

Investigación en Progreso: Propuesta de Proceso para Ingeniería de Explotación de Información Centrado en Métodos y Técnicas: Núcleo IMMAT

Ezequiel Baldizzoni^{1,2}

1. Programa de Maestría en Ingeniería de Sistemas de Información.

Escuela de Posgrado, Facultad Regional de Buenos Aires. Universidad Tecnológica Nacional. Argentina.

2. Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información (LIDIEI).

Grupo de Investigación en Sistemas de Información (GISI). Universidad Nacional de Lanús. Argentina.
ebaldizzoni@gmail.com

Resumen— Debido a que las organizaciones generan grandes cantidades de información, el éxito de los negocios depende de la habilidad para entender las nuevas tendencias o cambios ocultos en sus datos. Gracias a la ingeniería de explotación de información se pueden identificar estas tendencias y cambios. Con el fin de mejorar este proceso se intenta adaptar a esta rama de la ingeniería los métodos y técnicas propuestas por Semat para la ingeniería de software partiendo de los procesos existentes.

Palabras clave—Explotación de Información, Modelo de Procesos, Herramientas y técnicas

I. JUSTIFICACIÓN

Las organizaciones generan grandes cantidades de información sobre sus procesos productivos, desempeño operacional, mercados y clientes, sin embargo, éxito de los negocios depende de la habilidad de parte de los gestores para entender las nuevas tendencias o cambios ocultos en sus datos. A esto se le suma la gran cantidad de información que produce hoy en día Internet lo que hace que el trabajo de extracción, procesamiento o explotación de dicha información resulte cada vez más arduo. La utilización de la ingeniería de explotación de información se orienta a identificar tendencias, comportamientos y conocimiento de la organización y su entorno [2-13].

Hoy en día existen propuestas de procesos recomendados para llevar adelante proyectos basados en la ingeniería de explotación de información ([8-5-23]) donde se detallan cuáles son los pasos que hay que realizar, las actividades a desarrollar en cada paso e incluso técnicas recomendadas para que esta tarea tan costosa retorne los resultados necesarios con el fin de que el negocio crezca.

Las investigaciones en el campo disciplinar se orientan a mejorar los resultados, los tiempos y los recursos necesarios para este tipo de proyectos de ingeniería partiendo de los lineamientos del “que debe hacer” el equipo quitando el foco al “cómo gestionar” lo que se hace para que sea más efectiva [37]. Estas teorías expresan su interés sobre el proceso de explotación de información dejando de lado a los actores como el cliente y al equipo de ingeniería de explotación de información [6-29-30].

Mantener los recursos humanos actualizados es una realidad compleja. La rotación de personal y el crecimiento en la cantidad de proyectos requeridos por las organizaciones hacen que sea cada vez más complicado encontrar recursos humanos capacitados en las necesidades y procesos internos

debido a la gran variedad de alternativas y, más allá de las alternativas, cada organización ajusta los procesos propuestos a su estructura interna y necesidades.

Existe una línea de investigación en la ingeniería de software reciente [19] que se dedican a redefinir y a detectar las mejores prácticas dentro de proyectos de desarrollo basados en la ingeniería del software. Si bien en Ingeniería de Explotación de Información existe una gran cantidad de herramientas y resultados de investigación [5-37-23-2] que aportan las herramientas que se debe utilizar en cada paso de los proyectos referidos a la ingeniería de explotación de información no existe cuerpos de conocimiento que aporte a los equipos de ingeniería de explotación de información cómo deben gestionar los proyectos, por lo que se puede postular como un problema abierto establecer buenas practicas que ayuden a facilitar la ejecución de este tipo de proyectos, buscando realizar mejoras a la forma de trabajo y en la gestión de las personas que están involucradas.

II. FUNDAMENTOS

Un grupo de investigadores en el cual forman parte Ivar Jacobson, Bertrand Meyer y Richard Soley crearon en 2009 la iniciativa SEMAT como un movimiento de cambio o actualización de la forma de pensar los proyectos de ingeniería del software [2-9-17]. Propusieron lo que se denominó el llamado a la acción, lo cual parte de una serie de problemas críticos de ingeniería de software y a partir de esto proponen dar un giro a las prácticas tradicionales y modernas con el fin de actualizar la forma de pensarlas y gestionarlas. Ellos dan una propuesta del por qué actuar y cómo hacerlo para solucionar estos problemas. Entre estos los detectados son a grandes rasgos los siguientes [19]:

- La prevalencia de bogas más típicas en la industria de la moda que en una disciplina ingenieril.
- La carencia de una base teórica sonora y ampliamente aceptada.
- La gran cantidad de métodos y variantes de métodos, con diferencias que poco se entienden y que se magnifican artificialmente.
- La carencia de evaluación y validación experimentales y creíbles.
- La separación entre la práctica industrial y la investigación académica.

En [18] puede verse que este grupo no intenta eliminar o dejar de lado las teorías y técnicas creadas hasta el momento, si no que expresan que todo lo relacionado a lean o ágil. Crean

que estar en contra de esto se debe a que la gente adopta soluciones inapropiadas por estar de moda o por adaptarse al mercado.

Según indica [19], para redefinir la ingeniería del software, Semat aporta teorías sólidas, principios probados y mejores prácticas que:

- Incluyan un núcleo de elementos ampliamente aceptados y que se pueda extender a usos específicos.
- Traten asuntos tecnológicos y humanos.
- Los apoyen la industria, la academia, los investigadores y los usuarios.
- Apoyen la extensión ante los requisitos cambiantes y la tecnología.

Semat encontró un terreno común a todo proyecto de desarrollo de software. Este terreno común se dividió en dos grupos, las cosas con las que siempre trabajamos y las cosas que siempre hacemos

El núcleo Semat a través de sus definiciones provee [19]:

- Un marco de pensamiento para que los equipos razonen sobre el progreso que están haciendo y la salud de sus esfuerzos.
- Un terreno común para la discusión, mejoramiento, comparación e intercambio de métodos y prácticas de ingeniería de software.
- Un marco para que los equipos ensamblen y mejoren continuamente su forma de trabajo, mediante la composición de prácticas definidas por separado y de diverso origen.
- Un fundamento para la definición de medidas que no dependan de las prácticas, para evaluar la calidad del software producido y los métodos que se usan para producirlo.
- Más importante aún, una forma de ayudarle a los equipos a comprender dónde están, qué deberían hacer luego y dónde necesitan mejorar.

El núcleo Semat cuenta con una serie de principios fundamentales. Estos principios son: El núcleo es accionable, extensible y práctico.

A. Principio fundamental: el núcleo es accionable.

A través de las cosas con las que siempre trabajamos, el núcleo Semat propone una serie de elementos llamados alfas los cuales son elementos esenciales de esfuerzos de ingeniería del software [19]. Ejemplos de estos son: Interesado, oportunidad, requisitos, sistema software, trabajo, equipo y forma de trabajo.

Cada alfa cuenta con una serie de estados que sirven para evaluar el progreso y la salud del momento. Por ejemplo, el sistema de software cuenta con los siguientes estados: con arquitectura seleccionada, demostrable, usable, listo, operacional y retirado.

A su vez, cada estado cuenta con una lista de chequeos que ayudan al equipo a saber que debe cumplir cada alfa para poder culminar las tareas y que el estado sea cumplido.

Dichos estados y sus listas de chequeo hacen al núcleo accionable.

B. Principio fundamental: el núcleo es extensible.

El núcleo Semat soporta múltiples proyectos, esto lo hace de alguna manera extensible. Pero no es solo eso lo que representa este principio sino que también es extensible en cuanto a sus alfas. Cada alfa representa, como fue mencionado con anterioridad, un esfuerzo de proyectos de ingeniería de software. Estos alfas pueden ser divididos en sub-alfas

representando una parte de dicho esfuerzo con lo cual el seguimiento y control de salud es más específico. Por ejemplo el alfa requisitos puede ser dividido en múltiples requisitos los cuales serán controlados como unidad y que a su vez juntos forman un todo (el alfa requisitos total). Esto hace que el núcleo sea extensible.

Además el alfa forma de trabajo puede contener subalfas de diferentes formas de trabajo y técnicas que pueden ser gestionadas por separado con lo cual queda en manos del equipo decidir que usar y que no. De esta forma son ellos quienes deciden con que se queda y que desecha de cada una. De alguna manera esto aporta a la creación de métodos adaptados.

C. Principio fundamental: el núcleo es práctico.

Este principio determina la manera en que se usa en la práctica el núcleo. Los enfoques tradicionales apuntan a resolver el problema de proyectos de ingeniería de software en cuanto a los enfoques de la ingeniería de procesos e ingeniería de calidad. Este núcleo propone un marco de pensamiento tangible y práctico que apoya a los profesionales de software a medida que realizan su trabajo dado que se puede tocar y usar [16-35] empleando tarjetas. Estas tarjetas proporcionan recordatorios concisos y señales para los miembros del equipo, a medida que realizan sus tareas diarias [19]. Cada tarjeta representa un estado de un alfa y cuenta con una lista de chequeos que proporciona una guía de lo que es necesario hacer para que dicho estado sea concluido. El líder con una carta en su mano lee la lista de chequeos y plantea una lo que hay que hacer para cumplir el ítem. Esto da una visión del estado actual y no libra al azar de la memoria del líder. En pocas palabras el núcleo se convierte en algo que el equipo usa diariamente. A diferencia de los enfoques tradicionales que apuntan a la documentación exagerada de los procesos, el núcleo se enfoca en el uso diario y no en la utilización del método solo cuando se integran nuevos recursos al equipo. Las tarjetas proporcionan una descripción concisa que sirve como recordatorios para los miembros del equipo. Ellos pueden mantener el núcleo como un pequeño mazo de tarjetas en sus bolsillos, que fácilmente pueden extraer para discutir el estado actual del desarrollo, la asignación de trabajo y la colaboración entre los miembros del equipo. Los equipos también pueden discutir áreas de mejoramiento refiriéndose a las tarjetas. Así, el núcleo no es sólo una descripción pesada de lo que un equipo necesita hacer. En su lugar, forma una parte esencial de lo que hacen cada día [19].

III. OBJETIVO

Esta tesis busca desarrollar la adaptación del núcleo Semat [19], que fue pensado para proyectos de ingeniería de software, a proyectos de ingeniería de explotación de información con el fin de aportar a esta última, buenas practicas que colaboren en obtener los mejores resultados, de mejor calidad y optimizando los recursos. Esta línea de trabajo busca potenciar los resultados partiendo de mejorar la forma en que se ejecutan los proyectos basados en ingeniería de explotación de información.

Esta tesis intenta aportar a los equipos que ejecutan proyectos de ingeniería de explotación de información una manera de mejorar los procesos que hoy en día llevan a cabo. Esto indica que el núcleo IMMAT será pensado no para eliminar toda teoría hasta ahora realizada sino para que en conjunto se utilicen y mejoren los proyectos de ingeniería de explotación de información. El núcleo IMMAT colaborará en la mejora de los procesos existentes disminuyendo la distancia

entre la implementación de técnicas y procesos a la forma de trabajo con las nuevas mejoras propuestas por el equipo o por investigadores o consultores.

Esta tesis ofrece plantear una forma de trabajo en los equipos que desarrollan proyectos de ingeniería de explotación de información proponiendo un marco de pensamiento en forma de núcleo con el fin de balancear riesgos y mejorar la forma de trabajo.

IV. METODOLOGÍA DE DESARROLLO

Para construir el conocimiento de la presente investigación, se seguirá un enfoque de investigación clásico [40-7] con énfasis en la producción de tecnologías [42]; identificando métodos y materiales necesarios para desarrollar el proyecto.

A. Métodos

A continuación se definen los métodos que se llevarán a cabo en el presente trabajo. Ellos son:

1) Revisiones sistemáticas

Las revisiones sistemáticas [1] de artículos científicos siguen un método explícito para resumir la información sobre determinado tema o problema. Se diferencia de las revisiones narrativas en que provienen de una pregunta estructurada y de un protocolo previamente realizado.

2) Prototipado evolutivo experimental (método de la Ingeniería)

El prototipado evolutivo experimental [4] consiste en desarrollar una solución inicial para un determinado problema, generando su refinamiento de manera evolutiva por prueba de aplicación de dicha solución a casos de estudio (problemáticas) de complejidad creciente. El proceso de refinamiento concluye al estabilizarse el prototipo en evolución.

B. Materiales

A continuación se detallan los materiales que se utilizarán para el desarrollo de la tesis:

- El núcleo Semat [19].
- Información recolectada por la comunidad científica en su sitio web [20].
- El modelo de proceso KDD [8].
- Metodología CRISP-DM de [6].
- Metodología Catalyst (P3TQ) [37].
- Metodología SEMMA [29-30]
- MoProPei [24].
- Procesos de Explotación de Información basados en Sistemas Inteligentes [5].
- Modelo para Evaluar la Viabilidad de Proyectos de Explotación de Información [37].
- Derivación del proceso de explotación de información [23].
- Propuesta de proceso de transformación de datos para proyectos de explotación de información [3].

C. Metodología

Para alcanzar los Objetivos trazados se propone:

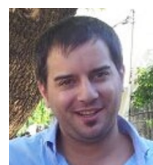
- Realizar una investigación documental sobre Metodologías y procesos de explotación de información. También realizar la investigación documental del núcleo Semat,
- Identificar casos de estudio y casos de validación,
- Desarrollar una propuesta de núcleo IMMAT, incluyendo los alfas, sus relaciones, sus estados y sus listas de chequeo,

- Aplicar el núcleo IMMAT sobre los diferentes casos de estudio identificados,
- Aplicar el núcleo IMMAT sobre los casos de prueba y validación,
- Realizar las simulaciones correspondientes para analizar el comportamiento de los alfas, sus estados y las diferentes listas de chequeo. Analizar los resultados obtenidos. En caso de encontrar problemas, aplicar la solución a la propuesta del núcleo IMMAT, y
- Analizar los resultados obtenidos de las métricas y compararlos con métricas existentes.

REFERENCIAS

- [1] Argimón J. (2004). "Métodos de Investigación Clínica y Epidemiológica". Elsevier España, S.A. ISBN 9788481747096.
- [2] Azoff, M. Apt Methods and Tools, Fujitsu. Ovum Technology Report. Reference Code O100032-002 (Jan. 2011).
- [3] Baldizzoni, E. Un Proceso de Transformación de Datos Para Proyectos de Explotación de Información. Revista Latinoamericana de Ingeniería de Software, 1(2): 57-70, ISSN 2314-2642. 2013.
- [4] Basili, V. (1993). "The Experimental Paradigm in Software Engineering". En Experimental Software Engineering Issues: Critical Assessment and Future Directions (Ed. Rombach, H., Basili, V., Selby, R.). Lecture Notes in Computer Science, Vol. 706. ISBN 978-3-540-57092-9.
- [5] Britos, P. Procesos de Explotación de Información Basados en Sistemas Inteligentes. Tesis Doctoral. Facultad de Informática. Universidad Nacional de La Plata. 2008.
- [6] Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. CRISP-DM 1.0 Step by step Bgguide. Edited by SPSS. 2000.
- [7] Creswell, J. (2002). "Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research". Prentice Hall. ISBN 10: 01-3613-550-1.
- [8] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uhturudsamy, R. Advances in Knowledge Discovery and Data Mining, (editors). AAAI Press. 1996.
- [9] Fujitsu, Ivar Jacobson International AB, Model Driven Solutions. Essence—kernel and language for software engineering. Initial submission, 2012, version 1.0.
- [10] García Martínez, R., Britos, P. Ingeniería de Sistemas Expertos. Editorial Nueva Librería. ISBN 987-1104-15-4. 2004.
- [11] García Martínez, R., Lelli, R., Merlino, H., Cornachia, L., Rodríguez, D., Pytel, P., Arbolea, H. Ingeniería de Proyectos de Explotación de Información para PYMES. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Artículo 3761. 2011.
- [12] García Martínez, R., Merlino, H., Diez, E., Rodríguez, D., Pytel, P., Baldizzoni, E., Arbolea, H., Martins, S. Instrumentos para la Gestión de Proyectos de Explotación de Información. Proceedings del XV Workshop de Investigadores en Ciencias de la Computación, Pág. 137-141. ISBN 978-9-872-81796-1. 2013.
- [13] García-Martínez, R., Britos, P., Martins, S., Baldizzoni, E. 2015. Explotación de Información. Ingeniería de Proyectos. Editorial Nueva Librería ISBN 978-987-1871-34-6.
- [14] Gómez, A., Juristo, N., Montes, C., Pazos, J. Ingeniería del Conocimiento. Centro de Estudios Ramón Areces. S.A., Madrid. 1997.
- [15] Hossian, A. Modelo de Proceso de Conceptualización de Requisitos. Tesis Doctoral. Tesis Doctoral. Universidad Nacional de La Plata, Facultad de Informática. La Plata, Argentina. 2012.
- [16] Jacobson, I., Pan-Wei Ng and Spence, I. Enough of process—let's do practices. Journal of Object Technology 6, 6, 41- 67. 2007.

- [17] Jacobson, I., Meyer, B. Methods need theory. Dr. Dobb's Journal, (2009).
- [18] Jacobson, I., Meyer, B., Soley, R. The SEMAT vision statement, 2009.
- [19] Ivar Jacobson, Pan-Wei Ng, Paul E. McMahon, Ian Spence y Svante Lidman. La Esencia de la Ingeniería de Software: El Núcleo de Semat. Integrantes de la iniciativa Semat (Método y Teoría de la Ingeniería de Software). Traducido por: Carlos Mario Zapata Jaramillo. 2012.
- [20] Web de de la comunidad científica Kdnuggets, <http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>. Página web vigente al 11/11/2015. 2015.
- [21] León León, O., Asato España, J. La Importancia del Modelado de Procesos de Negocio como Herramienta para la Mejora e Innovación. Revista Panorama Administrativo, 4(7). <http://admon.itc.mx/ojs/index.php/panorama/article/view/155/156>. Página web vigente al 11/11/2015. 2009.
- [22] Mariscal, G., Marbán, Ó., González, A., Segovia, J. Hacia la Ingeniería de Data Mining: Un Modelo de Proceso para el Desarrollo de Proyectos. Proceedings V Taller de Minería de Datos y Aprendizaje (TAMIDA '07). Pág. 139-148. ISBN 978-84-9732-602-5. 2007.
- [23] Martins, S. Derivación del Proceso de Explotación de Información desde el Modelado del Negocio. Trabajo Final de Licenciatura en Sistemas. Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús. 2013.
- [24] Martins S., Pesado P., García-Martínez R., Propuesta de Modelo de Procesos para una Ingeniería de Explotación de Información: MoProPEI. Revista Latinoamericana de Ingeniería de Software, 2(5): Pág. 313-332, ISSN 2314-2642. 2014.
- [25] Michalski, R. Bratko, I. Kubat, M. Machine Learning and Data Mining, Methods and Applications (Editores) John Wiley & Sons. 1998.
- [26] Mendez, P., Rodriguez, A. Herramienta de Estudio de Viabilidad para Proyectos que Utilizan la Metodología P3 TQ. Trabajo Profesional de Ingeniería en Informática. Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. Universidad de Buenos Aires. 2009.
- [27] Michalski, R. A Theory and Methodology of Inductive Learning. Artificial Intelligence, 20: 111-161. 1983.
- [28] Michalski, R. Bratko, I. Kubat, M. Machine Learning and Data Mining, Methods and Applications (Editores) John Wiley & Sons. Mobasher, B, R Cooley and J Srivastava (1999). Creating adaptive web sites through usage based clustering of URLs. Proceedings Workshop on Knowledge and Data Engineering Exchange, Pág. 19-25. 1998.
- [29] Moine, J., Gordillo, S., Haedo, A. Análisis Comparativo de Metodologías para la Gestión de Proyectos de Minería de Datos. Proceedings VIII Workshop Bases de Datos y Minería de Datos (WBDDM). Pág. 931-938. 2011.
- [30] Moine, J., Gordillo, S., Haedo, A. Estudio Comparativo de Metodologías para Minería de Datos. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Pág. 278-281. ISBN 978-950-673-892-1. 2011.
- [31] Moss, L., Atre, S. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. AddisonWesley Information Technology Series. 2003.
- [32] Morales, M., Cancino, C. La RSE como Herramienta Estratégica del Negocio. Revista Economía & Administración. Facultad de Economía y Negocios. Universidad de Chile. Nro. 158. Pág. 48-57. ISSN 076-4793. 2009.
- [33] Ochoa, M. Propuesta de Técnicas de Prototipo Metodológica para la Comprensión del Negocio. Reportes Técnicos en Ingeniería del Software, Vol. 7, Nro. 1. Pág. 21-25. 2005.
- [34] Ortín, M. J., García Molina, J., Moros, B., Nicolás, J. El Modelo del Negocio como base del Modelo de Requisitos. Grupo de Investigación de Ingeniería del Software. Departamento de Informática y Sistemas. Facultad de Informática. Universidad de Murcia. España. 2001.
- [35] Pan-Wei Ng and Magee, M. Lightweight application lifecycle management using state cards. Agile Journal. 2010.
- [36] Pollo-Cattaneo, F., Britos, P., Pesado, P., García-Martínez, R. Ingeniería de Procesos de Explotación de Información. En Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica (Editores: R. Aguilar, J. Díaz, G. Gómez, E-León). Pág. 252-263. Alfaomega Grupo Editor. ISBN 978-607-707-096-2. 2010.
- [37] Pyle, D. Business Modeling and Data Mining. Morgan Kaufmann Publishers. 2003.
- [38] Pytel, P. Método de Estimación de Esfuerzo para Proyectos de Explotación de Información. Herramienta Para Su Validación. Tesis de Magister en Ingeniería del Software. Convenio Universidad Politécnica de Madrid e Instituto Tecnológico Buenos Aires. 2011.
- [39] Pytel P., Amatriain H., Britos P., García-Martínez R. Estudio del Modelo para Evaluar la Viabilidad de Proyectos de Explotación de Información. Grupo Investigación en Sistemas de Información. Universidad Nacional de Lanús, Grupo GEMIS. UTN-FRBA y Grupo de Investigación en Explotación de Información. Laboratorio de Informática Aplicada, Universidad Nacional de Río Negro. Argentina. 2012.
- [40] Riveros, H. y Rosas, L. (1985). "El Método Científico Aplicado a las Ciencias Experimentales". México: Editorial Trillas. ISBN 96-8243-893-4.
- [41] Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información. Anales del XVI Congreso Argentino de Ciencias de la Computación. Pág. 664-673. ISBN 978-950-9474-49-9. 2010.
- [42] Sabato J, Mackenzie M. (1982). "La Producción de Tecnología: Autónoma o Transnacional". Instituto Latinoamericano de Estudios Transnacionales - Technology & Engineering. ISBN 9789684293489.



Ezequiel Baldizzoni. Es Analista Programador Universitario y Licenciado en Sistemas por la Universidad Nacional de Lanús. Es Candidato del Programa de Magister en Ingeniería de Sistemas de Información de la Escuela de Postgrado de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional. Es Investigador Tesista del Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información del Grupo de Investigación en Sistemas de Información e Instructor en la Licenciatura en Sistemas de la Universidad Nacional de Lanús.