# KALCAS: Un frameworK para evaluación de ALineamiento de arquiteCturAS de información y de negocio

Trabajo de grado para ostentar el t<br/>tulo de Ingenieros de Sistemas de la Universidad Francisco de Paula Santander. Trabajo de Tesis presentado al Departamento de Ingeniería de Sistemas

por

# **Estudiante UFPS**

Asesor: Darío Ernesto Correal Torres

Para optar al título de Magister en Ingeniería

# KALCAS: Un frameworK para evaluación de ALineamiento de arquiteCturAS de información y de negocio

Aprobado por:
Darío Ernesto Correal Torres, Asesor
Rubby Casallas Gutiérrez
Alexis Eduardo Ocampo Ramirez
Fecha de Aprobacin

## Resumen

El alineamiento entre procesos de negocio y tecnologías de información (IT) es la preocupación principal en los estudios de gerencia de IT, debido al impacto directo que ejerce sobre la agilidad y flexibilidad de las organizaciones. La arquitectura empresarial (EA) es un valioso instrumento para evaluar y alcanzar tal alineamiento. Los marcos de EA describen la organización en dominios como Arquitectura de Negocio, Información, Aplicaciones y Tecnología. Diferentes trabajos han propuesto marcos y metodologías para evaluar la alineación a través de los elementos contenidos en los dominios de EA. Sin embargo, suponen tareas manuales como aplicar encuestas o comparar artefactos de EA. Estas tareas manuales aplicadas sobre artefactos heterogéneos son costosas en tiempo y recursos, propensa a errores e imprácticas en extensas EAs. Presentamos KALCAS, una propuesta para soportar alineamiento entre BA e IA a través de comparación automática de sus elementos constitutivos y soportado en Arquitectura Dirigida por Modelos (MDA) y Matching de Ontologías. Nuestros principales objetivos son: i) Soportar el proceso de evaluación de alineamiento entre BA e IA y ii) Automatizar la detección de potenciales alineamientos y desalineamientos entre BA e IA. Presentamos los resultados obtenidos en la aplicación de esta aproximación en el Instituto Colombiano para la Evaluación de la Educación (ICFES).

Keywords: Alineamiento de Negocio y Tecnología, Detección Semi-Automatica de Trazabilidad; Heurísticas de Alineación, Matching de Ontologías, Modelo de EA.

# ABSTRACT

The alignment of Business Processes and Information Technologies (IT) is among the top concerns of IT management surveys, because it has a direct impact on the organization's agility and flexibility to change in response to business needs. Enterprise Architecture (EA) is a valuable instrument to assess and achieve such alignment. EA frameworks describe the organization in domains like Business (BA), Information (IA), Application (AA) and Technology Architecture (TA). Different works have proposed frameworks and methodologies for alignment evaluation across elements contained in EA domains; however, they suppose manual tasks such as applying surveys or comparing EA's artifacts. These manual tasks applied over heterogeneous artifacts are time-costly, error-prone and impractical on large EAs. We introduce KALCAS, a framework to support the alignment between BA and IA via the automatic comparison of their constituent elements supported in Model Driven Architecture (MDA) and Ontology Matching techniques. Our key objectives are: i) To support the process of evaluating BA-IA alignment; ii) to automatically discover traceability among the elements in the BA and IA domains; iii) to detect potential alignments and misalignments between BA and IA. We present the obtained results of this approach applied in the Institute responsible for assessing the quality of education in Colombia.

Keywords: Business-Technology Alignment, Semi-Automatic Traceability Detection; Alignment Heuristic, Ontology Matching, Enterprise Architecture Model.

A Dios quien, ha puesto todas las cosas importantes en mi camino, incluida esta.

A mis padres que con mucho esfuerzo lograron brindarme las bases que hoy me permiten culminar otra etapa.

A mi esposa, que ha sido el soporte sobre el cual he podido construir todo lo que me he propuesto, y que ha invertido tanto o más esfuerzos que yo en la consecución de este logro.

A mi hijo Sebastián, que aún sin saberlo, me aportó el sacrificio, la fuerza y la serenidad necesaria para superar las pruebas que se presentan cuando se persiguen valiosos objetivos

.

# RECONOCIMIENTOS

A Dios y a mi familia que han dispuesto las herramientas que hoy me permiten desarrollar esta trabajo.

Doy gracias a *Darío*, como asesor de tesis me brindó confianza para desarrollar este trabajo, por la dedicación y la guía que me permitieron encontrar la dirección adecuada para llevar a buen puerto esta investigación.

Doy gracias a la *Universidad de los Andes* por ayudarme a mejorar mi capacidades profesionales y desarrollarme académicamente.

Doy gracias al ICFES, a sus directivos que facilitaron el desarrollo de mi maestría, y al valioso grupo de colegas que generosamente me apoyaron en el proceso de validación de esta propuesta.

Finalmente, a los amigos que me brindaron su apoyo y aliento en momentos complicados durante el desarrollo de toda mi maestría.

# Contents

Res	sume	e <b>n</b>
$\mathbf{AB}$	STR	ACTiv
DE	DIC	ATORIA
$\mathbf{RE}$	CON	NOCIMIENTOS
LIS	STA :	DE TABLAS viii
LIS	STA I	DE FIGURAS ix
Ι	INT	TRODUCCIÓN
	1.1	Planteamiento del Problema
	1.2	Objetivos y Contribuciones
	1.3	Estructura del Documento
II	CO	NTEXTO
	2.1	Alineación de Negocio-IT
	2.2	Arquitectura Empresarial
	2.3	Arquitectura Dirigida por Modelos
	2.4	Tartarus
		2.4.1 Arquitectura de Negocio (BA)
		2.4.2 Arquitectura de Información (IA)
	2.5	Ontologías
		2.5.1 Matching de Ontologías
III	SIN	${f ZA}$
	3.1	Trabajo Futuro
	3.2	Publicaciones

# List of Tables

# List of Figures

1	Vista General del Metamodelo Tartarus	
2	Detalle del Metamodelo Tartarus y las Extensions Kalcas	

# Chapter I

# INTRODUCCIÓN

La alineación entre negocio y tecnología puede ser definida como la forma de cuantificar el nivel de coherencia entre las necesidades del negocio y la respuesta ofrecida por las Tecnologías de Información (IT) [?]. Esta alineación es un tema clave en todas las organizaciones. Cada año, cuando los directores de tecnología son encuestados para identificar sus principales prioridades, la necesidad de alinear negocio y IT aparece en los primeros lugares [?]. Gestionar y evaluar la alineación negocio-IT no es fácil, ni en su conceptualización ni en su realización [?]. La falta de alineación es una de las razones fundamentales por la cual las empresas no pueden alcanzar todo el potencial de sus inversiones en IT [?]. Información desactualizada, procesos repetitivos no automatizados, silos de información, procesos y entidades redundantes son ejemplos comunes de dicha desalineación.

Estudios anteriores [?, ?, ?, ?, ?] han propuesto marcos teóricos y metodologías de alineación centrados en la aplicación de encuestas sobre la percepción y tabulación de resultados; pero en general no abordan un análisis apoyado en herramientas automatizadas.

La Arquitectura Empresarial (EA) se presenta como un elemento importante para alcanzar la alineación negocio-IT [?]. Una EA describe la organización en dimensiones o arquitecturas como Negocio (BA), Información (IA), Aplicaciones (AA) y Tecnología (TA) y se han propuesto frameworks para evaluación del alineamiento a través de los elementos contenidos en estos dominios.

Varias propuestas [?, ?, ?, ?] abordan el alineamiento desde la perspectiva de la correspondencia entre los elementos o componentes de los dominios de EA ( i.e. BA, IA, AA y TA). Para determinar esta correspondencia se requiere identificar los diferentes elementos

de la EA, compararlos, establecer las relaciones o trazas entre las diferentes dimensiones y evaluarlos aplicando reglas heurísticas [?, ?, ?] que detectan síntomas de potenciales desalineaciones. Algunas de las heurísticas de alineación BA-IA propuestas son: i) Redundancia de procesos de negocio y activos de información, ii) procesos que no acceden a ninguna entidad y iii) entidades que no son accedidas por ningún proceso.

Según [?, ?, ?], dado que uno de los propósitos de la EA es alinear negocio e IT, el concepto de trazabilidad es esencial para hacer explícita la manera como esta integración es alcanzada en todos los niveles organizacionales. Por tanto [?] define la trazabilidad como la capacidad para hacer explícitas las relaciones entre elementos que se encuentran en diferentes niveles de la EA y que refleja un nivel de alineación e integración entre ellos.

Nuestra aproximación está en línea con los trabajos previos y entiende la trazabilidad como un síntoma o indicador de alineación de Negocio-IT que puede ser inferido partir de los elementos contenidos en una EA. Estudios anteriores sobre evaluación de alineación Negocio-IT a través de los artefactos de EA coinciden en un conjunto de tareas necesarias para implementar tal evaluación: i) Identificar los elementos de cada dimensión de la EA, ii) comparar los elementos de cada dimensión de la EA, iii) definir relaciones entre estos elementos (trazabilidad), iv) identificar alineaciones y desalineaciones y v) asignar un nivel de alineación.

#### 1.1 Planteamiento del Problema

A pesar que uno de los principales objetivos de los marcos de EA es proveer representaciones integradas entre los diferentes niveles organizacionales, es difícil establecer claramente la trazabilidad entre ellos [?]. La verificación de la alineación y la trazabilidad con el paradigma de modelado por dimensiones se ha convertido en un problema importante [?].

Hasta donde conocemos, algunas tareas involucradas en la evaluación de alineamiento como: Comparar los elementos de EA e identificar trazabilidad entre elementos de diferentes niveles se realizan tradicionalmente de forma manual. Esto implica revisar y comparar manualmente un conjunto de artefactos heterogéneos (diagramas, documentos de texto, hojas de cálculo, imágenes) que describen una EA. Entre más elementos tiene cada dominio

de EA, más complejo es el concepto y la evaluación del alineamiento, porque más reglas y heurísticas necesitan ser definidas y aplicadas para gobernar las relaciones entre dichos elementos [?]. Un procedimiento manual de revisión, comparación y asociación implica una alta probabilidad de error, gran inversión de tiempo y recursos [?]. El problema se profundiza en grandes organizaciones con complejas y extensas EA que contienen cientos de artefactos, por lo que esta tarea no solo es compleja, sino a veces inviable en la práctica.

La preguntas de investigación que abordamos en este trabajo son:

- RQ1: ¿Es posible automatizar tareas de evaluación de alineación entre BA-IA y de que manera?
- RQ2: ¿Existe un beneficio en términos de precisión y agilidad al apoyar las tareas de evaluación de alineación entre BA-IA con una aproximación automatizada orientada por modelos y matching de ontologías?

# 1.2 Objetivos y Contribuciones

Los principales objetivos de nuestra propuesta son: i) Apoyar el proceso de evaluación de alineación entre BA-IA. ii) Descubrir automáticamente trazabilidad entre elementos de los dominios de BA e IA. iii) Detectar alineaciones y desalineaciones potenciales entre BA-IA en el marco de una EA.

Las contribuciones centrales de este trabajo se resumen así: Extendemos un metamodelo de EA para formalizar las asociaciones BA-IA. Definimos un procedimiento para inferir trazabilidad entre elementos de BA e IA, apoyado en matching de Ontologías. Construimos Kalcas Query Language (KQL), un Domain Specific Language (DSL) gráfico que permite realizar consultas de alineaciones y desalineaciones entre BA-IA consignadas en el modelo de EA.

#### 1.3 Estructura del Documento

El resto de este documento esta organizado de la siguiente manera: El Capítulo 2 describe el contexto asociado a la problemática y en el cual podemos basar nuestra estrategia. El Capítulo ?? ofrece un caso de estudio para motivar nuestra aproximación. En el Capítulo

?? presentamos nuestra propuesta de solución. El lenguaje de consulta KQL es detallado en el Capítulo ??. El Capítulo ?? ofrece en detalle la implementación de esta propuesta. La experimentación y evaluación realizada se expone en el Capítulo ??. El Capítulo ?? describe el trabajo relacionado. Y finalmente el Capítulo 3 reporta las conclusiones y el trabajo futuro.

# Chapter II

# **CONTEXTO**

En este capítulo se realiza una revisión de los conceptos empleados durante el desarrollo de este documento.

# 2.1 Alineación de Negocio-IT

La alineación entre negocio y tecnología puede ser definida como la forma de cuantificar el nivel de coherencia entre las necesidades del negocio y la respuesta ofrecida por las Tecnologías de Información (IT) [?]. Henderson y Venkatraman [?] definen el alineamiento estratégico a partir de cuatro componentes: i) Estrategía de Negocio, ii) estrategía de IT, iii) procesos e infraestructura organizacional, y iv) procesos e infraestructura de IT. Las diferentes relaciones que se dan entre estos componentes son definidas como Alineamiento Estratégico, Alineamiento Funcional y Alineamiento Transversal.

Otros trabajos [?, ?, ?, ?] han abordado la gestión y evaluación del alineamiento en términos de los componentes contenidos en la EA. En la misma línea, para el framework SEAM [?] el alineamiento negocio-IT corresponde a la trazabilidad entre los niveles de negocio, operación e IT.

#### 2.2 Arquitectura Empresarial

Una Arquitectura Empresarial (EA) ofrece la descripción integral y estructurada de la organización, sus Sistemas de Información (IS), y la forma en que estos se integran a fin de alcanzar los objetivos de negocio apoyados en Tecnologías de Información (IT). Esta descripción se compone de documentos, diagramas y demás artefactos que formalizan diferentes puntos de vista de la organización, de tal manera que sean un referente y soporte

para la toma de decisiones. Los frameworks tradicionales de EA como [?, ?, ?], tienen en común la desagregación en dimensiones: i) La Arquitectura de Negocio (BA) define la estrategia, gobernabilidad, organización y procesos claves de negocio. ii) La Arquitectura de Datos (IA) describe la estructura de los activos de datos lógicos y físicos de la organización y los recursos de gestión de datos. iii) La Arquitectura de Aplicaciones provee un modelo para las aplicaciones a ser desplegadas, sus interacciones y sus relaciones con los principales procesos de negocio de la organización. iv) En la Arquitectura de Tecnología se describen las capacidades de software y hardware que son requeridas para el despliegue de servicios de negocio, datos y aplicaciones.

# 2.3 Arquitectura Dirigida por Modelos

Model-Driven Architecture (MDA) es una propuesta de la OMG para abordar el desarrollo de software proporcionando un conjunto de guías para estructurar especificaciones expresadas en modelos. Es neutral en cuanto a tecnología y proveedor, y busca reducir significativamente el esfuerzo de desarrollo, separando la arquitectura del sistema, de las arquitecturas de plataforma. Uno de los elementos claves de MDA es el Modelo de Plataforma Independiente (*PIM*) que describe la estructura y el comportamiento de un sistema, pero no su implementación. La implementanción en la plataforma particular (JEE, .NET, WS, etc) está definida en un Modelo de Plataforma Específica (*PSM*), el cual es originado a partir del PIM. Para materializar esta conversión, se realizan transformaciones basadas en plantillas detalladas para cada plataforma, que mapean elementos del PIM hacia elementos PSM.

## 2.4 Tartarus

Tartarus es un acercamiento MDA para el análisis de EAs [?]. Tartarus surge como una opción de solución ante la actual variedad de frameworks, estándares, herramientas y formatos que hacen parte de la definción de una EA [?]. El metamodelo descrito en la Figura 1 está compuesto por cinco paquetes: *Enterprise* contiene la estructura, cadena de valor, principios, incentivos organizacionales y demás elementos estratégicos. *Continuum* reúne las definiciones para describir la manera en que la EA evoluciona. *Management* tiene los

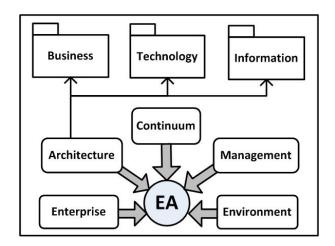


Figure 1: Vista General del Metamodelo Tartarus

factores necesarios para evaluar los artefactos que conforman una arquitectura. *Environment* comprende el conjunto de elementos que describen el entorno en el que funciona la empresa.

Architecture agrupa los conceptos clave para visualizar y estructurar la EA. Architecture contiene los cuatro dominios: Business Domain: Describe los procesos de negocio. Technology Domain: Comprende las capacidades de software y hardware que soportan los servicios de negocio e información. Information Domain: Estructura los componentes de datos que conforman la información de la empresa.

Vamos a describir el metamodelo, detallando el dominio de información (izquierda), procesos de negocio (derecha) como parte del contexto de nuestro trabajo. La Figura 2 muestra el metamodelo de los dominios mencionados.

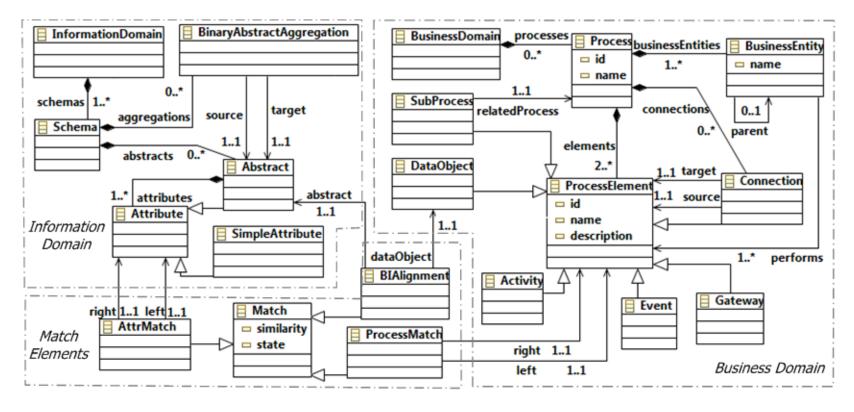


Figure 2: Detalle del Metamodelo Tartarus y las Extensions Kalcas

### 2.4.1 Arquitectura de Negocio (BA)

Este dominio define los procesos de negocio de la compañía. En el metamodelo se resaltan los elementos de proceso (ProcessElements), Entidades de Negocio, Objetos de Flujo y Conexiones. El metamodelo aborda los diferentes tipos de actividades, eventos y flujos derivados de la nomenclatura BPMN (Business Process Modeling Notation). El concepto DataObject asocia las entidades de datos leídas y/o producidas por las actividades.

Para nuestro caso el Proceso de Registro corresponde a un elemento Process el cual contiene las once actividades (Activity) conectadas por elementos de clase Connection y/o Gateway. Los objetos de datos como Payment Format y Card se consignan como instancias de tipo DataObject.

### 2.4.2 Arquitectura de Información (IA)

Nuestro metamodelo de arquitectura de información es una adaptación del trabajo propuesto en [?], enriquecido con las definiciones de las relaciones inferidas entre entidades, comentarios de tablas y comentarios de columnas. La metaclase Schema representa los esquemas contenidos en la EA. Para nuestro caso, el esquema S1 se convierte en la instancia Schema:S1. La metaclase Attribute está especializada en dos subclases: SimpleAttribute la cual define columnas en la base de datos o tipos primitivos en esquemas XML, tienen un tipo de dato (INTEGER, DOUBLE, STRING, etc). Por otro lado, Abstract se refiere a entidades en un modelo relacional o tipos de dato complejos en XML Schema.

Por ejemplo, la entidad USER del esquema S1 se convierte en un objeto Abstract:S1.USER y cada uno de sus campos (name, document, etc) son objetos de clase SimpleAttribute con sus respectivos tipos de dato. En BinaryAbstractAggregation, se definen las relaciones existentes entre cada par de elementos Abstract. La relación entre las entidades USER e REGISTRATION se representa con la asociación BinaryAbstractAggregation: USER\_REGISTRATION

# 2.5 Ontologías

Una ontología, básicamente, es una descripción explícita de un dominio de conocimiento específico, definida en términos de sus conceptos, propiedades, atributos, restricciones e individuos [?]. Formalmente podemos definir una ontología como:  $O = \{C, P, H^C, H^P, A^O, I, R^I\}$ . Donde C es el conjunto de conceptos, P el conjunto de propiedades.  $H^C$  es la jerarquía de relaciones entre los conceptos tal que  $H^C \subset C \times C(c_i, c_j) \in H^C$  denota que el concepto  $c_i$  es subconcepto de  $c_j$ . De la misma manera  $H^P$  define las relaciones jerárquicas entre propiedades.  $A^O$  es el conjunto de axiomas. I comprende el conjunto de Individuos, es decir, instancias de conceptos y propiedades quienes se asocian a través de instancias relacionales  $R^I$ . Una de las principales ventajas de las ontologías, es proveer características útiles para sistemas inteligentes, representación e ingeniería de conocimiento [?].

#### 2.5.1 Matching de Ontologías

La función de alineación de ontoloías ha sido definida formalmente [?, ?]:  $f(O_1, O_2) = \{e_{i1}, e_{i2}, i_i, r_i\}$ . Donde  $O_1$  y  $O_2$  son los esquemas/ontologías de entrada, comúnmente llamados origen y destino respectivamente,  $e_{i1}$  y  $e_{i2}$  son las dos entidades comparadas,  $i_i$  corresponde al índice de similitud o confianza (medido entre 0 y 1) y  $r_i$  la relación (igualdad, especialización, generalización) que puede haber entre  $e_{i1}$  y  $e_{i2}$ . Detectar elementos similares entre diferentes fuentes de información es también una necesidad central en procesos de evaluación, migración, integración y evolución de SI, intercambio de información en sistemas P2P y composición de web services [?].

En 2004 surge la Ontology Alignment Evaluation Initiative (*OAEI*), una iniciativa que anualmente evalúa sistemas de alineamiento de ontologías. El objetivo de la OAEI es comparar diferentes propuestas, con el objetivo de ofrecer conclusiones sobre las mejores técnicas y estrategias, para lo cual, provee unos casos de prueba sobre los cuales los diferentes sistemas experimentan. Entre los temas evaluados se encuentran: *The benchmark track, The directories and thesauri track, Instance matching*.

Existen diferentes métodos y técnicas para implementar alineamiento automático de ontologías [?, ?, ?] y nuestra propuesta incluye algunos de ellos. Las principales técnicas

de alineamiento son basadas en esquema, basadas en contenido y combinadas. Las basadas en esquema sólo tienen en cuenta la información estructural del esquema, no su contenido. Dentro de este grupo se aplican comparaciones lingüísticas, textuales, de restricciones y estructurales. Las estrategias basadas en contenido involucran estadísticas, patrones o incluso los mismos datos para inferir correspondencias. Las técnicas combinadas aplican, en conjunto, las anteriores aproximaciones en busca de mejores resultados. Esta combinación puede ser configurada manual o automáticamente utilizando aprendizaje de máquina.

Este trabajo no busca determinar la mejor forma de realizar alineamientos, sino adaptar y aplicar técnicas y avances en el área de matching de ontologías para inferir trazabilidad entre componentes de procesos e información.

# Chapter III

# **SINZA**

En este capítulo discutimos SOBRE SINZA

# 3.1 Trabajo Futuro

Las siguientes son algunas propuestas de trabajo futuro que hemos visualizado como continuación o extensiones a nuestra propuesta:

- Futuras investigaciones pueden incorporar la definición de métricas de alineación que permitan evaluar una EA frente a niveles de madurez presentados en trabajos anteriores como [?].
- La inclusión de los demás dominios de la EA como Aplicaciones (AA) y Tecnología (TA) mejoraría la completitud de esta propuesta. Inferir trazabilidad con elementos que hacen parte de la estrategia de la organización como drivers, principios, objetivos aportaría acercaría mas nuestra propuesta a la alineación negocio-IT. En este punto hay consideraciones importantes a investigar, como el hecho de alinear elementos a diferente nivel de granularidad, como lo son por ejemplo un driver de negocio y una entidad de información.
- Los mapeos generados con este framework podrían exportarse a lenguajes estándar de integración de EA como ArchiMate [?]. De tal manera que diferentes herramientas puedan reutilizar las inferencias alcanzadas con Kalcas. Otros tipos de consultas que resuelvan diferentes preguntas sobre el modelo pueden enriquecer el editor KQL, por ejemplo obtener todas las actividades o entidades que no estén alineadas.

- Las tareas de alineamiento en la fase de experimentación han utilizado algunos algoritmos de Alignment API, extender las pruebas en cuanto a motores y algoritmos para evaluar mas resultados puede mejorar la exactitud promedio de Kalcas. Para el momento de nuestro trabajo no hay disponible un algoritmo lingüístico aplicable al idioma español, por tanto un avance en esa dirección es muy valioso no solo para esta propuesta sino para el campo del matching de ontologías en general.
- En este trabajo se utilizó la trazabilidad inferida para apoyar análisis de alineación,
   pero nuevas investigaciones pueden emplear la trazabilidad para soportar el análisis
   de impacto en una EA. Nuestra propuesta podría ser extendida o modificada para ser
   usada con otros metamodelos de EA diferentes a Tartarus.

#### 3.2 Publicaciones

Durante toda la realización de este trabajo de tesis, fueron aceptadas publicaciones en modalidad *full research papers* en los siguientes eventos internacionales:

- 16th East-European Conference on Advances in Databases and Information Systems (ADBIS 2012): KALCAS: A framework for semi-Automatic aLignment of data and business proCesses ArchitectureS. Poznan, Polonia. Septiembre 2012.
- XXX International Conference of the Chilean Computer Science Society (SCCC 2011):
   An Ontology-Matching based Proposal to Detect Potential Redundancies on Enterprise Architectures. Curicó, Chile. Noviembre 2011.
- XXXVII Conferencia Latinoamericana de Informática (XXXVII CLEI): Detección de Elementos Redundantes en Arquitecturas de Información: Un Enfoque Apoyado en Alineación de Ontologías. Quito, Ecuador. Octubre 2011.