

Mejora de Procesos de Negocio recuperados mediante Ingeniería Inversa

María Fernández-Ropero¹, Hajo A. Reijers², Ricardo Pérez-Castillo¹ and Mario Piattini¹

¹Instituto de Tecnologías y Sistemas de la Información, University of Castilla-La Mancha
Paseo de la Universidad 4, 13071 Ciudad Real, Spain
{marias.fernandez, ricardo.pdelcastillo, mario.piattini}@uclm.es

²Department of Mathematics and Computer Science, Eindhoven University of Technology
Den Dolech 2, 5612 AZ Eindhoven, The Netherlands
h.a.reijers@tue.nl

Resumen. Con el fin de realizar una adecuada gestión de sus procesos, las empresas cada vez están más interesadas en disponer de modelos de sus procesos de negocio que presenten grados de calidad óptimos. Estos procesos de negocio que son ejecutados por las empresas son a menudo obtenidos utilizando técnicas de ingeniería inversa a partir de sus propios sistemas de información. Sin embargo, estos modelos de procesos de negocio suelen tener un grado menor de calidad debido a la pérdida semántica que conllevan, y no siempre corresponden a representaciones exactas de los procesos de negocio actuales. En este trabajo se presentan todos los problemas detectados que deben ser abordados para mejorar la calidad de los modelos de procesos de negocio especialmente recuperados mediante ingeniería inversa. Entre estos dichos problemas se encuentran la no recuperación automática de elementos relevantes en el modelo o la recuperación de elementos que no son relevantes, el exceso de elementos de granularidad fina que disminuye el nivel de abstracción, entre otros. Este trabajo también sugiere una técnica para mejorar un modelo de procesos de negocio a lo largo de tres fases: reparación, refactorización y mejora semántica.

Palabras Clave: Modelos de procesos de negocio; Desafíos de procesos de negocio; Refactorización; Entendibilidad; Modificabilidad.

1 INTRODUCTION

La gestión de procesos de negocio permite a las organizaciones ser cada vez más eficientes, eficaces y más fácilmente adaptable a los cambios que los enfoques de gestión tradicional, aportando así notables mejoras competitivas [1]. Estos procesos de negocio definen la secuencia de actividades de negocio que son llevadas a cabo por una organización, así como los roles y recursos involucrados, con el fin de alcanzar un objetivo común [2]. Con el fin de facilitar su gestión, los procesos de negocio pueden ser representados mediante modelos, siguiendo notaciones estándares como BPMN (*Business Process Modeling and Notation*) [3]. Sin embargo, estos modelos de procesos de negocio pueden no estar explícitamente presentes en la organización o, en el caso de estarlo, pueden no estar actualizados. En estos casos, los procesos de nego-

cio pueden ser obtenidos a partir de los sistemas de información que los soportan [4] mediante técnicas de ingeniería inversa. Aunque la ingeniería inversa produce procesos de negocio más técnicos y en menor tiempo que, por ejemplo, hacerlo desde cero por expertos, por contra es más propensa a provocar errores [5]. Así, los modelos de procesos de negocio pueden entrañar ciertos problemas que afecten su grado de calidad, ya que toda técnica de ingeniería inversa implica una pérdida semántica [6]. Los trabajos presentes en la literatura identifican los problemas que pueden estar presentes en los modelos de procesos de negocio descubiertos mediante técnicas de minería (por ejemplo, a partir de los eventos automáticamente registrados [7]) o aquellos que han sido diseñados manualmente [8]. Sin embargo, los problemas presentes en este tipo de modelos de procesos de negocio (obtenidos mediante ingeniería inversa) aún no han sido identificados. Estos modelos pueden estar incompletos o contener información no relevante, o incluso contener ciertas ambigüedades que provoquen un difícil entendimiento de ellos y, por tanto, un nivel menor de calidad. En estos casos se hace necesario mejorar dichos modelos de procesos de negocio, abarcando los problemas detectados mientras se acerca el modelo de procesos de negocio cada vez más a la realidad que representa [9].

Con dicho objetivo este artículo presenta el conjunto de problemas detectados en este tipo de modelos de procesos de negocio, los cuales es necesario abordar para mejorar su calidad. Dichos problemas o retos han sido obtenidos tras una revisión de la literatura y tras una serie de experiencias prácticas con modelos de procesos de negocio obtenidos a partir de varios sistemas de información industriales. Además, este trabajo introduce una propuesta para abordar dichos problemas. La propuesta combina la ingeniería inversa con otros enfoques de análisis con el fin de mitigar la pérdida semántica, ya que algunas técnicas de ingeniería inversa usan el código fuente de un sistema de información como fuente para obtener los modelos de procesos de negocio, mientras existen otras fuentes de conocimiento para obtener conocimiento no presente en el propio código fuente. El enfoque propuesto se divide en tres etapas: reparación, refactorización y mejora semántica. Cada una de las etapas hace uso de diversas fuentes de conocimiento adicionales como los eventos registrados durante su ejecución, las líneas guías con buenas prácticas que han sido propuestas en la literatura, heurísticas, la decisión de expertos, entre otras. Aunque las técnicas de refactorización se han convertido en el método más ampliamente usado para mejorar la calidad de modelos de procesos de negocio [10, 11], este trabajo también propone dos etapas adicionales para auxiliar a la etapa de refactorización: reparación y mejora semántica.

El resto del documento se organiza de la siguiente forma: la Sección 2 recoge los problemas presentes en los modelos de procesos de negocio recuperados mediante ingeniería inversa; la Sección 3 presenta el enfoque propuesto para abordar los problemas citados en la sección anterior. Finalmente, Sección 4 discute las conclusiones extraídas del presente trabajo, así como las líneas futuras que deben ser abordadas.

2 PROBLEMAS IDENTIFICADOS

Esta sección presenta los desafíos de la mejora de modelos de procesos de negocio, es decir, los problemas comunes que son identificados en los modelos de procesos de

negocio obtenidos mediante ingeniería inversa. Estos problemas han sido recogidos tras una revisión de la literatura y gracias a experiencias prácticas con modelos obtenidos desde diversos sistemas de información reales. Para ello se ha utilizado MARBLE, una herramienta para recuperar los modelos de procesos de negocio que subyacen en los sistemas de información gracias al análisis de su código fuente [4]. Los modelos de procesos de negocio obtenidos en los casos de estudio industriales realizados usando MARBLE fueron analizados con el fin de descubrir errores comunes y frecuentes. Dichos problemas encontrados son listados a continuación.

- **Complejidad.** Los modelos obtenidos mediante ingeniería inversa pueden estar incompletos. Esto es debido a que, como se mencionó en la introducción, el conocimiento se encuentra distribuido en varias fuentes, no sólo en el propio código fuente, de tal forma que mediante un único análisis estático no se pueden conseguir modelos totalmente completos. Nodos como tareas de negocio, gateways, eventos o datos pueden no haber sido recuperados, así como las conexiones entre dichos elementos. La pérdida de flujos de secuencia entre los elementos dificulta la identificación del orden de ejecución de las distintas tareas, de tal forma que repercute en la completitud semántica del modelo [12]. Además, los puntos de comienzo y fin pueden no estar definidos ya que no se dispone de suficiente información para determinar el orden. Una de las posibles causas de esta pérdida de elementos puede ser que ellos no hayan sido instanciados en tiempo de diseño y, por tanto, no hayan sido recuperados. Por tanto, uno de los retos es redescubrir estos elementos que no han sido automáticamente recuperados por la ingeniería inversa.
- **Granularidad.** De acuerdo al enfoque propuesto por *Zou et al.* [13], cada unidad invocable es considerada por las técnicas de ingeniería inversa como candidata a ser una tarea de negocio. Sin embargo, un sistema de información contiene miles de unidades invocables de diferentes tamaños que son habitualmente considerados como tareas, teniendo diferentes niveles de granularidad [7, 14]: (1) unidades invocables que soportan la funcionalidad principal del sistema como aquellas relativas a la capa de dominio o controlador; (2) unidades invocables pequeñas (como los get y set en programación orientada a objetos) que no desempeñan ninguna tarea de negocio real, tan solo leen o escriben variables; (3) un conjunto de pequeñas unidades invocables que tienen comportamiento similar y desempeñan conjuntamente una tarea de negocio; o (4) un conjunto de unidades invocables pequeñas que conjuntamente dan soporte a otra mayor (tarea padre). El problema de los distintos niveles de granularidad debe ser abordado considerando las tareas de grano fino y descartando aquellas de grano fino que hacen que el modelo sea muy cercano a la perspectiva de código fuente. Autores como *Polyvyanyy et al.* [15] proponen dos técnicas: (1) Eliminación de aquellas tareas que son consideradas irrelevantes, de grano fino, y (2) Agrupación de ciertas tareas en una única tarea de forma que la información sea preservada.
- **Relevancia.** En contraste con la completitud, este problema es relativa a los elementos que sí han sido recuperados cuya eliminación no provoca pérdida semántica ni altera el comportamiento. La relevancia es un aspecto importante ya que asegura que el modelo de procesos de negocio contiene los suficientes elementos para

transmitir su información [12]. Este reto debe ser abordado mediante la eliminación de los elementos no relevantes presentes en el modelo.

- **Incertidumbre.** Una pobre entendibilidad de un modelo de procesos de negocio puede inducir a errores y conclusiones erróneas. Como se comentó anteriormente, la entendibilidad normalmente empeora cuando estos modelos han sido obtenidos mediante ingeniería inversa ya que los identificadores y nombres (etiquetas) de los elementos pueden ser poco descriptivos. Esto se debe a que la mayoría de ellos son heredados del código fuente, que sigue convenciones de nombrado propias de la programación y no convenciones apropiadas para una interpretación intuitiva [12]. Por ejemplo, una etiqueta de una tarea puede consistir en la concatenación de varias palabras que comienzan con mayúsculas. Este problema debe ser abordado mediante el renombrado de los elementos de los modelos de procesos de negocio de forma que representen lo más fielmente posible la semántica que desempeñan.
- **Ambigüedad.** Otro reto es relativo a la ambigüedad de algunos elementos del proceso de negocio. Por ejemplo, dos fragmentos de código fuente pueden construir dos tareas distintas, siendo en realidad la misma tarea. Además, pueden diseñarse modelos de procesos de negocio redundantes a partir de varios fragmentos de código. La ambigüedad, por tanto, afecta negativamente a la entendibilidad y modificabilidad del modelo ya que es relativa a cómo de intuitivamente son formulados los elementos [12]. Un modelo se considerará no ambiguo cuando esté libre de redundancias y no contenga elementos que contradigan la lógica de negocio de otro elemento. Este reto debe ser abordado mediante la detección y eliminación de redundancias e inconsistencias en el modelo de procesos de negocio.

3 ENFOQUE DE MEJORA DE PROCESOS DE NEGOCIO

Con el fin de abordar los retos mencionados anteriormente, este trabajo presenta una propuesta para mejorar los modelos de procesos de negocio que han sido obtenidos mediante ingeniería inversa, con el fin de que reflejen lo más fielmente posible la realidad de negocio con unos niveles óptimos de calidad. Para ello, se propone un marco basado en tres etapas: reparación, refactorización y mejora semántica. Cada etapa aborda algunos de los retos antes mencionados y hace uso de ciertas fuentes de conocimiento para llevar a cabo su propósito. La fase de reparación se considera a menor nivel de abstracción mientras que las fases de refactorización y mejora semántica se representan un nivel de abstracción mayor.

La fase de reparación es considerada a nivel de ingeniería inversa ya que hace uso de fuentes de conocimiento tales como los registros de eventos para completar el modelo. Esta fase asegura que los modelos de procesos de negocio reflejan la ejecución real del sistema de información. Resultados preliminares son propuestos en [16], donde se muestra una serie de pasos que son llevados a cabo para que, partiendo de un modelo de procesos de negocio y el registro de eventos, se obtenga un modelo mejorado que incorpore ciertos flujos de secuencia que no habían sido anteriormente recuperados.

La fase de refactorización, al contrario que la anterior, es la encargada de modificar la estructura interna del modelo de procesos de negocio sin alternar ni modificar su

comportamiento externo. Esta fase mantiene por tanto el nivel de abstracción y mantiene la semántica. Esta técnica da lugar a modelos de procesos de negocio más entendibles, más fáciles de mantener y más reutilizables [17]. Los retos abordados por esta fase son la relevancia, incertidumbre, granularidad y completitud. Para ello, esta fase hace uso de líneas guía propuestas por la literatura, heurísticas, así como la propia experiencia de expertos en modelado de procesos de negocio. En [18] se muestran varios operadores de refactorización que han sido propuestos para aquellos modelos de procesos de negocio obtenidos mediante ingeniería inversa. Por ejemplo, operadores de refactorización que abordan la relevancia mediante la eliminación de nodos aislados o anidaciones innecesarias, la granularidad mediante la agrupación de elementos, entre otras. En ese trabajo se aportan además datos relativos a la mejora obtenida tras la aplicación de cada uno de los operadores propuestos en base a varias métricas que evalúan la entendibilidad y modificabilidad. Además, dicho trabajo revela que el orden de aplicación de los distintos operadores de refactorización influye en el resultado final, de forma que es necesario conocer qué operadores de refactorización es necesario aplicar en cada modelo, así como el orden en el cual aplicarlos.

Finalmente, la fase de la mejora semántica aborda la ambigüedad y la relevancia mediante la decisión de un experto, debido a que no todos los problemas pueden ser abordados por las anteriores fases. Así, en esta fase el nivel de abstracción será mantenido pero, sin embargo, la semántica no será preservada.

4 CONCLUSIONES

La ingeniería inversa se ha convertido en una solución para obtener los modelos de procesos de negocio que subyacen en un sistema de información. Desafortunadamente, estos modelos obtenidos entrañan algunos problemas que deben ser abordados a fin de incrementar su nivel de calidad.

La completitud es uno de dichos retos presentes en los modelos de procesos de negocio obtenidos mediante ingeniería inversa. Esto es debido a que los datos pueden estar distribuidos en varias fuentes. Además, la presencia de elementos de granularidad fina conlleva un nivel de calidad inferior. Similarmente, la recuperación de elementos no relevantes es otro de los problemas ya que un modelo no debería contener elementos adicionales que no desempeñan ninguna lógica de negocio en la organización. La incertidumbre es otro de los problemas ya que afecta negativamente a la entendibilidad por lo que convenciones apropiadas deben ser seguidas por los nombres e identificadores de los elementos del modelo. Otro de los retos es relativo a la ambigüedad, de tal forma que el modelo de procesos de negocio debe de estar libre de redundancias e inconsistencias.

Este trabajo propone un enfoque basado en tres fases (reparación, refactorización y mejorar semántica) para abordar dichos problemas mediante la utilización de fuentes de conocimiento adicionales como los registros de eventos en el caso de la reparación y el uso de heurísticas, guías, etc. para el caso de la refactorización. Entre las líneas futuras de este trabajo se encuentran: (1) refinar, ampliar y mejorar ambas técnicas que ya han sido previamente definidas (reparación y refactorización) con el fin de evitar los problemas anteriores; (2) definir la última fase de mejora semántica con el

fin de utilizar las decisiones de expertos para detectar y eliminar ambigüedades en los modelos, así como, elementos no relevantes no detectados por las anteriores fases.

Agradecimientos. Este trabajo ha sido soportado por el programa FPU así como los siguientes proyectos: MAGO /PEGASO (Ministerio de Ciencia e Innovación [TIN2009-13718-C02-01]) y GEODAS-BC (Ministerio de Economía y Competitividad & Fondos FEDER [TIN2012-37493-C03-01]).

REFERENCES

1. Jeston, J., J. Nelis, and T. Davenport, *Business Process Management: Practical Guidelines to Successful Implementations*. 2nd ed. 2008, NV, USA: Butterworth-Heinemann. 469.
2. Weske, M., *Business Process Management: Concepts, Languages, Architectures*. 2007.
3. OMG. *Business Process Modeling Notation Specification 2.0*. 2011; Available from: <http://www.omg.org/spec/BPMN/2.0/PDF/>.
4. Pérez-Castillo, R., et al., *MARBLE. A Business Process Archeology Tool*, in *27th IEEE International Conference on Software Maintenance*. 2011: Williamsburg, VI. p. 578 - 581
5. Antunes, P., et al., *An End-User Approach to Business Process Modeling*. Journal of Network and Computer Applications, 2013.
6. Fernández-Ropero, M., R. Pérez-Castillo, and M. Piattini, *Refactoring Business Process Models: A Systematic Review*, in *7th ENASE 2012*. Wrocław, Poland. p. 140-145.
7. van der Aalst, W., *Process Mining: Overview and Opportunities*. ACM Transactions on Management Information Systems (TMIS), 2012. **3**(2): p. 7.
8. Indulska, M., et al. *Business process modeling: Current issues and future challenges*. in *Advanced Information Systems Engineering*. 2009: Springer.
9. Fahland, D. and W.M.P.v.d. Aalst, *Repairing Process Models to Reflect Reality*. 2012.
10. Weber, B. and M. Reichert, *Refactoring Process Models in Large Process Repositories*, in *Proceedings of the 20th international conference on Advanced Information Systems Engineering*. 2008, Springer-Verlag. p. 124-139.
11. Dijkman, R., M.L. Rosa, and H.A. Reijers, *Managing large collections of business process models—Current techniques and challenges*. Computers in Industry, 2012. **63**(2): p. 91.
12. Overhage, S., D.Q. Birkmeier, and S. Schlauderer, *Quality Marks, Metrics, and Measurement Procedures for Business Process Models*. Business & Information Systems Engineering, 2012: p. 1-18.
13. Zou, Y. and M. Hung, *An Approach for Extracting Workflows from E-Commerce Applications*, in *Proceedings of the 14th International Conference on Program Comprehension*. 2006, IEEE Computer Society. p. 127-136.
14. Pérez-Castillo, R., et al., *Generating Event Logs from Non-Process-Aware Systems Enabling Business Process Mining*. Enterprise Information System Journal, 2011.
15. Polyvyanyy, A., S. Smirnov, and M. Weske, *Business process model abstraction*. Handbook on Business Process Management 1, 2010: p. 149-166.
16. Fernández-Ropero, M., et al., *Repairing Business Process Models as Retrieved from Source Code*, in *14th BPMDS, in conjunction with CAiSE 2013*: Valencia, Spain.
17. Dijkman, R., et al., *Identifying refactoring opportunities in process model repositories*. Information and Software Technology, 2011.
18. Fernández-Ropero, M., et al., *Assessing the Best-Order for Business Process Model Refactoring*, in *28th Symposium On Applied Computing 2013*: Coimbra, Portugal.