirements Engineering for Sociotechnical Systems", prefacio, Information Science Publishing, Maté & Silva editores, Londres, 2005.
- Nistala Padmalata, Kumari Priyanka. "An approach to carry out consistency analysis on requirements: Validating and tracking requirements through a configuration structure". 21st IEEE International Requirements Engineering Conference, RE 2013 - Proceedings (2013)
- Palmer J.D., "Traceability", en Software Engineering, editores M. Dorfman y R.H. Thayer, IEEE Computer Society Press, 1996, pp.266-276. Reimpreso en "Software Requirements Engineering", editores R.H. Thayer y M. Dorfman, IEEE Computer Society Press, 2º edición, Los Alamitos, CA, 1997, pp.364-374.
- Pinheiro FAC, "Requirements Traceability", en el libro "Perspectives on Software Requirements", Kluwer Academic Publishers, Estados Unidos, ISBN: 1-4020-7625-8, capítulo 5, 2004, pp.91-113.
- Ridao, M.N., Doorn, J.H., "Detección de Agrupamientos en Glosarios del Universo de Discurso", 2do Congreso Nacional de Ingeniería Informática / Sistemas de Información, CONAISI 2014, San Luis, 2014.

- Riera G.A., “Estudio Comparativo de Métodos de Priorización de Requisitos”, CONAIIISI 2014 – 2º Congreso Nacional de Ingeniería Informática / Sistemas de Información, en el rubro trabajo estudiantil de investigación Universidad Nacional de San Luis, San Luis, 2014.
- Welke RJ, Kumar K: "Method Engineering: a proposal for situation-specific methodology construction". In: Cotterman, Senn (eds.) Systems Analysis and Design: A Research Agenda. Wiley, Chichester. pp. 257–268, 1992.
- Ralyté J., Deneckère R., Rolland C. Towards a generic model for situational method engineering, In Advanced Information Systems Engineering, 1029–1029, 2003.
- Robertson, J., Robertson, S.: Volere Requirements Specification Template. The Atlantic Systems Guild Inc., edición 16 (2012) <http://www.volere.co.uk/template.htm> accedido 23/7/14
- Walia GS, Carver JC (2008) Evaluation of Capture-Recapture Models for Estimating the Abundance of Naturally Occurring Defects. 2nd ACM-IEEE Intl Symposium of Empirical Software Engineering and Measurement, Alemania, pp158-167. ISBN: 978-1-59593-971-5.
- Young R.R., “The Requirements Engineering Handbook”, Artech House, Norwood, 2004.

Anexo I

Formulario FPI-015



FORMULARIO DE RENDICIÓN DE GASTOS Y ADMINISTRACIÓN DE FONDOS

Unidad Académica que acredita el proyecto: DIIT							
Código de Proyecto: C175							
Título del Proyecto: Control de Calidad de requisitos escritos en Lenguaje Natural							
Director: Doorn, Jorge Horacio							
Fecha de inicio: 1/1/2015							
Fecha de finalización: 31/12/2016							
Periodo de la rendición: 2016							
1.- Insumos							
Nº de Orden	Folio Nº	Fecha	Proveedor o Prestador	Nº de Comprobante	Nº de CUIT	Descripción/ Concepto	Importe
1	1	22/3/16	LIBRERÍA CREACIÓN 2	00006599	30712358943	Librería	398,50
2	2	09/4/16	LIBRERÍA CREACIÓN 2	00006744	30712358943	Librería	24,00
3	3	11/4/16	TERCER MILENIO	00243040	30694124247	Librería	62,00
4	4	26/7/16	LIBRERÍA LUNAR	00055012	27134104703	Librería	69,00
5	5	2/12/16	TERCER MILENIO	0004-00071504	30694124247	Librería	339,00
6	6	12/12/16	TERCER MILENIO	0005-00006605	30694124247	Librería	1440,00
Total							2332,50
2.- Equipamiento							
Nº de Orden	Folio Nº	Fecha	Proveedor o Prestador	Nº de Comprobante	Nº de CUIT	Descripción/ Concepto	Importe
7	7	4/03/16	DiamondSystem S. A.	0016-00002984	30711050767	4 GB de memoria para Computadora de escritorio	447,00
8	8	07/12/16	TERCER MILENIO	0005-00006699	30694124247	Pen Drive	2160,00
Total							2607,00
3.- Contratación de Servicios Técnicos Especializados							
Nº de Orden	Folio Nº	Fecha	Proveedor o Prestador	Nº de Comprobante	Nº de CUIT	Descripción/ Concepto	Importe
Total							
4.- Viáticos							
Nº de Orden	Folio Nº	Fecha	Proveedor o Prestador	Nº de Comprobante	Nº de CUIT	Descripción/ Concepto	Importe
9	9	14/4/16	Universidad Nacional de Entre Ríos	00090030680	30562252157	Inscripción WICC 2016	300,00
10	10	13/4/16	Servicios de Taxi Bs. As.	27539	20111987042	Taxi de ida	78,78
11	11	13/4/16	Peaje Caminos del Uruguay S.A.	00353942	30639935872	Peaje viaje ida	20,00
12	12	13/4/16	AxionEnergy S.A.	00353942	30639935872	Bebida viaje ida	46,00

13	13	13/4/16	Charata SRL	00466346	3064702293	Almuerzo ida	58,00
14	14	13/4/16	Bar Ideal	00057972	2317662432	Café	45,00
15	15	13/4/16	Catering La Capilla S.A.	0080949	30700986604	Cena 13/4/16	115,00
16	16	14/4/16	Hotel Salto Grande	0040033513	30520435251	Hotel	573,00
17	17	14/4/16	Malleret José Luis	00211297	20163604655	Agua	15,00
18	18	14/4/16	Catering La Capilla S.A.	0081007	30700986604	Cena 14/4/16	135,00
19	19	14/4/16	Est. De Serv. YPF Campana Oeste	320600322123	30678774495	Nafta regreso	370,15
20	20	15/4/16	Petrovalle S.A.	00200609	30572365391	Café Agua	33,00
21	21	15/4/16	Peaje Caminos del Río Uruguay S.A.	00353942	30639935872	Peaje viaje vuelta 1	35,00
22	22	15/4/16	Peaje Caminos del Río Uruguay S.A.	02005436	30639935872	Peaje viaje vuelta 2	20,00
23	23	15/4/16	Servicios de Taxi Bs. As.	4078	2021674460	Taxi de vuelta	82,82
24	24	26/9/16	Aerolíneas Argentinas S. A.	357189	30641405554	Pasaje a Salta	2387,91
25	25	29/11/16	Universidad Católica de Salta	0001-00000685	30518591327	Inscripción CONAIISI 2016	700,00
26	26	21/11/16	Amerian Salta Hotel	0003-00004966	30714376469	Hotel	1630,89
27	27	16/11/16	Confitería City Bar	0002-00389451	30710354347	Almuerzo	124,00
28	28	16/11/16	Lo de Juana Manuela	0003-00006067	33709371059	Cena	223,00
29	29	17/11/16	El solar del convento SRL	00003251	30707957200	Almuerzo	65,00
30	30	18/11/16	Lo de Juana Manuela	0003-00006223	33709371059	Cena	193,00
31	31	16/11/16	Taxi a Aeroparque	002004	Patente PLW-343	Taxi	276,08
32	32	22/11/16	Remis a Aeropuerto en Salta	84	Patente LJJ-013	Taxi	175,38
33	33	17/11/16	Remis a UCASAL campus	84	Patente PMB-773	Taxi	88,98
34	34	18/11/16	Remis a UCASAL campus	8285	Patente PAM-742	Taxi	86,82
35	35	16/11/16	Carga Monedero para regresos de UCASAL	000000001313	Lote nro 1895	Red bus salta	30,00
Total							7907,81

5.- Bibliografía

Nº de Orden	Folio Nº	Fecha	Proveedor o Prestador	Nº de Comprobante	Nº de CUIT	Descripción/ Concepto	Importe

							Total	
6.- Otros								
Nº de Orden	Folio Nº	Fecha	Proveedor o Prestador	Nº de Comprobante	Nº de CUIT	Descripción/ Concepto	Importe	
36	36	30/8/16	Cerrajería GUILLER	0001-00000425	20086328284	5 llaves copias Pramin	200,00	
							Total	200,00

Lugar y fecha.....

.....
Firma del Director de Proyecto

Aclaración:.....

DNI:.....

Anexo II

Documentación de alta/baja de integrantes del equipo de investigación

San Justo, 10 de marzo de 2016

At. Secretario de Investigaciones
Departamento de Ingeniería e Investigaciones Tecnológicas.
Dr. Daniel Giulianelli

De nuestra mayor consideración:

Por la presente solicitamos la incorporación de la **Mg. Kaplan, Gladys Noemí, número de DNI: 16.086.376**, al proyecto de investigación **C175 “Control de Calidad de requisitos escritos en Lenguaje Natural” a partir del 1/1/2016** del cual somos Director y Codirector respectivamente.

Debido a la experiencia que la docente presenta en proyectos similares (ha co-dirigido el C124 "Contextualización del Proceso de Requisitos" y dirigido el C144 "Estabilidad de los Documentos en el Proceso de Requisitos", ambos íntimamente relacionados con el presente Proyecto), su inclusión mejorará las conclusiones a las que se arribe, colaborará en la capacitación de los RRHH e incorporará una línea de trabajo sobre jerarquías conceptuales en la construcción del LEL, que será incluida en los trabajos realizados.

Sin otro particular lo saluda atentamente.

Firma de Director

Aclaración

Firma de Codirector

Aclaración

Anexo III

Copias de certificados de participación de integrantes
en eventos científicos.

WICC 2015



UNIVERSIDAD
NACIONAL DE SALTA

WICC 2015



RedUNCI

XVII Workshop de Investigadores en Ciencias de la Computación

Se certifica que **Claudia Litvak (UNLaM), Graciela Hadaad (UNO), Jorge Doorn (UNO)** ha/n participado en calidad de autor/es del artículo ***Influencia de las nominalizaciones sobre la completitud de modelos de requisitos***, aceptado en el **XVII Workshop de Investigadores en Ciencias de la Computación - WICC 2015**, llevado a cabo los días 16 y 17 de Abril de 2015 en la ciudad de Salta, Argentina.

Ing. Armando E. De Giusti
Coordinador RedUNCI

Ing. Carlos E. Puiga
Decano de la Facultad de Ciencias Exactas
Universidad Nacional de Salta



UNIVERSIDAD
NACIONAL DE SALTA

WICC 2015

XVII Workshop de Investigadores en Ciencias de la Computación



RedUNCI

Se certifica que **Graciela Hada** (UNO, UB), **Jorge Doorn** (UNO), **Claudia Litvak** (UNO, UB), **Carlos Imparato** (UNO) ha/n participado en calidad de autor/es del artículo **Interacción de Atributos al Priorizar Requisitos a través de Objetivos**, aceptado en el XVII Workshop de Investigadores en Ciencias de la Computación - WICC 2015, llevado a cabo los días 16 y 17 de Abril de 2015 en la ciudad de Salta, Argentina.

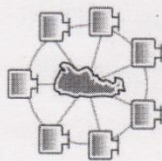
Ing. Armándo E. De Giusti
Coordinador RedUNCI

Ing. Carlos E. Puga
Decano de la Facultad de Ciencias Exactas
Universidad Nacional de Salta



UNIVERSIDAD
NACIONAL DE SALTA

WICC 2015



RedUNCI

XVII Workshop de Investigadores en Ciencias de la Computación

Por cuanto **Litvak Claudia Silvia, DNI 16.131.447** asistió en calidad de **Expositor** al **XVII Workshop de Investigadores en Ciencias de la Computación - WICC 2015**, llevado a cabo los días 16 y 17 de Abril de 2015 en la ciudad de Salta, Argentina.

Se extiende el presente certificado.

Ing. Armando E. De Giusti
Coordinador RedUNCI

Ing. Carlos E. Puga
Decano de la Facultad de Ciencias Exactas
Universidad Nacional de Salta

WICC 2016



Facultad de Ciencias
UNER de la Administración

WICC
2016

XVIII Workshop
de Investigadores
en Ciencias de la
Computación



Se certifica que **Claudia S. Litvak (UNLaM), Graciela Hada (UB), Jorge Doorn (UNLaM)**, han participado en calidad de AUTORES del artículo **“Procesamiento de Lenguaje Natural para Estudiar Complejidad de Requisitos”**, aceptado en el XVIII Workshop de Investigadores en Ciencias de la Computación, llevado a cabo los días 14 y 15 de abril de 2016 en la ciudad de Concordia, Entre Ríos.

Concordia, 15 de abril de 2016.

LIC. GUILLERMO E. FEIRHERD
COORDINADOR
RED DE UNIVERSIDADES
CON CARRERAS EN INFORMÁTICA

CP. HEROLDO B. FINK
MECANO
FACULTAD DE CIENCIAS DE LA ADMINISTRACIÓN
UNIVERSIDAD NACIONAL DE ENTRE RÍOS



UNER Facultad de Ciencias
de la Administración

WICC
2016

XVIII Workshop
de Investigadores
en Ciencias de la
Computación



Se certifica que **Graciela Hada (UB), Alberto Sebastián (UB), Claudia S. Litvak (UB), Jorge Doorn (UNO)**, han participado en calidad de AUTORES del artículo **“Mapas Conceptuales para tratar Nominalizaciones en Modelos de Requisitos”**, aceptado en el XVIII Workshop de Investigadores en Ciencias de la Computación, llevado a cabo los días 14 y 15 de abril de 2016 en la ciudad de Concordia, Entre Ríos.


Concordia, 15 de abril de 2016.

LIC. GUILLERMO E. FEIRHERD
COORDINADOR
RED DE UNIVERSIDADES
CON CARRERAS EN INFORMÁTICA

GRACIELA HADA
DECANO
FACULTAD DE CIENCIAS DE LA ADMINISTRACIÓN
UNIVERSIDAD NACIONAL DE ENTRE RÍOS

Por la presente dejo constancia que **CLAUDIA LITVAK (UNLaM)**, ha participado en carácter de asistente a las **XVIII WORKSHOP DE INVESTIGADORES EN CIENCIAS DE LA COMPUTACIÓN**, organizadas por la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos, los días 14 y 15 de abril de abril de 2016, en Concordia, Entre Ríos.

Se extiende la presente constancia para presentar ante quien corresponda, en la ciudad de Concordia, a los 14 días del mes de abril de 2016.

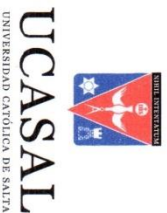

Lic. GUSTAVO P. LEONARDI
Secretario Académico
Facultad de Ciencias de la Administración
Universidad Nacional de Entre Ríos

CONAISI 2016

4to Congreso Nacional de Ingeniería en Informática /
Sistemas de Información

CONAISI

2016 17 y 18 de Noviembre
Salta | Argentina



Se certifica que

**Claudia Silvia Litvak, Jorge Horacio Doorn, Facundo Darío Velazquez Santillán,
Diego Andrés Lo Giudice, Ximena Anabela Milla, María Silvina Cáceres**

Ha/n participado como autores del trabajo titulado

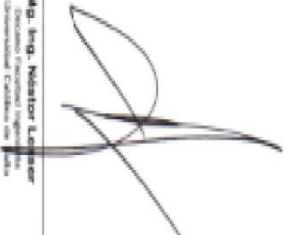
Aporte del Procesamiento de Lenguaje Natural a la Ingeniería de Requisitos

En el IV CONGRESO NACIONAL DE INGENIERÍA INFORMÁTICA / SISTEMAS DE INFORMACIÓN (CONAISI 2016), organizado por la Red de Carreras de Ingeniería Informática / Sistemas de Información (RIISIC) perteneciente al CONFEDI y realizado en la UNIVERSIDAD CATÓLICA DE SALTA, los días 17 y 18 de noviembre del 2016.

Se extiende el presente a los dieciocho días del mes de noviembre del 2016 en la ciudad de Salta, Argentina.


Mg. Ing. Esteban F. de Gallo
Catedrático de Ingeniería de Software
Universidad Católica de Salta


Mg. Dra. Guillermina Milena
Jefa Dept. Ficc. Ingeniería en Informática
Universidad Católica de Salta


Mg. Ing. Roberto Lanzetta
Director General Ingeniería de Software
Universidad Católica de Salta

4^{to} Congreso Nacional de Ingeniería en Informática /
Sistemas de Información

CONAISI

2016 17 y 18 de Noviembre
Salta | Argentina



Se certifica que : **CLAUDIA SILVIA LITVAK**

ha participado del **4º Congreso Nacional de Ingeniería en Informática /
Sistemas de Información**

en carácter de **ASISTENTE**


Mg. Ing. **Beatriz P. de Gallo**
Coordinadora del Congreso
Universidad Católica de Salta

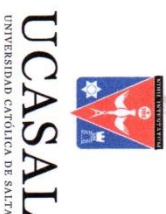

Ing. **Guillermina Nieva**
Jefa Dpto. Fac. Ingeniería en Informática
Universidad Católica de Salta


Mg. Ing. **Néstor Leisser**
Decano Facultad Ingeniería
Universidad Católica de Salta

4^{to} Congreso Nacional de Ingeniería en Informática /
Sistemas de Información

CONAIISI

2016 17 y 18 de Noviembre
Salta | Argentina



Se certifica que

Claudia Silvia Litvak,


Ha/n participado como EXPOSITOR del trabajo titulado

Aporte del Procesamiento de Lenguaje Natural a la Ingeniería de Requisitos

En el IV CONGRESO NACIONAL DE INGENIERÍA INFORMÁTICA / SISTEMAS DE INFORMACIÓN (CONAIISI 2016), organizado por la Red de Carreras de Ingeniería Informática / Sistemas de Información (RIISIC) perteneciente al CONFEDI y realizado en la UNIVERSIDAD CATÓLICA DE SALTÁ, los días 17 y 18 de noviembre del 2016.

Se extiende el presente a los dieciocho días del mes de noviembre del 2016 en la ciudad de Salta, Argentina.


Mg. Ing. Beatriz P. de Gallo
Coordinadora del Congreso
Universidad Católica de Salta


Ing. Guillermina Nieva
Jefa Dpto. Fac. Ingeniería en Informática
Universidad Católica de Salta


Mg. Ing. Nestor Lesser
Decano Facultad Ingeniería
Universidad Católica de Salta

1ER JORNADA PARA DIRECTORES Y CODIRECTORES DE PROYECTOS DE INVESTIGACIÓN DE UNLAM



II JORNADA DE INVESTIGACIÓN INTERDEPARTAMENTAL

¿Para qué investigamos en la UNLaM?

Hacia un modelo de investigación propio

1º Taller dirigido a Directores y Co-directores de proyectos de investigación UNLaM

Se certifica que **Litvak Claudia Siccia** participó en calidad de Co-Director/a en la

II Jornada de Investigación Interdepartamental

¿Para qué investigamos en la UNLaM? Hacia un modelo de investigación propio.

1º Taller dirigido a Directores y Co-directores de proyectos de investigación UNLaM,
organizada por la Secretaría de Ciencia y Tecnología, el día jueves 19 de mayo de 2016.




Mag. Ana Badiña

Secretaría de Ciencia y Tecnología
Universidad Nacional de La Matanza

Anexo IV

Copia de artículos presentados en publicaciones periódicas, y ponencias presentadas en eventos científicos

WICC 2015

Influencia de las nominalizaciones sobre la completitud de modelos de requisitos

Claudia S. Litvak^{1,2}, Graciela D. S. Hadad¹, Jorge H. Doorn^{1,2}

¹Escuela de Informática, UNO

Córdoba 1055, Merlo, Buenos Aires. (0220) 482-0799

²Departamento de Ingeniería e Investigaciones Tecnológicas, UNLaM

e-mail: claudialitvak@gmail.com, gracielahadad@gmail.com, jdoorn@exa.unicen.edu.ar

Resumen

Dentro de la Ingeniería de Software surge como área altamente relevante, el estudio de los requisitos de un sistema de software. Definir todos los requisitos de un sistema en un contexto determinado es prácticamente imposible. Meramente estimar el grado de completitud logrado es un problema muy complejo. Para abordar el problema de la completitud se suelen generar modelos, que son validados con clientes y usuarios, siendo de particular interés los modelos en lenguaje natural, ya que facilitan la comprensión por parte del usuario. Pero aún contando con esta facilidad de comprensión por parte del usuario y contando con ingenieros de requisitos expertos, el problema de completitud de los requisitos desarrollados es notorio y debe ser atendido. En el presente trabajo se propone considerar específicamente las ventajas e inconvenientes que acarrea el uso de las nominalizaciones de los verbos en los modelos construidos, ya que se supone que un uso poco cuidado de las mismas puede desencadenar omisiones significativas.

Palabras clave: Ingeniería de Requisitos, Completitud de Modelos, Nominalizaciones.

Contexto

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

Introducción

El proceso de la Ingeniería de Requisitos tiene como primera actividad conocer cabalmente el dominio de la aplicación, a fin de definir correctamente los servicios que deberá prestar el sistema de software que está siendo desarrollado.

Usualmente esta actividad involucra el desarrollo de modelos que reflejan dicho dominio.

La utilización del lenguaje natural en el modelado del contexto y de los requisitos de un sistema software facilita la comunicación con todos los involucrados dado que es comprendido por todos, y de ese modo los clientes y

usuarios pueden participar más eficazmente en ese proceso.

El presente proyecto se encuentra enmarcado en un proceso de requisitos desarrollado a lo largo de años y que ha sido utilizado en numerosos casos de estudio, tanto a nivel académico de grado y postgrado como en el ámbito profesional [Leite97] [Leite00] [Leite04]. Los modelos presentados por Leite et al. [Leite97] son el Léxico Extendido del Lenguaje (LEL) y los Escenarios, modelos que han probado ser idóneos para facilitar la comunicación entre los involucrados. El modelo LEL es un glosario con términos que corresponden al contexto donde el software se implantará, mientras que los escenarios son descripciones de situaciones en el universo de discurso, es decir, en dicho contexto.

El vocabulario intrínseco a un dominio particular transporta información. Esto se debe a que la cultura del dominio de la aplicación se materializa en un alto grado en su vocabulario [Nettle00] [Fishman99]. Es por ello fundamental, como parte de las actividades destinadas a conocer el dominio de la aplicación, adaptarse lo más posible a ese vocabulario. Leite enfatiza [Leite01] que el proceso de definición de los requisitos es un proceso inherentemente incompleto, dada la complejidad del mundo real. Por ello plantea que al no poderse alcanzar la completitud es deseable obtener modelos "lo más completos posible". Esto debe ocurrir con todos los modelos del proceso, comenzando por el modelo que inicialmente se construye: LEL.

Si bien estos conceptos han estado presentes desde las primeras etapas de definición del presente proceso de requisitos, en etapas avanzadas del desarrollo del mismo se ha hecho cada vez más evidente la importancia de la completitud de los modelos utilizados. El

proceso de construcción del modelo LEL ha sido ampliamente desarrollado en [Leite95] [Leite97] [Leite04], puntualizando en el estudio de la completitud de dicho modelo [Doom03] [Ridao06] [Doom08] [Hadad12] [Litvak12] [Hadad13] [Litvak13a] [Litvak13b] [Hadad14] [Litvak14] de modo de hallar la mayor cantidad de términos posibles de dicho glosario.

En la construcción del LEL, se define para cada término del LEL un nombre, con todos los sinónimos involucrados, su denotación y connotación. Luego se catalogan los términos símbolos según el tipo: sujeto, objeto, verbo o estado.

Con la experiencia de décadas trabajando con el LEL, utilizado en cursos de grado, cursos de posgrado y en aplicaciones reales, se han observado inconvenientes en el tratamiento de los verbos. Estos inconvenientes están relacionados principalmente con el uso de un término como verbo en su forma verbal o en su forma nominalizada.

La nominalización consiste en la construcción de sustantivos, a partir de verbos o adjetivos [Alexiadou01] [Azpiazu02] [Bisetto05] [NGRAE09]. El presente trabajo se aboca específicamente a las nominalizaciones de los verbos. La forma más sencilla de describir las nominalizaciones es mediante la frase *acción y efecto de*. Aunque en algunos casos las nominalizaciones son sólo la *acción de* o sólo el *efecto de*. El punto central es que no es lo mismo registrar en un glosario el verbo que sus nominalizaciones. Este aspecto es más notable a la hora de registrar los impactos. Un impacto describe la connotación del término del LEL, es decir, describe cómo el símbolo actúa en el dominio de la aplicación.

La primera idea sería confirmar que las nominalizaciones son en algunos casos más precisas que el verbo mismo,

en infinitivo o conjugado. Además, muchos verbos pueden producir más de una nominalización. Entonces queda por estudiar si realmente es específico del dominio de aplicación el verbo, el verbo más su nominalización (sustantivo obtenido o estado generado) o solamente la nominalización.

Líneas de Investigación, Desarrollo e Innovación

Se ha observado que al estudiar el mismo problema, diferentes personas utilizan el verbo mismo mientras que otras utilizan su nominalización. Esto es potencialmente una causa de ambigüedad, ya que el verbo en sí mismo puede arrastrar impactos del dominio de la aplicación diferentes de los que arrastra su versión nominalizada, especialmente cuando esta adquiere un significado de objeto o estado.

Posiblemente esto lleve a descubrir homónimos ocultos o a registrar por separado el verbo por un lado y el estado o el objeto producido por el otro. Seguramente, de ser este el caso se obtendrían una mayor cantidad de impactos y una mejora en la especificidad de los mismos.

A partir de todas estas razones ha surgido la presente línea de investigación, que aspira a mejorar la completitud del LEL mediante un mejor tratamiento de los verbos y sus nominalizaciones.

Resultados y Objetivos

Hasta la fecha se ha realizado un estudio del concepto de nominalización [Alexiadou01] [Azpiazu02] [Bisetto05] [NGRAE09] y se han avizorado las consecuencias que su uso no controlado es una posible fuente de deficiencias en el proceso de requisitos.

Las nominalizaciones se clasifican en dos grandes grupos: a) Eventivas y b) No eventivas. Las nominalizaciones Eventivas denotan la existencia de un proceso que ocurre o tiene lugar, de ahí que en la oración se combinen con predicados como *tener lugar*, *durar*, *ocurrir*, etc., o bien con adjuntos temporales sin la preposición *de*.

Las nominalizaciones No eventivas no indican un proceso. Pueden categorizarse en varios tipos. Uno de estos son las nominalizaciones de resultado, definidas en distintos trabajos a partir de la negación de las propiedades que caracterizan una nominalización Eventiva [Grimshaw90]. En segundo lugar, junto con las nominalizaciones de resultado, existen otras nominalizaciones No eventivas que no indican un resultado de la acción, sino que expresan diversos contenidos significativos, como objetos y estados, no resultantes de la acción significada en la base verbal, lugares o instrumentos.

La primera conclusión obtenida ha sido que las nominalizaciones son en algunos casos más precisas que el verbo en infinitivo y en otros igual de ambiguas. Son más precisas cuando su condición de Eventiva o No eventiva no es ambigua (ejemplo: agrupación/ agrupamiento), pero pueden ser también ambiguas cuando la condición de Eventiva o No eventiva depende del contexto (ejemplo: construcción). Eso es un aspecto importante en la reducción de la ambigüedad.

El objetivo del presente trabajo es estimar cómo influye en la completitud la presencia de la nominalización de un verbo, del verbo mismo o la presencia de ambas. Además, se pretende determinar si los impactos definidos pertenecen al verbo mismo o a su nominalización. Es posible que la migración de impactos entre verbos y sus nominalizaciones sea

una de las causas de la aparición de omisiones en el modelo LEL.

Asimismo, se aspira a continuar con el estudio de la completitud del LEL, avanzando hacia la completitud de los modelos posteriores como los Escenarios y el documento de Especificación de Requisitos de Software.

Esto último incluye mejorar los resultados ya logrados acerca de la influencia de la completitud del LEL sobre los modelos posteriores.

Formación de Recursos Humanos

El LITIC de la Universidad Nacional del Oeste, fue creado en 2014, con investigadores de amplia trayectoria en otras universidades. A la fecha, el Proyecto de Ingeniería de Requisitos del LITIC cuenta con tres investigadores. La propuesta es agregar mayor cantidad de investigadores, formando a docentes de la Universidad Nacional del Oeste como investigadores. También se ha solicitado la incorporación de becarios para que participen en este proyecto.

En la Universidad Nacional de La Matanza, desde hace 10 años existe un grupo de investigación del área de Ingeniería de Requisitos, perteneciente al Departamento de Ingeniería e Investigaciones Tecnológicas. Una de las ramas actuales de investigación de este grupo es la completitud. En particular actualmente hay integrantes del grupo avocados al estudio de las nominalizaciones, sobre dos casos prácticos específicos.

Además, este trabajo forma parte de la Tesis de Doctorado que la Magister Claudia Silvia Litvak está desarrollando en UNLP, la cual trata sobre la completitud de los modelos de requisitos y se denomina "Gestión de la completitud en la Ingeniería de Requisitos".

Referencias

[Alexiadou01] Alexiadou, A. (2001). *Functional Structure in Nominals. Nominalization and ergativity*. Amsterdam, The Netherlands. John Benjamins Publishers, ISBN 9789027227638.

[Azpiazu02] Azpiazu S. (2002). *Las estrategias de nominalización y el adverbio: estudio contrastivo del caso español*. Salamanca. Ediciones Universidad de Salamanca, ISBN 8478007873.

[Bisetto05] Bisetto A., Melloni C. (2005). *Result Nominals: a Lexical-Semantic Investigation*. Geert Booij et al. (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*, Fréjus 15-18 September 2005, University of Bologna, pp. 393-412.

[Doom03] Doom, J.H., Ridaio, M. (2003). *Completitud de Glosarios: Un Estudio Experimental*. 6th Workshop on Requirements Engineering Paracicaba, Brasil: Universidade Metodista de Paracicaba, pp. 317-328.

[Doom08] Doom, J.H., Ridaio, M.N. (2008). *Completeness Concerns in Requirement Engineering*. In Mehdi Khosrow-Pour (Ed.) *Encyclopedia of Information Science and Technology*, Second Edition Hershey, PA: IGI Global, pp. 619-624.

[Fishman99] Fishman, J. (1999). *Handbook of Language and Ethnic Identity*. Oxford: Oxford University press. ISBN10: 0195374924.

[Grimshaw90] Grimshaw, J. (1990). *Argument structure*. *Linguistic Inquiry Monographs* 18. Cambridge, MIT Press. ISBN10: 0262570904.

[Hadad12] Hadad G.D.S., Litvak C.S., Doom 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>.

[Hadad13] Hadad, G.D.S., Litvak, C.S., Doom, J.H., Ridaio 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, 3ª edición, cap. 279, pp. 2854-2863, Julio 2014 (10384 p.). ISBN13: 9781466658882.

- [Hadad14] Hadad G.D.S., Litvak C.S., Doorn J.H. (2014). Problemas y Soluciones en la Complejidad de Modelos en Lenguaje Natural. CADI 2014 – II Congreso Argentino de Ingeniería, Universidad Nacional de Tucumán, San Miguel de Tucumán, ISBN: 978-987-1662-51-7, T366, Septiembre 2014. Artículo destacado mediante la aceptación para su exposición oral, además de su publicación en los anales electrónicos. <https://cadi.org.ar/index.php/trabajos-seleccionados/16-tecnologia-de-la-informacion-y-comunicacion>.
- [Leite95] 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.
- [Leite97] 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.
- [Leite00] 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>
- [Leite01] 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.
- [Leite04] Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridaó, 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.
- [Litvak12] 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>.
- [Litvak13a] 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.
- [Litvak13b] Litvak, C.S., Hadad, G.D.S., Doorn, J.H. (2013). Mejoras semánticas para estimar la completitud de modelos en lenguaje natural. CoNaITSI 2013, 1er Congreso Nacional en Ingeniería Informática / Sistemas de Información. Córdoba, Argentina: Universidad Tecnológica Nacional.
- [Litvak14] 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>.
- [Nettle00] Nettle, D., Romaine, S. (2000). Vanishing Voices. Oxford: Oxford University Press. ISBN10: 0195152468.
- [NGRAE09] REAL ACADEMIA ESPAÑOLA y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009). Nueva gramática de la lengua española, Madrid, Espasa Libros.
- [Ridaó06] Ridaó, 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.

**Enciclopedia
IGI GLOBAL
A PUBLICARSE EN 2017**

Nominalizations in Requirements Engineering Natural Language Models

Claudia S. Litvak

Universidad Nacional de La Matanza, Argentina & Universidad Nacional del Oeste, Argentina

Graciela D. S. Hadad

Universidad Nacional del Oeste, Argentina & Universidad de Belgrano, Argentina

Jorge H. Doorn

Universidad Nacional de La Matanza, Argentina & Universidad Nacional del Oeste, Argentina

INTRODUCTION

The first activity of the development of a software system is to define its requirements. Requirements engineer must interact mainly with clients and users among other stakeholders (Macaulay, 1993). He or she has to understand the context in which the future system will act and he or she must carefully consider the reasons that conducted to the decision of developing such system. Requirements engineer's responsibilities include establishing a fluid communication with all stakeholders to produce reliable documents that will be used later in the software development process (Leite et al., 2004). Usually the culture, knowledge and skills of clients and users are rather different from those of the software development experts. As part of the communication with clients and users, the requirements engineer must clearly show them the characteristics of the software system that he or she is conceiving to attend clients and users' issues. Requirements Engineering process has two main activities: to understand the application domain, and to correctly define the services that the future software system will provide (Leite et al., 2004). As a general practice, both activities involve the development of models that describe such domain. It is also a common practice to develop such models in natural language to enhance the communication among stakeholders (Rolland & Ben Achour, 1998; Leite et al., 2004; Seiff et al., 2009); however, this introduces some obstacles, such as ambiguity, incompleteness and poor information structuring (Zowghi & Gervasi, 2002; Berry & Kamsties, 2004; Leite et al., 2005; Doorn & Ridao, 2009; Hadad et al., 2015). All of these inconveniences come from the natural language itself. As a consequence, Requirements Engineering has become more and more involved with linguistic considerations. Furthermore, it should be kept in mind that language conveys culture and knowledge (Nettle & Romaine, 2000; Fishman, 1999). Thereby, the terminology of clients and users holds application domain knowledge. Therefore, to document and to slightly formalize the relevant words or phrases heard from clients and users or read from documents is a valuable practice. In other words, creating a glossary of such terminology helps the Requirements Engineering process in two relevant ways: it eases the understanding of the application domain and it reduces the ambiguity of the oral communication with clients and users and the ambiguity of every produced document (Hadad, Doorn, & Kaplan, 2009). However, the glossary construction itself introduces some drawbacks. The most important of these drawbacks, not yet treated, is the presence of nominalizations, either in the clients and users' terminology or in the produced glossary. The former is a possible source of ambiguity while the latter is a hint of an adequate or inadequate creation of glossary symbols by the requirements engineer. Both may cause defects in the glossary. Nominalization refers to the construction of nouns from verbs or adjectives. Linguistic authors have largely studied nominalization in many languages such as English, French, German, Russian, Spanish, etc (Alexiadou, 2001; Bisetto & Melloni, 2005; Grimshaw, 1990; Rothmayr,

2009; Rozwadowska, 1997). The simplest way to describe verb nominalization is by means of the phrase *action of and effect of*. In some cases, nominalization occurs only by *action of*; while in others only occurs by *effect of*.

In this chapter, the influence of nominalization on the quality of Requirements Engineering documents is analyzed. The requirements engineer should be aware of the substantial differences in meaning, that sometimes arise when using the nominal mode of a verb or its verbal mode, since the action and the effect of a verb nominalization may produce synonyms or even homonyms.

BACKGROUND

The work presented in this chapter is based on the lessons learned in several research projects where many study cases were created using Scenarios and Language Extended Lexicon (LEL) models (Leite et al., 1997; Leite et al., 2000; Leite et al., 2004). LEL is a glossary proposed by Leite and Franco (1993). Scenarios are structured natural language descriptions of situations that occur in the application domain (Leite et al., 2000).

The LEL is itself a glossary with roles and structure different from the usual one. It is composed by a set of symbols, which are words or phrases peculiar and frequently used in the application domain. Each symbol is identified by a name or names. An acronym or an abbreviation may be also a name of a term, only if present in the application domain. In case of synonyms the more relevant name is used as the main key entry. Every symbol has two types of descriptions; this particular structure makes the difference with other glossaries. The first type, called Notion, is the usual one and describes the denotation of the word or phrase, that is, it defines *what the symbol is*. The second, called Behavioral Response, describes the connotation of the word or phrase, that is, it describes *how the symbol acts in the application domain*; this description is not usually present in other glossaries and enriches the knowledge about the symbol and the context at hand. LEL symbols contain hypertext links pointing to directly related entries.

When the denotation of the term acquires several meanings, it indicates the existence of homonyms, which forces the creation of more than one entry in the lexicon. The absence of any behavioral response indicates that the symbol does not belong to the LEL.

LEL entries are classified in four types according to its general use in the application domain. The types are: Subject, Object, Verb and State. Table 1 shows the LEL model.

Table 1. Language Extended Lexicon model (Leite et al., 2000)

LEL: representation of the symbols in the application domain language. Syntax: {Symbol} _i ^N
Symbol: entry of the lexicon that has a special meaning in the application domain. Syntax: {Name} _i ^N + {Notion} _i ^N + {Behavioral Response} _i ^N
Name: identification of the symbol. More than one represents synonyms. Syntax: Word Phrase
Notion: denotation of the symbol. Syntax: Sentence
Behavioral Response: connotation of the symbol. Syntax: Sentence

Symbols of type object, verb and state may be affected by a linguistic transformation, called nominalization, which may hide the type of the symbol and may produce confusions between verbs and objects, or between verbs and states. Nominalization is a source of ambiguities, especially in behavioral responses. It has not received the necessary attention in the Requirements Engineering literature (Berry &

Kamsties, 2004; Kovitz, 1998). This chapter is devoted to analyze and control the influence of nominalization on the quality of Requirements Engineering documents. Proposals to reduce the possible negative influence of nominalization are described.

The scenario model is a structure composed of the entities: title, goal, context, resources, actors, episodes, exceptions and the attribute constraint (see Table 2). Actors and resources are an enumeration. Title, goal, context and exceptions are declarative sentences while episodes are a set of sentences expressed in a simple language that give an operational description of behavior. A scenario must satisfy a goal that is reached by performing its episodes. Episodes represent the main course of action but they include also variations or possible alternatives. While performing episodes an exception may arise, signaling an obstacle to goal achievement. The treatment of the exception may or may not satisfy the original goal. All descriptions in the Scenario must maximize the use of LEL symbols, as a way to reduce ambiguity. This is another reason that strengthens the necessity of a high quality LEL creation.

Table 2. Scenario model

<p>Scenario: description of a situation in the application domain. Syntax: Title + Goal + Context + {Resources}₁^N + {Actors}₁^N + {Episodes}₂^N + {Exceptions}</p>
<p>Title: identification of the scenario. In the case of a sub-scenario, the title is the same as the episode sentence (see below in the Episode definition), without the constraints. Syntax: Phrase ([Actor Resource] + Verb + Predicate)</p>
<p>Goal: aim to be reached in the application domain. The scenario describes the achievement of the goal. Syntax: [Actor Resource] + Verb + Predicate</p>
<p>Context: composed by at least one of the following sub-components: Geographical Location: physical set of the scenario. Temporal Location: time specification for the scenario development. Precondition: initial state of the scenario. Syntax: (Geographical Location) + {Temporal Location} + {Precondition} where Geographical Location is: Phrase + {Constraint} where Temporal Location is: Phrase + {Constraint} where Precondition is: [Subject Actor Resource] + Verb + Predicate + {Constraint}</p>
<p>Resources: relevant physical elements or information that must be available in the scenario. Syntax: Name + {Constraint}</p>
<p>Actors: persons, devices or organisation structures that have a role in the scenario. Syntax: Name</p>
<p>Episodes: set of actions that details the scenario and provides its behaviour. An episode can also be described as a scenario. Syntax (using partial BNF): <episodes> ::= <group series> <episode series> <group series> ::= <group> <group> <non-sequential group> <group series> <group> <group> ::= <sequential group> <non-sequential group> <sequential group> ::= <basic sentence> <sequential group> <basic sentence> <non-sequential group> ::= # <episode series> # <episode series> ::= <basic sentence> <basic sentence> <episode series> <basic sentence> <basic sentence> ::= <simple sentence> <conditional sentence> <optional sentence> <simple sentence> ::= <episode sentence> CR <conditional sentence> ::= IF <condition> THEN <episode sentence> CR <optional sentence> ::= [<episode sentence>] CR where <episode sentence> is described: (([Actor Resource] + Verb + Predicate) ([Actor Resource] + [Verb] + Title)) + {Constraint}</p>
<p>Exceptions: usually reflect the lack or malfunction of a necessary resource. An exception hinders the achievement of the scenario goal. The treatment of the exception may be expressed through other scenario. Syntax: Cause [(Solution)] where Cause is: Phrase ([Subject Actor Resource] + Verb + Predicate) where Solution is: Title</p>
<p>Constraint: a scope or quality requirement referring to a given entity. It is an attribute of Resources, basic Episodes or sub-components of Context. Syntax: ([Subject Actor Resource] + Must [Not] + Verb + Predicate) Phrase</p>
<p>+ means composition, {x} means zero or more occurrences of x, () is used for grouping, stands for or and [x] denotes that x is optional</p>

VERB NOMINALIZATIONS

Verb nominalization may be interpreted as *eventive* with the semantic of *action of* and *stative* when it takes the semantic of *effect of*. Eventive nominalizations denote an activity and they are frequently combined with indications of location, duration or date. Eventive nominalizations inherit the arguments of the verbal base in such a way that they are capable of expressing the same semantic roles that the verb itself. Stative nominalizations are related with the result of the action of the verb. In most cases this result is an entity or a state. The following sentences contain the nominalization of the verb *publish*. The first sentence contains an eventive nominalization and the second a stative one.

- **The software system must register the number of copies of every *publication*.**
- **The *publication of books*, and even of journals, requires resources that have to be available on time in order to minimize costs.**

Ambiguity, completeness, and structuring of natural language documents are related among them. A slightly formalized glossary contributes to improve completeness since it contains a set of application domain notions materialized through subjects, objects, activities and states and their interrelations. The perception by the requirements engineer of any of those glossary elements helps to discover more information about the related glossary symbols. This clearly contributes to reduce incompleteness. The experience acquired along many study cases led to observe the weaknesses that occur by nominalization of verb symbols. The semantic flaws produced due to nominalizations are of such nature that sometimes is quite different to register the verb itself or its nominalization in the LEL. The actual consequences of verb nominalization may be:

- **Verb and its nominalization have exactly the same meaning**
- **Verb and its nominalization have different meaning**
 - a. Just verb semantic is relevant in the application domain
 - b. Just nominalization semantic is relevant in the application domain
 - c. Both verb and its nominalization semantic are relevant in the application domain
- **There is more than one nominalization for the same verb**

The purpose of the LEL is to represent the vocabulary used in the application domain, therefore, nominalizations could appear or not as part of that vocabulary. Nominalizations are not a problem provoked by the Requirements Engineering process or by the requirements engineer himself or herself. It is an obstacle that he or she must manage by identifying the variants of the term and by establishing the meaning of those variants in the application domain. When capturing the vocabulary used in the application domain, it could happen and it usually happens that the requirements engineer does not perceive one or more of the possible meanings of a nominalized verb. Once this happens, it is very likely that he or she describes only the captured meaning keeping all the omissions undiscovered.

SOLUTIONS AND RECOMMENDATIONS

Each one of the cases enumerated in the previous section should be treated in a particular way, as shown in Table 3. When the requirements engineer runs across with a verb symbol, he or she does not know in which case the verb falls. To avoid omissions or ambiguities he or she has to find out how the symbol is used in the application domain, and then to which case it belongs in order to take out the correct action.

Table 3. Guidelines to manage verbs and nominalizations

Syntactic form	Meaning	Action
Used as a verb	Action of	Register as a verb symbol
	Effect of	N/A
	Both	N/A
Used as a nominalization	Action of	Register as a verb symbol
	Effect of	Register as an object or state symbol
	Both	Register both as a verb symbol and as an object or state symbol. Label both as homonyms
Used as a nominalization with more than one nominalization	Action of	Register each of them as a verb symbols. Consider that eventually any nominalization may be a synonym of another
	Effect of	Register each of them as object or state symbols. Consider that eventually any nominalization may be a synonym of another
	Both	Register as a verb symbol and as an object or state symbols. Consider that eventually any nominalization may be a synonym of another
Used both forms	Action of	Register both forms as a verb symbol, choose the most used form as key entry name and the least used as synonym (see Table 3)
	Effect of	N/A
	Both	Register both as a verb symbol and as an object or state symbol (see Table 4). Consider that the nominalization may be also a synonym of the verb form, then label both entries as homonyms
Used both forms with more than one nominalization	Action of	Register all forms as verb symbols. Consider that any nominalization may be a synonym of the verb form; choose the most used form as the key entry name and the other as synonym. Label homonyms if applicable
	Effect of	N/A
	Both	Register as a verb symbol and as object or state symbols. Consider that one of the nominalizations may be a synonym of the verb form; choose the most used form as the key entry name and the other as synonym. Consider that eventually any nominalization may be a synonym of another. Label homonyms if applicable (see Table 5)

Table 3 strictly considers the possible consequences of the use of nominalizations without paying attention to the particular semantic of the verb itself or its nominalizations. Actually a verb or a nominalization may have more than one interpretation in the application domain, producing eventually homonyms. Therefore, any of the cells of third column of Table 2 may produce more symbols than it is indicated. Watching how the nominalizations are used is not enough to detect such additional homonyms cases. This happens with any single symbol of the LEL, regardless if it is the nominalization of a verb or not. The particular semantic of each symbol should be clearly understood in order to discover hidden or partially hidden homonyms, and possible synonyms.

Table 4. Verb and its nominalization used as action of, being both synonyms. Study Case: Production Management of Cardboard Boxes

Symbol Name/s	Plan Production / Production Planning	Type: Verb
Notion	<ul style="list-style-type: none"> • Process to establish the production schedule for all the shifts of the next week. • It is performed every Tuesday. 	
Behavioral Response	<ul style="list-style-type: none"> • Manufacturing programs for each of the 21 shifts per week are generated. • Deployments boxes are organized to reduce cardboard scrap. • Production orders received from a collaborative supplier are considered urgent orders. 	

Table 4 shows an example about the Study Case Production Management of Cardboard Boxes. In this example, the requirements engineer detects two different synonyms, used both in that context, for the same action: Plan Production and Production Planning.

Table 5. Verb and its nominalization used as effect of. Study Case: Management of Treatments given by Medical Providers

Symbol Name/s	Pay	Type: Verb
Notion	<ul style="list-style-type: none"> • Action performed by the treasury staff producing an economic compensation to a medical provider. • It takes place after the payment has been approved by the medical director. 	
Behavioral Response	<ul style="list-style-type: none"> • The treasury staff issues the payment thru the system. • The treasury staff notifies the medical provider upon the issued payment. 	
Symbol Name/s	Payment	Type: Object
Notion	<ul style="list-style-type: none"> • Document issued by the treasury staff, establishing the cancellation of a debt acquired due to a treatment given by a medical provider. 	
Behavioral Response	<ul style="list-style-type: none"> • It is registered in the system. • It requires the director approval before given to the medical provider. • A copy signed by the medical provider is stored. 	

In Table 5 there are two different symbols, one is a Verb symbol, Pay, and the other is an Object symbol, Payment, belonging to the same study case, Management of Treatments given by Medical Providers, where the Object refers to the verb nominalization used as effect of.

Table 6. Verb and two nominalizations used as action of and effect of, producing synonyms and homonyms. Study Case: Management of Treatments given by Medical Providers

Symbol Name/s	Authorize Treatment / Treatment Authorization	Type: Verb
Notion	<ul style="list-style-type: none"> Action that allows the patient to receive treatment. It is performed at the patient service office. 	
Behavioral Response	<ul style="list-style-type: none"> The staff checks the medical credit of the patient thru the system. The staff checks the clinical order given by the patient. The staff issues an authorization. The staff delivers the authorization to the patient. The patient annexes the clinical order to the authorization form in order to get its treatment. 	
Symbol Name/s	Treatment Authorization / Authorization	Type: object
Notion	<ul style="list-style-type: none"> Document issued by the patient service staff, allowing a patient to get a treatment. 	
Behavioral Response	<ul style="list-style-type: none"> It is generated thru the system based on the clinical order exhibited by the patient. It is signed by the service director. It is assigned a sequential number and an expiration date. 	
Symbol Name/s	Authorized Treatment	Type: State
Notion	<ul style="list-style-type: none"> Condition that a treatment takes after treatment authorization. 	
Behavioral Response	<ul style="list-style-type: none"> The patient can get a treatment thru a medical provider before its expiration date. After the patient gets the treatment, the medical provider issues an invoice to receive the payment. 	

In Table 6 the symbol Authorize Treatment is a Verb, and there are three different nominalizations. The first one, Treatment Authorization, is a synonym of Authorize Treatment, being both verbs. Treatment Authorization as a Verb is also the homonym of Treatment Authorization as an Object. The third nominalization has generated a State symbol Authorized Treatment.

FUTURE RESEARCH DIRECTIONS

This chapter presents the results of an exploratory research which has allowed discovering, as a regular pattern, the presence of many nominalizations of verbs that are relevant in the application domain. Frequently, such nominalizations have created different types of defects in the glossaries. The knowledge of the existence of those defects and the understanding of their cause has allowed proposing guidelines in order to reduce quantitative and qualitative consequences of such defects. These guidelines have been proven to be effective in few and small study cases. It is planned to perform a set of more structured experiments to quantify the efficiency of the proposed guidelines and to obtain reliable statistics and metrics of the quality enhancement of the glossaries and the importance of this on the whole Requirements Engineering process.

CONCLUSION

The research whose results are reported in this chapter may be summarized as:

- The relevant words and phrases of any application domain include several verbs.
- The full understanding of the verbs is important for improving the knowledge about the application domain.
- Nominalization of verbs obscures the perception of their eventive or stative semantic.
- Frequently one of the possible verb nominalizations hides the perception of the others.
- Losing an eventive or a stative meaning prevents discovering important information.

REFERENCES

- Alexiadou, A. (2001). *Functional Structure in Nominals. Nominalization and ergativity*. Amsterdam, The Netherlands: John Benjamins Publishers.
- Berry, D. M., & Kamsties, E. (2004). Ambiguity in requirements specification. In *Perspectives on software requirements* (pp. 7-44). Springer US.
- Bisetto, A., & Melloni, C. (2005, September). Result Nominals: A Lexical-Semantic Investigation. In *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)* (pp. 393-412). Bologna: University of Bologna.
- Doom, J. H., & Ridao, M. (2009). Completeness Concerns in Requirements Engineering. In Mehdi Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology, Second Edition* (pp. 619-624). Hershey, PA: IGI Global, Information Science Reference.
- Fishman, J. (1999). *Handbook of Language and Ethnic Identity*. Oxford: Oxford University press.
- Grimshaw, J. (1990). *Argument structure. Linguistic Inquiry Monographs 18*. Cambridge, MIT Press.
- Hadad, G. D., Doom, J. H., & Kaplan, G. N. (2009). Creating Software System Context Glossaries. In Mehdi Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology, Second Edition* (vol.II, pp.789-794). Hershey, PA: IGI Global, Information Science Reference.
- Hadad, G.D.S., Litvak, C.S., Doom, J.H., & Ridao, M.N. (2015). Dealing with Completeness in Requirements Engineering. In Mehdi Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology, Third Edition* (pp. 2854-2863). Hershey, PA: IGI Global, Information Science Reference.
- Kovitz, B. L. (1998). *Practical software requirements: a manual of content and style*. Greenwich: Manning.

- Leite, J. C. S. P., Doorn, J. H., Hadad, G. D., & Kaplan, G. N. (2005). Scenario inspections. *Requirements Engineering*, 10(1), 1-21.
- Leite, J. C. S. P., Doorn, J. H., Kaplan, G. N., Hadad, G. D., & Ridao, M. N. (2004). Defining System Context using Scenarios. In J. S. C. P. Leite, & J. H. Doorn (Eds.), *Perspectives on Software Requirements* (pp. 169–199). Norwell, MA: Kluwer Academic Press.
- Leite, J. C. S. P., & Franco, A. P. M. (1993, January). A strategy for conceptual model acquisition. In *Proceedings of IEEE International Symposium on Requirements Engineering* (pp. 243-246). San Diego, CA: IEEE Computer Society Press.
- Leite, J. C. S. P., Hadad, G. D., Doorn, J. H., & Kaplan, G. N. (2000). A Scenario Construction Process. *Requirements Engineering*, 5(1), 38-61.
- Leite, J. C. S. P., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G. N., Hadad, G. D. S., & Oliveros, A. (1997). Enhancing a Requirements Baseline with Scenarios. *Requirements Engineering Journal*, 2(4), 184–198.
- Macaulay, L. (1993). Requirements capture as a cooperative activity. In *Proceedings of IEEE International Symposium on Requirements Engineering* (pp. 174-181). San Diego, CA: IEEE Computer Society Press.
- Nettle, D., & Romaine, S. (2000). *Vanishing Voices*. New York: Oxford University Press.
- Rolland, C., & Ben Achour, C. (1998). Guiding the construction of textual use case specifications. *Data & Knowledge Engineering*, 25, 125–160.
- Rothmayr, A. (2009). *The structure of stative verbs* (Vol. 143). John Benjamins Publishing.
- Rozwadowska, B. (1997). *Towards a unified theory of nominalizations: External and internal eventualities* (No. 2031). Wdawn. Uniwersytetu Wrocawskiego.
- Seyff, N., Maiden, N., Karlsen, K., Lockerbie, J., Grünbacher, P., Graf, F., & Ncube, C. (2009). Exploring how to use scenarios to discover requirements. *Requirements Engineering*, 14(2), 91-111.
- Zowghi, D., & Gervasi, V. (2002). The Three Cs of Requirements: Consistency, Completeness, and Correctness. In *Proceedings of 8th International Workshop on Requirements Engineering: Foundation for Software Quality* (pp. 155-164). Essen, Germany: Essener Informatik Beitiage.

ADDITIONAL READING

- Alexiadou, A., & Rathert, M. (Eds.). (2010). *The syntax of nominalizations across languages and frameworks* (Vol. 23). Berlin: Walter de Gruyter.

Alexiadou, A. (2011). *The aspectual properties of nominalization structures*. Amsterdam: John Benjamins.

Browne G.J., Pitts M.G., & Wetherbe J.C. (2007). Cognitive stopping rules for terminating information search in online tasks. *MIS Quarterly*, 31(1), 89-104.

Firesmith, D. (2005). Are Your Requirements Complete? *Journal of Object Technology*, 4(1), 27- 43.

Giannakidou, A., & Rathert, M. (Eds.). (2009). *Quantification, definiteness, and nominalization* (Vol. 24). Oxford: Oxford University press.

Koptjevskaja-Tamm, M. (2002). *Nominalizations*. New York: Routledge.

Pustejovsky, J. (Ed.). (2012). *Semantics and the Lexicon* (Vol. 49). Dordrecht: Springer Science & Business Media.

Rathert, M., & Alexiadou, A. (Eds.). (2010). *The semantics of nominalizations across languages and frameworks* (Vol. 22). Berlin: Walter de Gruyter.

Sager, N. (1981). *Natural language information processing*. Massachusetts: Addison-Wesley Publishing Company, Advanced Book Program.

Zucchi, A. (2013). *The language of propositions and events: Issues in the syntax and the semantics of nominalization* (Vol. 51). Dordrecht: Springer Science & Business Media.

KEY TERMS AND DEFINITIONS

Ambiguity: Phrase or sentence that does not have a single clear meaning. It provokes two or more possible interpretations.

Application Domain: The context in which the software system will be used.

Completeness: The state or condition of having all the necessary or appropriate parts of a model or document.

Language Extended Lexicon: A glossary composed by a set of symbols, which are words or phrases peculiar and frequently used in a given application domain. The lexicon symbols have an additional description, not present in other glossaries. It contains hypertext links that interconnect symbols.

Natural Language Model: A model in narrative text produced during Requirements Engineering process, with a structure and content easily read by all stakeholders.

Nominalization: The result of forming a noun from a verb, or the act of forming a noun from a verb. Nominalization itself is the nominalization of the verb *nominalize*. In this chapter it is used as a verb and as the effect of the verb (a state).

Omission: Information that has not been included in a model or document. It is the nominalization of the verb *omit*. In this chapter it is used as the effect of the verb (a state).

Scenario: The description of a situation that currently happens or will happen in the application domain. The former are called Current Scenarios and the latter are referred as Future Scenarios.

LIST OF RESPONSES TO THE REVIEWERS' COMMENTS

The reviewer 1 comment "2. Please clarify the difference between the natural language with high context culture and that with low context culture in developing Requirement Engineering" was ignored since there is not such thing as a comparison in any place of the chapter between 'high context culture' and 'low context culture'.

The reviewer 1 comment "3. Please make a comment in performing a set of more structures experiments to qualify the efficiency of the proposed guidelines and to obtain reliable statistics and metrics of the quality enhancement of the glossaries" was also ignored since the chapter contribution are the results of an exploratory research which has allowed discovering, as a regular pattern, the presence of many nominalizations of verbs that are relevant in the application domain. In other words no claim was made about improving experiences in LEL construction.

The reviewer 2 comment "The issue of providing good requirements does not require much elaboration. The paper discusses methods to avoid ambiguity and other issues. However, the effectiveness of these methods is not provided" was ignored for the same reason given for reviewer 1 comment 3.

The reviewer 2 comment "The effectiveness of the proposed method is not provided." was ignored for the same reason given for reviewer 1 comment 3.

All the other comments were modified at the chapter.

WICC 2016

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] [Loucopoulos 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 [Doom 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] [Doom 03] [Ridao 06] [Hadad 12] [Hadad 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 Magister 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>.
- [Doom 03] Doom, 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.iot.fm/issuues/issue_2005_01/column3.
- [Hadad 12] Hadad G.D.S., Litvak C.S., Doom 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., Doom, 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, 3ª edición, cap. 279, pp. 2854-2863, Julio 2014 (10384 p.). ISBN13: 9781466658882.
- [IEEE 20148] 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, Nº 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, Nº 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.unf.edu/hinertexnet/numero-5/pjn.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., Rumeson, P. (1998). Defect content estimations from Review Data. 20th Intl Conference on Software Engineering, Japón, pp. 400-409.

CONAIISI 2016

Aporte del Procesamiento de Lenguaje Natural a la Ingeniería de Requisitos

Claudia Litvak^{1,2}, Jorge Doorn^{1,2}, Facundo Velazquez Santillán¹, Diego Lo Giudice¹,
Ximena Milla¹, Silvina Cáceres¹

¹ DIT, Universidad Nacional de La Matanza, Argentina

² Escuela de Informática, Universidad Nacional del Oeste, Argentina

clitvak@unlam.edu.ar, jdoorn@exa.unicen.edu.ar,

{fvelazquezantillan, dlogiudice, xmillam, silcaceres}@unlam.edu.ar

Abstract

En la Ingeniería de Software, la etapa inicial corresponde a la definición de los requisitos del sistema de software a desarrollar. Los ingenieros de requisitos generan modelos para representar el conocimiento adquirido y las decisiones tomadas en el proceso de construcción de requisitos. En particular, los modelos escritos en Lenguaje Natural son muy utilizados ya que facilitan la interacción con los clientes y usuarios. Si bien el Lenguaje Natural tiene amplios beneficios, también presenta algunas dificultades. Realizando estudios de completitud de los modelos escritos en Lenguaje Natural denominados Léxico Extendido del Lenguaje y Escenarios ha surgido la necesidad de definir estrategias específicas, tanto para crear los modelos como para realizar sobre ellos estudios de completitud. En el presente artículo se presentan resultados preliminares de mejora en la calidad de los modelos obtenidos al aplicar una herramienta de Procesamiento de Lenguaje Natural y las expectativas para trabajos futuros cuando se apliquen con mayor extensión las herramientas.

1. Introducción

La Ingeniería de Requisitos es responsable por la definición de los servicios que el software deberá atender. Para atacar esta etapa inicial del desarrollo de un sistema de software, los ingenieros de requisitos desarrollan modelos, que registran el conocimiento adquirido en las diferentes etapas del proceso de definición de requisitos y las decisiones ya tomadas acerca de las características del software a construir.

Uno de los problemas que se enfrentan es la incompletitud. Esta ha sido ampliamente estudiada en la literatura, y diferentes autores coinciden que la incompletitud es uno de los principales problemas que afronta la Ingeniería de Requisitos [1-3]. Además la

completitud ha sido considerada en variados estándares, por ejemplo el estándar IEEE 830-1998 [4] la considera una de las 8 propiedades que debe cumplir una buena especificación de requisitos de software. A su vez el estándar IEEE 29148 [5], presenta la completitud como una propiedad fundamental que debe cumplir cada requisito individualmente y el conjunto de requisitos de software como un todo.

El presente estudio se centra en los modelos desarrollados por Leite et al. [6-10], que son: el Léxico Extendido del Lenguaje (LEL) y los Escenarios. Sobre esos modelos, varios autores han atacado la completitud de los mismos [11-20].

Diferentes estudios se han efectuado a fin de evaluar la completitud de los modelos, desde los primeros trabajos realizados por Doorn y Ridao en 2003 [11]. Posteriormente se efectuaron mejoras semánticas en las estrategias de estimación de la completitud a fin de obtener una mayor verosimilitud de los resultados. Luego se realizaron construcciones de los modelos con un mayor nivel de detalle, usando heurísticas más refinadas [15]. En todos los casos el nivel de completitud tuvo una alta coincidencia con estimaciones anteriores, donde se habían aplicado enfoques basados en aspectos sintácticos [15].

Las conclusiones a las que se arribó es este punto fueron que:

- i) Las estimaciones de completitud parecen ser confiables.
- ii) Las mejoras en las heurísticas resultaron poco relevantes, por lo que los abordajes utilizados no lograron una mejora en la eficacia en la construcción de los modelos.

A principios de 2016, Litvak et al. [21], proponen utilizar estrategias de Procesamiento de Lenguaje Natural (PNL), con el fin de mejorar el tratamiento de la completitud de los modelos.

En el presente artículo se describen las estrategias utilizadas para incorporar PNL al proceso de requisitos,

así como algunos resultados preliminares muy promisorios. Dichos estudios preliminares avizoran una importante mejora en la calidad de los modelos y han sugerido varios trabajos futuros. El núcleo de los trabajos futuros está orientado a crear heurísticas que aprovechen las herramientas de construcción de manera de mejorar la percepción del ingeniero de requisitos del problema que está siendo abordado.

2. Completitud de Modelos de Requisitos escritos en Lenguaje Natural

Es apropiado insistir que para mejorar la completitud de un documento escrito en lenguaje natural es necesario disponer de mecanismos para estimarlas. Esto ocurre escasamente, ya que en la mayoría de los casos, incluyendo las normas mencionadas, todo se reduce a expresiones de deseo en las que se recomienda construir modelos lo más completos posibles, sin brindar pista alguna acerca de cómo hacerlo o cómo medirlo.

Si bien se han realizado distintas propuestas para atacar la incompletitud de los modelos LEL y Escenarios, los resultados obtenidos son aún poco satisfactorios desde el punto de vista de la completitud obtenida, pero razonablemente válidos en lo que se refiere a la medición de la misma.

Los primeros trabajos realizados por Doorn y Ridao en 2003 [11] y en 2006 [12] aplicaron el método de captura - recaptura [22] para estimar la completitud. El método de captura - recaptura fue utilizado inicialmente en otras áreas de la ciencia, como el área de la Biología, Medicina y la Ecología. Posteriormente se aplicó en el campo de la Ingeniería de Software para estimar el número total de errores presentes en un artefacto de software, a partir de los resultados de un proceso de inspección sobre dicho artefacto [23-25]. El método usado por Doorn y Ridao fue una adaptación del método de captura - recaptura, llamado Detection Profile Method [26], utilizado para predecir el número de elementos faltantes en el LEL y en escenarios, donde concluyeron que al aumentar el número de ingenieros de requisitos la cantidad de elementos estimados se acercaría significativamente a la estimación realmente existente en el dominio del problema. El trabajo se realizó sobre 9 muestras correspondientes a un mismo caso de estudio, realizadas por ingenieros de requisitos independientes.

Los estudios posteriores, mencionados en la introducción, consistieron en procurar mejorar la estimación de completitud incorporando correcciones semánticas. El primero de ellos, presentado por Litvak et al. [13], incorpora previamente correcciones semánticas y luego aplica el método estadístico Detection Profile Method [26]. Dichas correcciones semánticas corresponden al estudio de las denotaciones y connotaciones de cada símbolo del LEL, para establecer

efectivamente los sinónimos entre los símbolos existentes y si los símbolos realmente eran importantes en el contexto. En el segundo estudio se realizaron correcciones semánticas más refinadas, estudiando la relevancia, pertenencia, sinonimia y homonimia del contenido textual del modelo; luego se compararon con los resultados de los trabajos precedentes, concluyéndose que los datos disponibles anteriormente acerca de la completitud eran esencialmente correctos. Lamentablemente la completitud de cada muestra individual seguía siendo baja.

En un intento de mejorar el grado de completitud logrado hasta el momento se desarrollaron nuevos léxicos del mismo caso de estudio, con un proceso de construcción más cuidadoso. Teniendo en ese momento 12 muestras de un mismo caso de estudio, generadas independientemente, en un nuevo trabajo [15], se comprobó que la completitud individual seguía en el mejor caso en el orden del 50%.

Todos estos estudios muestran que la replicación de un modelo por parte de ingenieros de requisitos independientes:

- i) Es una técnica muy valiosa para el estudio de la completitud.
- ii) Mejora notoriamente la completitud si se reúne la información de todas las repeticiones.
- iii) Provee datos significativos para estimar apropiadamente la completitud.
- iv) Es poco viable en desarrollos prácticos; ya que duplica, triplica o eventualmente cuadruplica los costos involucrados.

Es por esto que es imperativo abordar el problema de mejorar la completitud de los modelos sin necesidad de recurrir a repeticiones. Es este el objetivo central que ha motorizado todos los estudios de completitud realizados hasta la fecha.

3. Aportes del Procesamiento de Lenguaje Natural a la construcción de Modelos de Requisitos

Recién en el proyecto presentado en 2016 por Litvak et al. [21] aparece la propuesta innovadora de la incorporación de estrategias de Procesamiento de Lenguaje Natural por parte de los ingenieros de requisitos.

El Procesamiento del Lenguaje Natural (PNL o NLP en inglés) es una disciplina con una larga trayectoria. Nace en la década de 1960, como un subárea de la Inteligencia Artificial y la Lingüística, con el objeto de estudiar los problemas derivados de la generación y comprensión automática del lenguaje natural [27-28]. En los últimos años, las contribuciones que se han hecho

desde este dominio han crecido sustancialmente, permitiendo el procesamiento de ingentes cantidades de información, e incrementando su eficacia. Muestra de ello es la aplicación de estas técnicas como una componente esencial en los motores de búsqueda web, en las herramientas de traducción automática, o en la generación automática de resúmenes [29].

Las herramientas de Procesamiento de Lenguaje Natural permiten el reconocimiento de idiomas, la segmentación de oraciones, palabras y secciones, incluyendo en algunos casos un análisis morfológico, sintáctico y/o semántico [30]. Dentro del análisis morfológico se realiza la detección de palabras especiales, por ejemplo detección de nombres propios, fechas, distancias; la asignación de etiquetas para determinar para cada palabra su categoría gramatical correcta; y la desambiguación, que realiza la tarea de asignar el sentido correcto a las palabras, buscando la interpretación más adecuada en base al contexto. El análisis sintáctico permite hallar los constituyentes básicos que suelen llamarse en inglés “chunks”, como definir que el término “come” pertenece al Grupo Verbal o “pescado” al Grupo Nominal; permite además definir la estructura de la oración; establecer las funciones gramaticales y roles temáticos (como la división en sujeto, objeto y predicado). Luego el análisis semántico puede ser análisis léxico o proposicional [30].

3.1. Antecedentes de aplicación de PNL

Como se detalla en la sección 5, en base a las características y propiedades esperadas, se seleccionó la herramienta FreeLing [31] entre un conjunto de herramientas evaluadas.

Esta herramienta ha sido utilizada en varios proyectos industriales, como Ruby Reader, que es una aplicación para el iPhone que ayuda a los hablantes de japonés a comprender textos en inglés. Ha sido desarrollado en Tokio por la empresa CA-Mobile [32].

También se ha aplicado FreeLing en Vi-Clone, con el desarrollo de asistentes virtuales para páginas web corporativas. Algunos componentes de FreeLing se están integrando en el sistema de diálogo. Vi-Clone está financiando el desarrollo del módulo de corrección ortográfica que permitirá a FreeLing procesar frases del usuario escritas en variantes no estándar [33].

Otro producto que utiliza FreeLing es TextToSign, Traductor de texto en español al lenguaje de señas, que utiliza FreeLing para el procesamiento de texto [34].

Se emplea también en Dixio, que es un diccionario inteligente capaz de ayudar al lector de un texto, ofreciendo definiciones contextualizadas. Ha sido desarrollado por Semantix [35].

Además se aplica en Aport News, que es un portal de noticias que usa FreeLing como un preprocesador para

enriquecer texto en ruso. El resultado de la anotación se usa en la clasificación y agrupación de las noticias [36].

FreeLing se utilizó además en un software realizado para estudiar la opinión sobre los candidatos a presidente en Twitter en el año 2013 en Chile. En este caso se estudiaron dos técnicas desarrolladas para obtener el sentimiento asociado a un texto: un Método Estadístico y un Método Ontológico. Se eligió el Método Ontológico, para el que se construyen manualmente reglas, a fin de identificar el sentimiento, procesando texto libre con la librería FreeLing. Este trabajo fue desarrollado en la Universidad de Chile [37].

También la Universidad de Sonora, en diciembre de 2012, usó FreeLing para crear una herramienta que extrae conocimiento de prácticamente cualquier dominio en forma automática, principalmente de textos escritos en lenguaje natural en español [38].

Además se aplicó en la realización de la Primera aproximación para la extracción automática de Entidades Nombradas en corpus de documentos medievales castellanos, mediante FreeLing para español estándar y la variante diacrónica del español de los siglos XII al XVI [39] en un corpus formado por los documentos que componen el Libro Becerro de las Behetrías de Castilla del siglo XIV [40]. También se ha desarrollado un corpus de entrenamiento del etiquetador, y se ha usado la herramienta resultante en un estudio lingüístico sobre la evolución del uso del verbo haber [41-42].

A modo de resumen de estos trabajos mencionados anteriormente, se puede decir que todos comparten un núcleo común relacionado con la gestión del conocimiento; en algunos casos para su elicitación, en otros para su formalización y finalmente en otros para lograr un efecto pragmático.

Notoriamente gran parte de la Ingeniería de Requisitos consiste en extraer conocimiento, formalizarlo o usarlo en el diseño de los servicios que debe prestar un sistema de software. Es por eso que la inserción de estas estrategias en la Ingeniería de Requisitos ha resultado tan inmediata como eficaz.

3.2. Servicios necesarios de Procesamiento de Lenguaje Natural

Sólo es necesaria una mirada superficial al conjunto de herramientas de PNL para advertir la existencia de importantes diferencias en los servicios que prestan y en las propiedades de los mismos. Es así que se analizó a priori cuáles servicios son necesarios:

- Segmentar oraciones: Reconocimiento del punto al terminar una oración. Segmentación del texto en oraciones. Realiza la preparación y segmentación de la oración en el texto.
- Segmentar palabras: Segmentación del texto en palabras. Identifica las palabras en cada oración,

también se detectan números y signos de puntuación.

- Etiquetar con etiquetas POS: POS proviene de etiquetado de la oración (en inglés, "Part of speech"). Realiza el proceso de asignar a cada palabra su categoría gramatical correcta. Por ejemplo, para la frase: "El petróleo es un combustible", El (artículo masculino singular), petróleo (sustantivo masculino singular), es (verbo presente indicativo), un (determinante), combustible (sustantivo masculino singular).
 - Realizar análisis morfológico: Realiza tareas que determinan la forma, clase o categoría gramatical de cada palabra de una oración. Ejemplo de análisis morfológico: Se analiza la frase: "La niña llevaba sombrero rojo", obteniendo: La: artículo, femenino, singular; niña: sustantivo común, femenino, singular; llevaba: verbo llevar, tercera persona del singular del pretérito imperfecto del indicativo, primera conjugación; sombrero: sustantivo común, masculino, singular; rojo: adjetivo calificativo, masculino, singular.
 - Utilizar estimadores: Realizan la estimación de parámetros que establece el modelo estocástico del lenguaje a partir de un gran corpus de secuencias de palabras y de sus correspondientes secuencias de etiquetas semánticas. Un modelo estocástico de lenguaje es un modelo matemático en el cual se ha definido una función de probabilidad que calcula la probabilidad de ocurrencia de una frase en un lenguaje dado.
 - Permitir la desambiguación: Realiza la tarea de asignar el sentido correcto a las palabras. Selecciona la interpretación más adecuada teniendo en cuenta el contexto. Considerando el siguiente ejemplo: "Yo bajo con el hombre bajo a tocar el bajo bajo la escalera". Se presentan cuatro categorías diferentes: verbo, adjetivo, sustantivo y preposición. El desambiguador devolverá todas ellas para cada una de las apariciones de la forma "bajo" en la oración.
- Subsidiariamente se esperaba de la herramienta:
- Ser del dominio público de manera que las herramientas que se construyan puedan ser transferidas al medio productivo sin limitaciones.
 - Tener código abierto: Para facilitar la observación del funcionamiento interno de la herramienta.
 - Brindar procesamiento multilingüe automático: Dado que este proyecto de requisitos tiene trabajos interrelacionados con otros países del

mundo, como Brasil especialmente, y con la posibilidad de ampliar a idioma inglés.

- Permitir utilización local: Preferentemente poder utilizar PNL sin necesidad de conexión a Internet continua, para no depender del servidor del proveedor del servicio, y por lo tanto obtener una mayor disponibilidad.
- Aplicar diccionarios accesibles y modificables: La observación de los diccionarios y la posibilidad de modificarlos facilita el funcionamiento y permite ampliar la terminología a ser aplicada.
- Facilitar la integración con otros lenguajes de programación.
- Poder ser utilizada en diversas plataformas.
- Existencia de actualizaciones mediante nuevas versiones.

4. Comparación de los servicios que brindan diferentes herramientas de PNL

En la Tabla 1 se presentan un subconjunto de las herramientas más representativas analizadas. Se ha evaluado en total un conjunto de 10 herramientas, presentando cada una de ellas un conjunto de servicios diferentes.

Tabla 1. Comparación de servicios de herramientas de PNL

Función	Segmentación de oraciones	Segmentación de palabras	Etiquetado POS	Análisis morfológico	Estimadores	Desambiguación
FreeLing	Sí	Sí	Sí	Sí	Sí	Sí
NLTK	Sí	Sí	Sí	Sí	Sí	Sí
AS*	Sí	Sí	Sí	Sí	Sí	Sí
Grampal	Sí	Sí	Sí	Sí	No	No
WordNet	Sí	Sí	No	No	No	Sí
FrameNet español	Sí	Sí	Sí	No	No	No
Snowball	No	Sí	No	Sí	Sí	Sí
Apache OpenNLP	Sí	Sí	Sí	No	No	Sí
TreeTagger	No	Sí	Sí	No	No	Sí
SVMTool	Sí	Sí	Sí	No	No	No

*AS es el Analizador sintáctico on-line

En un análisis preliminar se observó que a las herramientas Grampal [43], WordNet [44], FrameNet español [45], Snowball [46], Apache OpenNLP [47],

TreeTagger [48] y SVMTool [49] les faltaban algunos de los servicios necesarios.

De entre las tres que cumplen todas las condiciones, FreeLing [31], NLTK [50] y AS [51] se seleccionó el FreeLing por los siguientes motivos.

Respecto de AS se consideró necesario descartarla dado que no tiene código abierto ni es multi-idioma. Se plantea a futuro trabajar con este mismo software en inglés y quizás en portugués.

Si bien NLTK resulta aceptablemente eficiente para procesar texto, en idioma español se pudo observar que no presenta un buen funcionamiento para la

identificación y clasificación de entidades nombradas (nombres de organizaciones, ciudades, personas, etc.), mientras que en la herramienta FreeLing, esta funcionalidad se realiza perfectamente, tanto para el español como para otros idiomas.

A continuación se muestra un ejemplo de prueba realizado en dichas herramientas. Se ingresó la siguiente oración en ambas herramientas: "Buenos Aires es una hermosa ciudad en Argentina".

El resultado obtenido en NLTK [50] se puede observar en la Figura 1.



Figura 1. Ejemplo de falta de reconocimiento de entidades nombradas en NLTK

El resultado obtenido con la herramienta FreeLing [52], se presenta en la Figura 2, con la salida obtenida al ingresar la misma frase y marcar las opciones Named Entity detection y Named Entity classification (detección y clasificación de Entidades Nombradas), seleccionando auto-detección de idiomas y PoS Tagging (etiquetado PoS).

Se puede observar entonces que en la herramienta NLTK [53] la detección de entidades nombradas no se realiza de manera correcta, y por lo tanto su clasificación tampoco resulta satisfactoria.

En la frase "Buenos Aires es una hermosa ciudad en Argentina" el NLTK a "Buenos Aires" lo toma por separado como dos sustantivos comunes, siendo:

- Buenos: ncmp000, donde: n: sustantivo c: común m: masculino p: plural.

- Aires: ncmp000, donde: n: sustantivo c: común m: masculino p: plural.

En cambio FreeLing si reconoce a "Buenos Aires" como un sustantivo propio, o sea que lo reconoce como una entidad nombrada. Es decir que la herramienta FreeLing, al evaluar texto en español, detecta y clasifica las entidades nombradas de manera correcta.

Lo mismo se repite en muchos otros casos probados, tanto en FreeLink como en NLTK, con frases similares, como por ejemplo "La Sagrada Familia es la mayor representación de arte de Barcelona.", donde NLTK no reconoce la Sagrada Familia como entidad nombrada. También en la frase "Mamel Antonio es la ciudad más bonita de Costa Rica", donde NLTK no reconoce la ciudad de Manuel Antonio ni a Costa Rica como

entidades nombradas. Ambas entidades nombradas son perfectamente reconocidas en FreeLing.

Se procedió al estudio de las propiedades y los servicios deseados, en base a lo planteado en la Sección 3.2. Inicialmente se evaluó el origen de la herramienta FreeLing, desarrollada en TALP [54], grupo encabezado por Lluís Padró [55-56]. Se trata de un proyecto financiado parcialmente por el Gobierno Español, la Unión Europea y la empresa Vi-Clone. A principios de mayo de 2016 se ha presentado la versión 4.0 del FreeLing, manteniendo la compatibilidad y disponibilidad de las versiones anteriores. El proyecto se estructura como una librería que puede ser llamada desde cualquier aplicación de usuario que requiera servicios de análisis del lenguaje [55]. El software se distribuye como código abierto bajo una licencia GNU General Public License 3 y bajo licencia dual a empresas que deseen incluirlo en sus productos comerciales [57]. Ser de código abierto, permitió incorporar el código de otros

proyectos similares, como el módulo de desambiguación del sentido de las palabras basado en UKB [58]. Soporta las lenguas asturiano, catalán, castellano, galés, gallego, inglés, italiano, portugués, y ruso, variando los servicios de análisis de cada lengua [55].

Permite construir diccionarios en otros idiomas como se muestra en la utilización del Checo en “Usando FreeLing en Checo” [59]. Además soporta el formato de codificación de caracteres en Unicode (UTF-8), servicio útil para su posterior aplicación.

Respecto a las características y servicios específicos en el Procesamiento de Lenguaje Natural permite la segmentación de oraciones, la segmentación de palabras, etiquetado POS, el análisis morfológico, uso de estimadores, la desambiguación de texto, y el soporte multi-idioma.

Analysis Results						
▼ Language identification						
Identified language is: Spanish (es)						
▼ Sentences						
Sentence 1						
Buenos_Airea	es	una	hermosa	ciudad	en	Argentina
buenos_aires	scr	uno	hermosa	ciudad	cn	argentina
NP00000	VB00000	DT00000	AD00000	NC00000	SP	NP00000

Figura 2. Ejemplo de reconocimiento de entidades nombradas en FreeLing

5. Resultados preliminares de aplicación de PNL a Ingeniería de Requisitos

Se ha realizado una primera aplicación de PNL para intentar obtener símbolos del Léxico Extendido del Lenguaje a partir de documentos escritos en Lenguaje Natural. Una de las heurísticas de construcción del Léxico Extendido del Lenguaje indica clasificar los elementos obtenidos en Sujeto, Objeto, Verbo o Estado.

La herramienta ha sido planteada de modo interactivo, con el propósito de lograr dirigir la atención del ingeniero de requisitos hacia la mayor cantidad posible de aspectos potencialmente útiles del texto procesado. En otras palabras, se intenta fuertemente evitar que el constructor del LEL omita analizar semánticamente parte del documento.

La diferencia entre el uso cotidiano del vocabulario en el dominio de aplicación y los conceptos que los ingenieros de requisitos van aprendiendo sobre un macrosistema que no se conoce profundamente, puede crear gaps semánticos entre las partes involucradas (ingenieros de requisitos y clientes-usuarios). Como consecuencia de dichos gaps semánticos, es muy posible que algunos términos sean descartados en forma prematura, al ser evaluados superficialmente.

En el proceso de aplicación de PNL a Ingeniería de Requisitos, se decidió definir una herramienta que permitiera, mediante PNL, proponer posibles Sujetos, Objetos, Verbos o Estados, a través de la aplicación de FreeLing a documentos escritos en Lenguaje Natural. El mismo puede ser una entrevista escrita, un manual de usuario, un manual técnico u otro documento de relevancia del contexto.

La herramienta construida usando FreeLing incentiva un análisis más uniforme de los términos detectados por la misma, lo que redundará en una reducción de las omisiones.

Se trabajó a partir de una transcripción literal de una entrevista realizada a clientes/usuarios sobre un caso de estudio específico. Se han comparado los resultados obtenidos por ingenieros de requisitos aplicando PNL con los resultados de ingenieros de requisitos sin utilizarlo. Los resultados preliminares se muestran en la Tabla 2.

Tabla 2. Resultados preliminares

	IngR sin usar PNL	IngR usando PNL	No vistos por IngR sin usar PNL	No vistos por IngR usando PNL
Sujetos	5	6	0	1
Objetos	24	29	9	4
Verbos	8	19	13	2
Estados	9	9	7	7

*IngR es ingenieros de requisitos

En el caso de los Estados se observa una diferencia notable entre los símbolos obtenidos de una u otra manera. Por ejemplo el símbolo de tipo Estado denominado "Insumo en stock" lo infieren los ingenieros de requisitos cuando se habla de insumos y luego del stock de los mismos.

No está en oraciones cercanas, por lo que la herramienta no lo detecta. Por otro lado los ingenieros de requisitos no han visto, por ejemplo, el símbolo de tipo Estado "Producto repintado", que si ha sido propuesto por la herramienta.

Con respecto a los verbos la aplicación de FreeLing permitió detectar más verbos.

Tabla 3. Porcentajes obtenidos aplicando FreeLing

	Total de símbolos	Porcentaje de símbolos no vistos por los ingenieros de requisitos
Sujetos	6	17 %
Objetos	33	27 %
Verbos	21	62 %
Estados	18	44 %
Total de símbolos	78	39 %

Los resultados son sumamente promisorios dado que la cantidad de símbolos hallados utilizando PNL es mucho mayor a la obtenida por los ingenieros de requisitos que no lo utilizaron. Los mismos se observan en la Tabla 3.

Lograr obtener un 39% de símbolos que no han sido vistos por los ingenieros de requisitos, los cuales pudieron ser sólo detectados aplicando FreeLing y los conceptos de PNL es una mejora sustancial de la calidad del modelo.

6. Conclusiones

Luego del trabajo realizado, y como consecuencia de los estudios preliminares, se avizoran los beneficios de aplicación de PNL en la generación de modelos de Requisitos escritos en Lenguaje Natural.

Haber aplicado FreeLing a entrevistas reales realizadas a los usuarios de un contexto dado, ha permitido obtener una amplia mejora de la completitud.

Esto es, por un lado una confirmación del grado de incompletitud de los modelos creados con las heurísticas actuales sin usar PNL, pero también es una evidencia de la herramienta que fue utilizada.

Como trabajo futuro se propone la finalización de la herramienta de generación de modelos de requisitos en lenguaje natural, aplicando FreeLing, no sólo en la búsqueda de símbolos del LEL sino en los modelos posteriores, como el modelo de Escenarios y la especificación de requisitos misma. Además se espera realizar una profundización de los estudios de completitud al aplicarla en diferentes contextos, evaluando también la completitud de los modelos posteriores del proceso.

Por otro lado se planea perfeccionar aún más las heurísticas de construcción de los modelos, de modo de generar modelos directamente más completos, con heurísticas refinadas.

La conjunción de mejores heurísticas, con el aprovechamiento de PNL, avizora grandes mejoras en la completitud de los modelos y por ende, en la calidad de los mismos.

7. Referencias

- [1] Kotonya G., y Sommerville I., *Requirements engineering – processes and techniques*, John Wiley & Sons UK, 1998.
- [2] Loucopoulos P., y Karkostas V., *System requirements engineering*, McGraw-Hill, Book company Europe, 1995.
- [3] Firesmith D.G., "Are Your Requirements Complete?", *Journal of Object Technology*, Vol. 4, N° 1, January-February 2005, pp. 27-43, http://www.jot.fm/issues/issue_2005_01/column3.

- [4] IEEE Computer Society, *Software Engineering Standards Committee, & IEEE-SA Standards Board, IEEE recommended practice for software requirements specifications*, Institute of Electrical and Electronics Engineers, 1998.
- [5] IEEE *Systems and software engineering - Life cycle processes - Requirements engineering*, IEEE, Nueva York, 2011.
- [6] Leite, J.C.S.P., y Oliveira, A.P.A., "A Client Oriented Requirements Baseline", *Second IEEE International Symposium on Requirements Engineering*, IEEE Computer Society Press, 1995, pp. 108-115.
- [7] Leite, J.C.S.P., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G., Hadad, G., y Oliveros, A., "Enhancing a Requirements Baseline with Scenarios", *Requirements Engineering Journal*, Springer-Verlag London Ltd., Vol. 2, N° 4, 1997, pp. 184-198.
- [8] Leite J.C.S.P., Hadad G.D.S., Doorn J.H., y Kaplan G.N., "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>
- [9] 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, 2001, pp. 238-246.
- [10] Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., y Ridaio, M.N., "Defining System Context using Scenarios", en el libro "Perspectives on Software Requirements", Kluwer Academic Publishers, EEUU, ISBN: 1-4020-7625-8, Cap. 8, 2004, pp. 169-199.
- [11] Doorn, J.H., y Ridaio, M., "Complejidad de Glosarios: Un Estudio Experimental", en *Proceedings: 6th Workshop on Requirements Engineering Paracibaba, Universidade Metodista de Piracicaba*, Brasil, 2003, pp. 317-328.
- [12] Ridaio, M.N., y Doorn, J.H., "Estimación de Complejidad en Modelos de Requisitos Basados en Lenguaje Natural", en *Proceedings: 9th Workshop on Requirements Engineering*, Rio de Janeiro, Brasil: Pontificia Universidade Católica do Rio de Janeiro, 2006, pp. 151-158.
- [13] Litvak C.S., Hadad G.D.S., y Doorn J.H., "Un abordaje al problema de completitud en requisitos de software", en *anales de CACIC 2012 - XVIII Congreso Argentino de Ciencias de la Computación*, Universidad Nacional del Sur, Bahía Blanca, ISBN: 978-987-1648-34-4, Octubre 2012, pp. 827-836, <http://hdl.handle.net/10915/23715>.
- [14] Litvak, C.S., Hadad, G.D.S., y Doorn, J.H., "Correcciones semánticas en métodos de estimación de completitud de modelos en lenguaje natural", en *Proceedings 16th Workshop on Requirements Engineering*, ORT University, Montevideo, Uruguay, 2013, pp. 105-117.
- [15] Litvak, C.S., Hadad, G.D.S., y Doorn, J.H., "Mejoras semánticas para estimar la completitud de modelos en lenguaje natural", en *anales de CoNaISI 2013, 1er Congreso Nacional en Ingeniería Informática / Sistemas de Información*, Universidad Tecnológica Nacional, Córdoba, Argentina, 2013.
- [16] Litvak C.S., Hadad G.D.S., y Doorn J.H., "Heurísticas para el modelado de requisitos escritos en lenguaje natural", en *anales de 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>.
- [17] Litvak C.S., Hadad G.D.S., y Doorn J.H., "Influencia de las nominalizaciones sobre la completitud de modelos de requisitos", en *anales de WICC 2015 - XVII Workshop de Investigadores en Ciencias de la Computación*, anales electrónicos, Universidad Nacional de Salta, Salta, ISBN: 978-987-633-134-0, 2015.
- [18] Hadad G.D.S., Litvak C.S., y Doorn J.H., "Agregando semántica a técnicas de predicción de completitud en modelos de requisitos", en *anales de WICC 2012 - XIV Workshop de Investigadores en Ciencias de la Computación*, anales electrónicos, Universidad Nacional de Misiones, Posadas, ISBN: 978-950-766-082-5, 2012, pp. 392-396, <http://hdl.handle.net/10915/18863>.
- [19] Hadad, G.D.S., Litvak, C.S., Doorn, J.H., y Ridaio M.N., "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, 3ª edición, cap. 279, ISBN13: 9781466658882, Julio 2014 (10384 p.), pp. 2854-2863.
- [20] Hadad G.D.S., Litvak C.S., y Doorn J.H., "Problemas y Soluciones en la Completitud de Modelos en Lenguaje Natural", en *anales de CADI 2014 - II Congreso Argentino de Ingeniería*, Universidad Nacional de Tucumán, San Miguel de Tucumán, ISBN: 978-987-1662-51-7, T366, Septiembre 2014, artículo destacado mediante la aceptación para su exposición oral, además de su publicación en los anales electrónicos. <https://cadi.org.ar/index.php/trabajos-seleccionados/16-tecnologia-de-la-informacion-y-comunicacion>.
- [21] Litvak C.S., Hadad G.D.S., y Doorn J.H., "Procesamiento de Lenguaje Natural para Estimar Completitud de Requisitos", en *anales de WICC 2016 - XVIII Workshop de Investigadores en Ciencias de la Computación*, anales electrónicos, Universidad Nacional de Entre Ríos, Entre Ríos, ISBN: 978-950-698-377-2, 2016.
- [22] Otis, D.L., Burnham, K.P., White, G.C., y Anderson, D.R., *Statistical inference from Capture on Closed Animal Populations*, Wildlife Monograph, 62, 1978.
- [23] Petersson, H., Thelin, T., Rimeson, P., y Wohlin, C., "Capture-recapture in software inspections after 10 years research—theory, evaluation and application", *The Journal of Systems and Software*, 72, 2003, pp. 249-264.
- [24] Kamel, A., y Sorenson, P., "The Application of CaptureRecapture Log-Linear Models To Software Inspections Data", en *Proceedings: ISESE '03 - International Symposium on Empirical Software Engineering*, 2003, pp. 213-222.

- [25] Thelin, T., "Team-based Fault Content Estimation in the Software Inspection Process", en *Proceedings ICSE '04 - 26th International Conference on Software Engineering*, 2004, pp. 263-272.
- [26] Wohlin, C., y Runeson, P., "Defect content estimations from Review Data", en *Proceedings 20th Intl Conference on Software Engineering*, Japón, 1998, pp. 400-409.
- [27] Méndez, E., y Moreiro González, J. A., "Lenguaje natural e indexación automatizada", *Ciencias de la Información*, 30(3), 1999, pp. 11-24.
- [28] Vilares Ferro, J., *Aplicaciones del procesamiento del lenguaje natural en la recuperación de información en español*, Universidad de Coruña, Departamento de Computación, Tesis doctoral, Mayo de 2005, <http://colegweb.dc.fi.udc.es/cole/library/ps/Vil2005a.pdf>, revisado el 4/7/16.
- [28] Vallez M., y Pedraza-Jimenez R., "El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines", [en línea], *Hipernext.net*, núm. 5, 2007, <https://www.upf.edu/hipertextnet/numero-5/pln.html>.
- [30] Alonso i Alemany, L., "Herramientas Libres para Procesamiento del Lenguaje Natural", en *Stas Jornadas Regionales de Software Libre*, Rosario, Argentina. Noviembre de 2005, <http://www.cs.famaf.unc.edu.ar/~luar/freeNLP>, revisado el 18/5/16.
- [31] FreeLing, <http://nlp.lsi.upc.edu/FreeLing-old/>, visited on 10/5/16.
- [32] CA-Mobile, <http://www.causobile.com>, revisado el 14/5/16.
- [33] Vi-Clone, <http://www.vi-clone.com>, revisado el 14/5/16.
- [34] TextToSign, <http://www.textosign.es>, revisado el 14/6/16.
- [35] Stana I., *Using the FreeLing library with Czech Bachelor's Thesis*, Masarykova univerzita, Fakulta informatiky, Bachelor's Thesis, 2013, http://is.muni.cz/th/256650/fi_bthesis.pdf, revisado el 28/7/16.
- [36] Aport News, <http://news.afort.ru>, revisado el 14/5/16.
- [37] Universidad de Chile- Facultad de Ciencias Físicas y Matemáticas- Departamento de Ciencias de la Computación. Presentaron el estudio Herramientas de Análisis de Opinión en Redes Sociales Virtuales en Marzo de 2014. http://repositorio.uchile.cl/bitstream/handle/2250/116987/cf-nkouchichai_i.pdf?sequence=1, revisado el 14/5/16.
- [38] Universidad de Sonora, Divisiones de Ingeniería, Ciencias Exactas y Naturales y Ciencias Biológicas y de la Salud, publicado en Diciembre de 2012, <http://www.enjstemas.uson.mx/revistas/pdf/numero13.pdf>, revisado el 14/5/16.
- [39] Sánchez-Marco, C., Gemma B., y Padró L., "Extending the tool, or how to annotate historical language varieties", en *Proceedings 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*.
- [40] *Janus*, Anevo 1, 2014, ISSN 2254-7290, <http://www.iamsdizital.es/ia/2427888079544686815390487/hd/b2013baia.pdf?sessionid=E80CB2D31139B6A7A9C4CA391738FA50>, revisado el 14/5/16.
- [41] Sánchez-Marco, C. y Stefan E., "Measuring semantic change: The case of spanish participial constructions", en *Proceedings 4th Conference on Quantitative Investigations in Theoretical Linguistics (QITL-4)*, Berlin, Germany, March 2011, pp. 79-83.
- [42] Sánchez-Marco, C., *Tracing the development of Spanish participial constructions: An empirical study of language change*, Universitat Pompeu Fabra, Barcelona, Spain, Ph.D. tesis, 2012.
- [43] Grampal, <http://cartago.lluf.uam.es/grampal/grampal.cgi>, revisado el 20/6/16.
- [44] WordNet, <http://grial.uab.es/synset/synset2.php>, revisado el 20/6/16.
- [45] FrameNet en español, <http://spanishfn.org/>, revisado el 20/6/16.
- [46] Snowball, <http://snowballstem.org/>, revisado el 20/6/16.
- [47] Apache OpenNLP, <http://opennlp.apache.org/>, revisado el 20/6/16.
- [48] TreeTagger, <http://chasen.org/~taku/software/vamcha/>, revisado el 20/6/16.
- [49] SVMTool, <http://alias-i.com/lingpipe/web/demo-ne.html>, revisado el 20/6/16.
- [50] NLTK, <http://www.nltk.org/>, revisado el 25/7/16.
- [51] Analizador sintáctico on line desarrollado por el Grupo de estructuras de datos y lingüística computacional, <http://www.zedlc.ulpgc.es>, revisado el 20/6/16.
- [52] Demo del FreeLing, <http://nlp.lsi.upc.edu/freeling/demo/demo.php>, revisado el 13/7/16.
- [53] NLTK demo, <http://text-processing.com/demo/tag/>, revisado el 25/7/16.
- [54] <http://www.talp.upc.edu/>, revisado el 1/7/16.
- [55] Padró, L., *Analizadores multilingües en FreeLing*, *Linguamática*, 3, 2011, pp. 13-20.
- [56] Padró, L. y Stanilovsky, E., "FreeLing 3.0: Towards Wider Multilinguality", en *Proceedings Language Resources and Evaluation Conference (LREC 2012)* ELRA, Estambul, European Language Resources Association (ELRA), 2012, pp. 2473-2479.
- [57] Licencias GNU, <http://www.gnu.org/copyleft/gpl.html>, revisado el 20/5/16.
- [58] Agirre, E., y Soroa A., "Personalizing pagerank for word sense disambiguation", en *Proceedings 12th Conference of the European chapter of the Association for Computational Linguistics (EACL-2009)*, Athens, Greece, 2009, pp. 33-41.

[59] Stana I., *Using the FreeLing library with Czech* Bachelor's Thesis, Masarykova univerzita, Fakulta informatiky, Bachelor's Thesis, 2013, http://is.muni.cz/th/256650/fi_b/thesis.pdf .
revisado el 28/7/16.