

Formalización de Dominios de Negocio para Proyectos de Explotación de Información basada en Técnicas de Ingeniería del Conocimiento

Vegega, C., Pytel, P., Ramón, H., Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R.

Grupo Investigación en Sistemas de Información. Departamento Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús.
Grupo de Investigación en Explotación de Información. Laboratorio de Informática Aplicada. Universidad Nacional de Río Negro.
Grupo de Estudio en Metodologías de Ingeniería de Software. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.
cinthia.vegega@gmail.com, fpollo@posgrado.frba.utn.edu.ar, paobritos@gmail.com, rgarcia@unla.edu.ar

Resumen. La documentación de requisitos, realizada durante el proceso de educación en un proyecto de Explotación de Información comprende el llenado de un conjunto de plantillas que permitirán tener una descripción detallada de cada requisito educido. Este trabajo busca satisfacer a la necesidad de estructurar el proceso de documentación, proponiendo, para ello, un proceso de formalización de los dominios del negocio basado en técnicas de representación del conocimiento provistas por la Ingeniería del Conocimiento. Este proceso contempla la utilización de un conjunto de técnicas y herramientas que agilicen la tarea de documentar los requisitos en este tipo de proyectos.

Palabras Clave: Explotación de Información. Ingeniería del Conocimiento. Educación de Requisitos. Documentación de Requisitos.

1. Introducción

Este trabajo continúa con la línea de investigación definida en [1], en donde se plantea la necesidad de adaptar el proceso tradicional de especificación de requisitos de sistemas software para proyectos de Explotación de Información (EdI) y se propone un proceso de educación de requerimientos que permite documentar los requisitos educidos. Para ello, se define el problema detectado en el ámbito de los proyectos de EdI respecto del llenado de las plantillas (sección 2) y se presenta una solución propuesta a esta problemática (sección 3). Se describen las técnicas de Ingeniería del Conocimiento utilizadas en la solución (sección 4) y se expone la prueba de concepto realizada (sección 5), para finalmente presentar las conclusiones obtenidas (sección 6).

2. Definición del Problema

La documentación de requisitos que se realiza durante el proceso de educación en proyectos de EdI, se basa en completar un conjunto de plantillas [2] que permiten formalizar los requisitos educidos, de forma de tener una descripción detallada de cada uno de ellos, teniendo presente qué es lo que se debe cumplir en el proyecto para satisfacer las expectativas del cliente. El llenado de las plantillas está a cargo de los ingenieros de EdI, quienes al no contar con un proceso estructurado que automatice esta documentación, pierden gran cantidad de tiempo en esta tarea. Este tiempo podría ser utilizado para enfocarse en el cumplimiento de los objetivos del proyecto, sin perder por ello, información relevante para el mismo.

En este contexto, se busca reducir la complejidad inherente al llenado de las plantillas, a través de la aplicación de técnicas de representación de conocimiento de la Ingeniería del Conocimiento, de forma de ofrecerles a los ingenieros de EdI un proceso estructurado que contemple la utilización de un conjunto de técnicas y herramientas que agilicen la tarea de documentar los requisitos y mejoren su performance.

3. Solución Propuesta

Para la resolución del problema planteado, se propone un proceso que aplica las técnicas de representación de conocimiento pertenecientes a la Ingeniería del Conocimiento en el llenado de las plantillas [2] empleadas para la documentación de requisitos en proyectos de EdI. El proceso propuesto se estructura en tres fases: (a) Identificación de los Términos Generales del Dominio, cuyo objetivo es definir y detallar el vocabulario utilizado en el negocio en que se está llevando a cabo el proyecto de EdI, (b) Identificación de las Relaciones entre los Conceptos del Dominio, cuyo objetivo es establecer las relaciones que existen entre los conceptos que se utilizan dentro del negocio objeto de estudio y (c) Documentación de Resultados, que tiene como objetivo el llenado de las plantillas correspondientes, según la información obtenida en las fases anteriores. La figura 1 representa el proceso propuesto, destacando las fases, actividades y productos confeccionados en cada actividad realizada.

La fase de Identificación de los Términos Generales del Dominio se estructura en dos actividades: (a) Armar el glosario de términos del dominio y (b) Describir las relaciones entre los términos del dominio. El contexto en el cual se desarrollará el proyecto constituye la entrada para la actividad “Armar el glosario de términos del dominio” que produce como resultado el glosario de términos, en el cual se detallan las características principales de cada uno de ellos. En la tabla 1 se describen las características de cada término y se explica el significado de las mismas. El glosario de términos es la entrada para la segunda actividad “Describir las relaciones entre los términos del dominio.” que produce como resultado la tabla de Conceptos, Atributos y Valores, tal como se detalla en la tabla 2.

La fase de “Identificación de las Relaciones entre los Conceptos del Dominio” se estructura en dos actividades: (a) Definir las relaciones existentes entre los conceptos del dominio y (b) Representar las relaciones existentes entre los conceptos del dominio. El glosario de términos, también, será la entrada para la actividad “Definir las

relaciones existentes entre los conceptos del dominio” que produce como resultado la tabla de Concepto-Relación, que se muestra en la tabla 3.

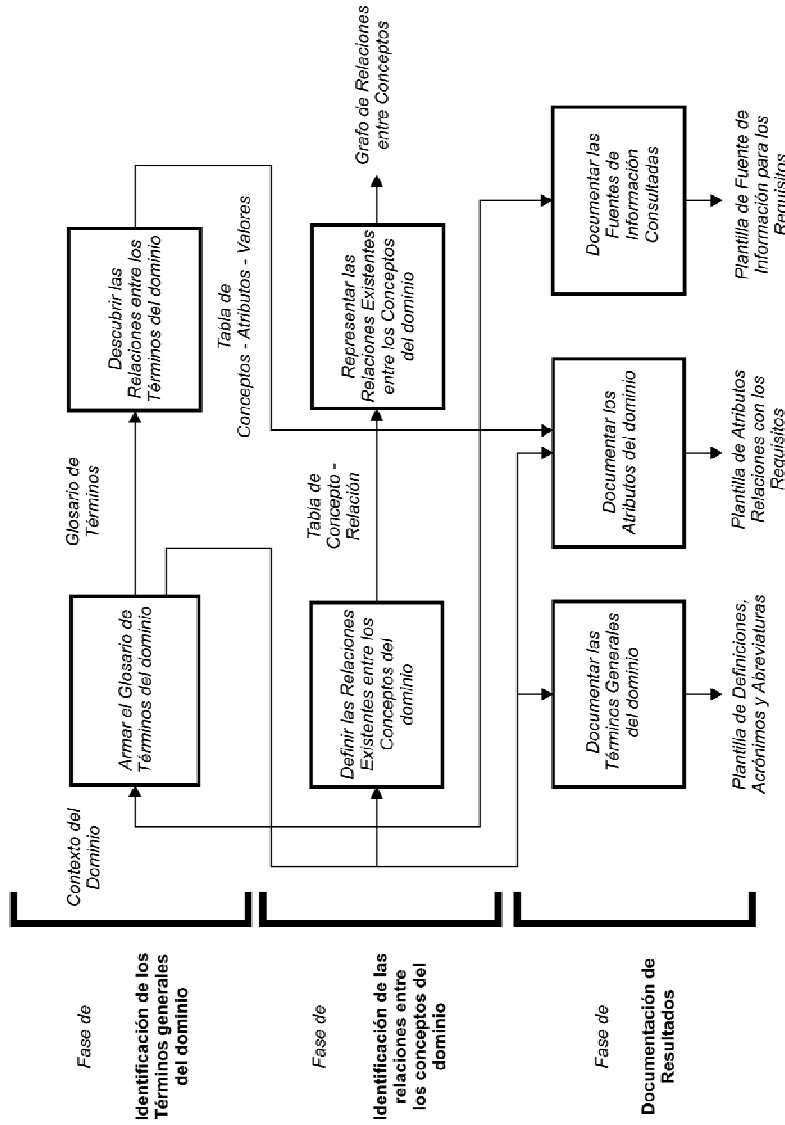


Figura 1. Proceso propuesto detallando fases, actividades y productos.

Tabla 1. Descripción de las características definidas por cada término del dominio.

Término:	Nombre del término.
Tipo:	Tipología: Concepto o Atributo.
Descripción:	Breve explicación del significado del término.
Función:	Función que cumple en el dominio.
Tipo de Valores:	En el caso de ser un atributo, se define la clase de valor que puede ejemplificar ese atributo: alfanumérico, numérico, etc.
Rango de Valores:	En el caso de ser un atributo, se define una lista de posibles valores descriptivos.
Sinónimos/Acrónimos:	Terminología análoga al nombre del término.
Fuente de Datos:	Forma de encontrar ese término dentro del dominio.
Referencia:	Descripción de los documentos y sesiones de adquisición de conocimientos (ej: entrevistas) que proporcionen la información acerca de ese término.

Tabla 2. Descripción del significado de los términos: Conceptos, Atributos, Valores.

Conceptos	Atributos	Valores
Nombre del concepto	Cualidades del concepto	Tipo de valores del atributo definido

Tabla 3. Descripción de la tabla Concepto-Relación.

Conceptos	Conceptos Asociados	Relación	Descripción
Nombre del concepto	Nombre del concepto asociado	Nombre de la relación existente entre los conceptos definidos	Breve explicación detallando la relación

La tabla de Concepto-Relación es la entrada para la segunda actividad “Representar las relaciones existentes entre los conceptos del dominio”, la cual busca graficar las relaciones entre los conceptos, para visualizar con más claridad toda conexión entre ellos. El producto de esta fase es el grafo de relaciones entre conceptos, el cual constituye uno de los productos finales de la solución propuesta, que le permite al ingeniero de EdI la comprensión del dominio, la cual es necesaria antes del llenado de las plantillas correspondientes. La figura 2 muestra la forma de armar el grafo, teniendo en cuenta la descripción de las relaciones ya identificadas en la primera actividad de esta fase.

La fase de Documentación de Resultados se estructura en tres actividades: (a) Documentar los términos generales del dominio, (b) Documentar los atributos del dominio y (c) Documentar las fuentes de información consultadas. El glosario de términos constituye la entrada para la actividad “Documentar los términos generales del domi-

nio”, que produce como producto el llenado de la plantilla de Definiciones, Acrónimos y Abreviaturas.

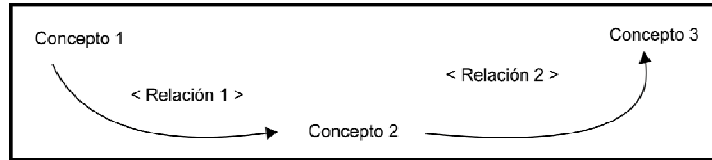


Figura 2. Ejemplo de armado del grafo de relaciones entre conceptos.

La figura 3 muestra la correspondencia de datos entre el glosario de términos y la plantilla correspondiente a esta actividad. El glosario de términos junto con la tabla de Concepto-Característica-Valor es la entrada para la segunda actividad “Documentar los atributos del dominio” que produce como producto el llenado de la plantilla “Atributos relacionados con los Requisitos”. La figura 4 muestra la correspondencia de datos entre el glosario de términos y la tabla de Concepto-Atributo-Valor, respecto de la plantilla. Por último, el contexto en el cual se desarrollará el proyecto constituye la entrada para la tercera actividad “Documentar las fuentes de información consultadas” que produce como resultado el llenado de la plantilla “Fuente de Información para los Requerimientos”, en la cual se registrarán las fuentes de datos consultadas en cada sesión de adquisición realizada, junto con el responsable de dicha fuente.

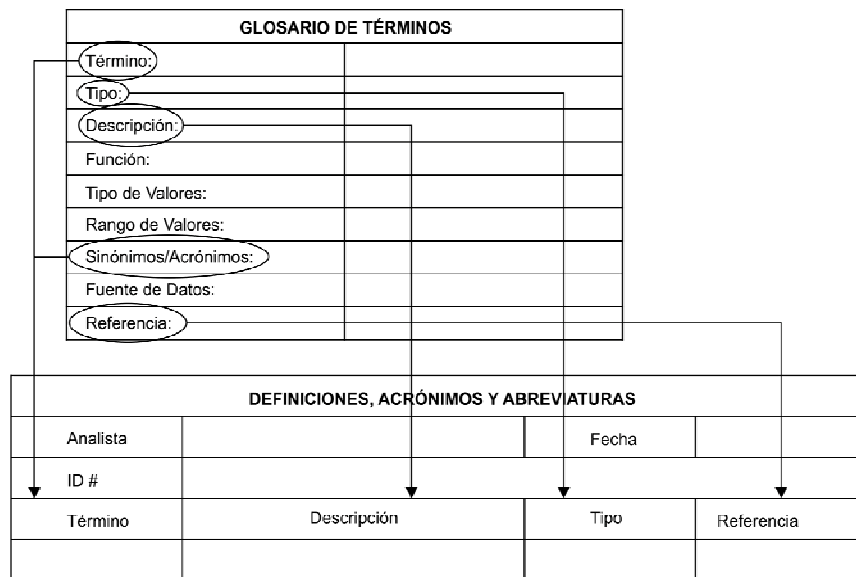


Figura 3. Correspondencia de datos entre glosario de términos y plantilla de Definiciones, Acrónimos y Abreviaturas.

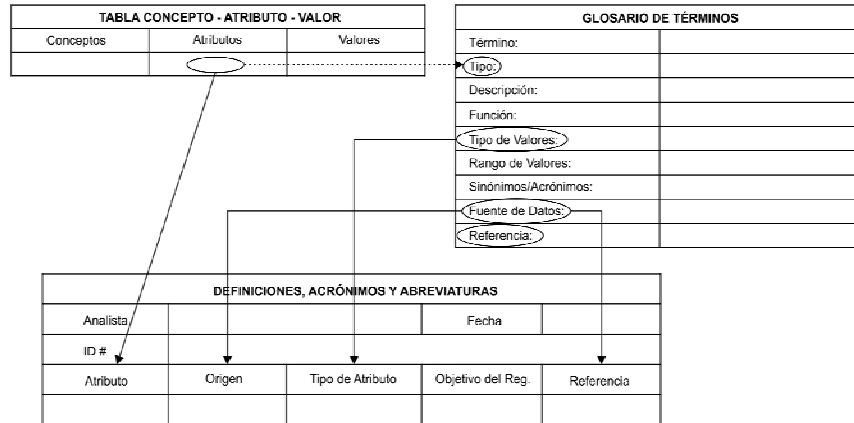


Figura 4. Correspondencia de datos entre glosario de términos, tabla de Concepto-Atributo-Valor y plantilla de Atributos Relacionados con los Requisitos.

Las técnicas de modelado utilizadas y las representaciones de las actividades con sus respectivos productos de entrada y salida, se resumen en la tabla 4.

Tabla 4. Fases, Actividades, Técnicas de Modelado y Productos.

FASE	ACTIVIDAD	PRODUCTOS DE ENTRADA	TECNICA DE MODELADO UTILIZADA	PRODUCTOS DE SALIDA
Identificación de los términos generales del dominio	Armar el glosario de términos del dominio	Contexto del Dominio	Diccionario de Conceptos Descripción de Atributos	Glosario de Términos
	Describir las relaciones entre los términos del dominio	Glosario de Términos	Tripleta Concepto-Atributo-Valor	Tabla de Concepto-Atributo-Valor
Identificación de las relaciones entre los conceptos del dominio	Definir las relaciones entre los conceptos del dominio	Glosario de Términos	Descripción Concepto-Relación	Tabla de Concepto-Relación
	Representar las relaciones existentes entre los conceptos del dominio	Tabla de Concepto-Relación	Red Semántica	Grafo de Relaciones entre Conceptos
Documentación de resultados	Documentar los términos generales del dominio	Glosario de Términos	Volcado de Información	Plantilla de Definiciones, Acrónimos y Abreviaturas
	Documentar los atributos del dominio	Glosario de Términos Tabla de Concepto-Atributo-Valor	Volcado de Información	Plantilla de Atributos relacionados con los requisitos
	Documentar las fuentes de información consultadas	Contexto del Dominio	Volcado de Información	Plantilla de Fuente de Información para los Requerimientos

4. Descripción de Técnicas de Ingeniería del Conocimiento

En el proceso propuesto, se utilizaron técnicas de representación de conocimiento provistas por la Ingeniería del Conocimiento. Algunas de estas técnicas, se aplicaron tal como se encuentran definidas [3], como por ejemplo, las Tripletas Concepto-Atributo-Valor y las Redes Semánticas. Otras, en cambio, son técnicas híbridas que surgieron de la conjunción de otras técnicas, como detallamos a continuación:

- *Glosario de Términos:* El Glosario de Términos surge a partir del Diccionario de Conceptos y la Descripción de Atributos. Dado que ambas técnicas modelaban un objeto específico del dominio, se detectó la necesidad de poseer en un solo repositorio, todos los términos que se utilizarán a lo largo del proyecto de EdI, distinguiendo el tipo de término que se documenta. De esa forma, se tendrían tanto los conceptos, como los atributos y sus valores, en forma unificada, facilitando el acceso a los mismos y sus descripciones.
- *Tabla de Concepto-Relación:* Esta técnica se utilizó para describir las relaciones existentes entre los conceptos del dominio, dado que se necesitaban especificar relaciones complejas, sin utilizar para ello las relaciones estándares propuestas en las Redes Semánticas.
- *Grafo de Relaciones entre Conceptos:* Este grafo representa en forma gráfica las relaciones especificadas en la tabla de Concepto-Atributo, de forma tal de que sea más ágil su visualización y utilización.

5. Prueba de Concepto

Hemos definido un proceso de formalización de los dominios del negocio, que contempla parte de la documentación de requisitos necesaria en proyectos de EdI, a partir de la aplicación de técnicas de representación del conocimiento, pertenecientes a la Ingeniería del Conocimiento.

La prueba de concepto se realizó aplicando el proceso propuesto, al caso de estudio “Detección de Patrones de Daños y Averías” [4]. Respecto de la primera fase “Identificación de los Términos Generales del Dominio”, en la tabla 5 se muestra un extracto del glosario de términos, como producto de la primera actividad “Armar el glosario de términos del dominio”, en el cual se detallan algunos de los términos más representativos del caso. Asimismo, en la tabla 6 se muestra la tabla de Concepto-Atributo-Valor, como producto de la segunda actividad “Describir las relaciones entre los términos del dominio”. Esta tabla muestra la relación entre los términos anteriormente descritos.

Respecto de la segunda fase “Identificación de las Relaciones entre los Conceptos del Dominio”, la tabla 7 es un extracto de la tabla de Concepto-Relación, que es el producto de la primera actividad “Definir las relaciones existentes entre los conceptos del dominio” y la figura 5 muestra parte del grafo de relaciones entre conceptos como producto de la segunda actividad “Representar las relaciones existentes entre los conceptos del dominio”, que constituye uno de los productos finales del proceso.

Respecto de la tercera fase, correspondiente a la Documentación de Resultados, las tablas 8, 9 y 10, hacen referencia a parte de las plantillas de “Definiciones, Acrónimos y Abreviaturas”, “Atributos relacionados con los Requisitos” y “Fuente de Información para los Requerimientos”, respectivamente.

Tabla 5. Extracto del Glosario de Términos

Término:	Vehículo
Tipo:	Concepto
Descripción:	Entidad que transita por un conjunto de lugares en los cuales puede sufrir daños que afecten su funcionamiento.
Función:	Es el objeto de estudio para determinar si hubo un daño o no.
Tipo de Valores:	----
Rango de Valores:	----
Sinónimos/Acrónimos:	Automóvil, Unidad.
Fuente de Datos:	Base de Datos.
Referencia:	Entrevistas.

Término:	Número de Identificación del Vehículo
Tipo:	Atributo
Descripción:	Código que se encuentra impreso en la documentación del vehículo, a través de un código de barras, mediante el cual, se pueden conocer datos del automóvil, tal como su origen, marca, modelo o color.
Función:	Permite la identificación del vehículo.
Tipo de Valores:	Numérico
Rango de Valores:	17 a 18 dígitos
Sinónimos/Acrónimos:	VIN (acrónimo de Vehicle Identification Number)
Fuente de Datos:	Base de Datos.
Referencia:	Entrevistas
Término:	Avería
Tipo:	Concepto
Descripción:	Tipo de daño producido en un vehículo.
Función:	Falla detectada en un vehículo que permite ser clasificada para obtener conclusiones dentro del proyecto a realizar.
Tipo de Valores:	----
Rango de Valores:	----
Sinónimos/Acrónimos:	Daño.
Fuente de Datos:	Base de Datos.
Referencia:	Entrevistas

Tabla 6. Tabla de Concepto-Atributo-Valor

Concepto	Atributo	Valor
Vehículo	Número de Identificación	Numérico
	Modelo	Numérico
Avería	Gravedad	Alfanumérico
	Código de Clave	Alfanumérico
	Observación	SI/NO

Tabla 7. Tabla de Concepto-Atributo-Valor

Conceptos	Conceptos Asociados	Relación	Descripción
Vehículo	Parte	está formado por	El vehículo está formado por partes.
Parte	Avería	tiene	Las partes del vehículo poseen un determinado daño que debe ser analizado, para observar su gravedad y tipo.

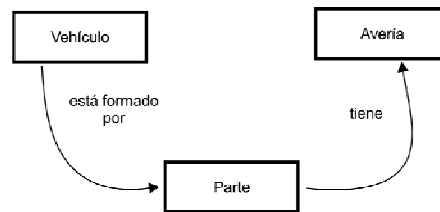


Figura 5. Grafo de relaciones entre conceptos.

Tabla 8. Plantilla de Definiciones, Acrónimos y Abreviaturas

DEFINICIONES, ACRÓNIMOS Y ABREVIATURAS			
<i>Analista</i>	Cynthia S. Vegega	<i>Fecha</i>	17/10/2011
<i>ID #</i>	PATRONES-AVERIAS		
<i>Término</i>	<i>Descripción</i>	<i>Tipo</i>	<i>Referencia</i>
Avería	Tipo de daño producido en un vehículo.	Concepto	Entrevistas
Número de Identificación del Vehículo (VIN)	Código que se encuentra impreso en la documentación del vehículo, a través de un código de barras, mediante el cual, se pueden conocer datos del vehículo, tal como su origen, marca, modelo o color.	Atributo	Entrevistas
Vehículo	Entidad que transita por un conjunto de lugares en los cuales puede sufrir daños que afecten su funcionamiento.	Concepto	Entrevistas

Tabla 9. Plantilla de Atributos relacionados con los Requisitos

ATRIBUTOS RELACIONADOS CON LOS REQUISITOS				
<i>Analista</i>	Cynthia S. Vegega	<i>Fecha</i>	17/10/2011	
<i>ID #</i>	PATRONES-AVERIAS			
<i>Atributo</i>	<i>Origen</i>	<i>Tipo de Atributo</i>	<i>Objetivo del Requisito</i>	<i>Referencia</i>
Número de Identificación del Vehículo	Base de Datos	Númerico	----	Entrevistas

Tabla 10. Plantilla de Fuente de Información para los Requerimientos

FUENTE DE INFORMACION PARA LOS REQUERIMIENTOS				
<i>Analista</i>	Cynthia S. Vegega	<i>Fecha</i>	17/10/2011	
<i>ID #</i>	PATRONES-AVERIAS			
<i>Origen</i>	<i>Tipo</i>	<i>Descripción</i>	<i>Responsable</i>	<i>Referencia</i>
Base de datos de los movimientos de importación y exportación	Base de Datos	Información vinculada a los movimientos de importación y exportación de vehículos (dos marcas del mercado argentino).	Empresas Involucradas	Entrevistas

6. Conclusiones

A partir del proceso propuesto, en este trabajo, se puede automatizar una parte de la documentación de requisitos en proyectos de Explotación de Información, utilizando un conjunto de técnicas de la Ingeniería del Conocimiento. Mediante este proceso, se pudieron completar en su totalidad las plantillas de “Definiciones, Acrónimos y Abreviaturas” y “Fuente de Información para los Requerimientos” y en parte la plantilla de “Atributos relacionados con los Requisitos”. El proceso genera además un producto final en la segunda fase, que no se relaciona directamente con las plantillas definidas en el proceso de educación de requisitos, pero que es necesario para el entendimiento del dominio por parte del ingeniero encargado del proyecto. A partir de esto, como línea futura de investigación se propone la realización de una nueva plantilla que permita documentar también este producto final, de forma de documentar en una misma estructura toda la información obtenida. Se propone también, la definición de nuevos procesos para completar el resto de las plantillas utilizadas en la educación de requerimientos y el estudio de cada una de ellas para detectar si en alguna plantilla faltan datos relevantes que podrían ser agregados, dado que las técnicas utilizadas aportan una gran cantidad de datos que las plantillas no contemplan en su estructura.

7. Referencias

1. F. Pollo-Cattaneo, P. Britos, P. Pesado, R. García-Martínez (2009). *Metodología para Especificación de Requisitos en Proyectos de Explotación de Información*. Proceedings XI Workshop de Investigadores en Ciencias de la Computación. Pág. 467-469. ISBN 978-950-605-570-7.
2. P. Britos, O. Dieste, R. García-Martínez (2008). *Requirements Elicitation in Data Mining for Business Intelligence Projects*. IFIP Series, 274: 139-150.
3. R. García-Martínez y P. Britos (2004). *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería.
4. H. Flores (2009). *Detección de Patrones de Daños y Averías en la Industria Automotriz*. Tesis de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.
5. R. García-Martínez (1994). *Adquisición de Conocimiento*. En Abecasis, S. y Heras, C. Metodología de la Investigación. Nueva Librería. ISBN 950-9088-65-x.
6. A. Gomez, N. Jurista, C. Montes y J. Pazos (1997). *Ingeniería del Conocimiento*. Editorial R. Areces. Madrid.