

INGENIERÍA DE SOFTWARE E INGENIERÍA DEL CONOCIMIENTO: DOS DISCIPLINAS INTERRELACIONADAS



UNIVERSIDAD DE MEDELLÍN



INGENIERÍA DE SOFTWARE E INGENIERÍA DEL CONOCIMIENTO: DOS DISCIPLINAS INTERRELACIONADAS

1a. edición: 2014

© Universidad de Medellín

© Antonio A. Aguilera

© Bell Manrique Losada

© Carlos Mario Zapata Jaramillo

© Darío Rodríguez

© Demetrio A. Ovalle Carranza

© Edison Spina

© Edwin H. Hincapié-Corrales

© Fabio Alberto Vargas

© Germán Urrego-Giraldo

© Gerzon E. Gómez

© Gloria Liliana Vélez

© Gloria Lucía Giraldo

© Gloria Piedad Gasca

© Guillermo González-Calderón

© Héctor J. Ortiz Pabón

© Hernán Merlino

© Jaime Alberto Echeverri

© Javier M. Reyes Vera

© John Branch

© John W. Castro

© Jonás Montilva

© Jorge Eliécer Giraldo Plaza

© Jovani Alberto Jiménez Builes

© Juan Carlos Hernández

© Juan P. Ucán

© Judith Barrios

© Liliana González-Palacio

© Lillyana María Giraldo

© Lina María Giraldo

© Luis Joyanes

© Marcel J. Simonette

© María Clara Gómez

© Mauricio González-Palacio

© Mónica Tentori-Espinosa

© Omar S. Gómez

© Óscar Dieste

© Óscar H. Arenas-Arenas

© Óscar Mauricio Salazar

© Paola-J Rodríguez-C

© Patricia Pesado

© Ramón García-Martínez

© Raúl A. Aguilar

© Roberto Manjarrés

© Rodrigo Zalapa-Cardiel

© Sandra Mateus

© Sebastián Martins

© Silvia T. Acuña

© Vianca Vega

ISBN: 978-958-8815-31-2

Catalogación bibliográfica - Universidad de Medellín. Biblioteca Eduardo Fernández Botero. María Isabel Quintero Bedoya.

Editor:

Leonardo David López Escobar

Dirección electrónica: ldlopez@udem.edu.co

Universidad de Medellín. Medellín, Colombia

Cra. 87 No. 30-65. Bloque 20, piso 2.

Teléfonos: 340 52 42 - 340 53 35

Medellín - Colombia

Distribución y ventas:

Universidad de Medellín

e-mail: selloeditorial@udem.edu.co; www.udem.edu.co

Cra. 87 No. 30-65 / Teléfono: 340 52 42 - Medellín, Colombia

Corrección de estilo:

Diseño portada:

Claudia Castrillón Álvarez

claudiadiseno Grafico@gmail.com

Diagramación:

Hernán D. Durango T.

hernandedurango@gmail.com

Impresión: Xpress Estudio Gráfico y Digital S.A.
Av. Américas No. 39-53 / PBX (+57 1) 602 0808
- Bogotá, Colombia

Todos los derechos reservados.

Esta publicación no puede ser reproducida, ni en todo ni en parte, por ningún medio inventado o por inventarse, sin el permiso previo y por escrito de la Universidad de Medellín.

Hecho el depósito legal.

CAPÍTULO X

Derivación del proceso de explotación de información desde el dominio de negocio

*Sebastian Martins**

*Darío Rodríguez***

*Ramón García-Martínez****

INTRODUCCIÓN

La inteligencia de negocio propone un abordaje interdisciplinario dentro del que se encuentra la Informática, que toma todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento para generar, con base en estos recursos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [1].

La explotación de información es la sub-disciplina de los sistemas de información que aporta a la inteligencia de negocio las herramientas para la transformación de información en conocimiento [2]. Se define como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información.

Un Proceso de Explotación de Información se define como un grupo de tareas relacionadas lógicamente [3] que, a partir de un conjunto de informa-

* Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información Grupo Investigación en Sistemas de Información Universidad Nacional de Lanús Remedios de Escalada, Buenos Aires, Argentina. Dirección electrónica: martinssbastian@yahoo.com

** Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información Grupo Investigación en Sistemas de Información Universidad Nacional de Lanús Remedios de Escalada, Buenos Aires, Argentina. Dirección electrónica: ar, djhr_1977@yahoo.com.ar

*** Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información Grupo Investigación en Sistemas de Información Universidad Nacional de Lanús Remedios de Escalada, Buenos Aires, Argentina. Dirección electrónica: rgm1960@yahoo.com

ción con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del aprendizaje automático [4], susceptibles de utilizar en cada uno de estos procesos.

En [5] se proponen los siguientes cinco procesos de explotación de información: (a) descubrimiento de reglas de comportamiento, que aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema; (b) descubrimiento de grupos, que aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema; (c) descubrimiento de atributos significativos, que aplica cuando se requiere identificar cuáles son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema; (d) descubrimiento de reglas de pertenencia a grupos, que aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida "a priori", pero presente en la masa de información disponible sobre el dominio de problema; (e) ponderación de reglas de comportamiento o de pertenencia a grupos, que aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean éstas las que, en mayor medida, inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo.

En este Capítulo se aborda el problema de determinar el proceso de explotación de información requerido para descubrir la pieza de conocimiento (patrón o regularidad) que aporta a la resolución de un problema de inteligencia de negocio (véase la Sección 2), se propone una solución consistente en un procedimiento de derivación del proceso de explotación de información a partir del modelado del negocio y del problema de negocio (véase la Sección 3), se presenta una prueba de concepto del procedimiento presentado (véase la Sección 4) y se formulan algunas conclusiones preliminares (véase la Sección 5).

10.1 DELIMITACIÓN DEL PROBLEMA

Las metodologías de desarrollo de proyectos de explotación de información difieren de las usadas para los proyectos de desarrollo de software tradicionales. Las clásicas etapas de análisis, diseño, desarrollo, integración y pruebas no encajan con las etapas naturales de los procesos de desarrollo

de este tipo de proyectos. En particular, las metodologías de explotación de información destacan la importancia de la comprensión del negocio y la vinculación entre el problema de negocio con el patrón o regularidad a descubrir en la información a partir del proceso de explotación de información a utilizar. Sin embargo, es un problema abierto determinar de manera sistemática cual es el proceso de explotación de información que se debe aplicar para resolver el problema de negocio.

10.2 SOLUCIÓN PROPUESTA

La solución propuesta tiene dos componentes: la identificación de formalismos de representación de conocimiento que permitan abordar el modelado del dominio de negocio, el problema del negocio y el problema de explotación de información (véase la Sección 3.1) y un procedimiento que permite derivar del proceso de explotación de información a partir del modelado del dominio de negocio (véase la Sección 3.2).

10.2.1 Formalismos de Representación de Conocimiento para Modelado del Dominio

Los formalismos de ingeniería de conocimiento que se utilizan para modelar el dominio de negocio y de explotación de información son las redes semánticas [6] y las tablas concepto-atributo-valor [7].

Con base en los formalismos señalados, en este capítulo se proponen doce formalismos para el modelado del dominio:

Tabla concepto-atributo-valor del dominio (CAVD) se utiliza para la representación sistemática de los conceptos del dominio, con detalles de los atributos y valores.

Tabla concepto-relación del dominio (CRD) se utiliza para modelar las relaciones entre los conceptos del dominio.

Red semántica del modelo de negocio (RSMN) modela el dominio del negocio.

Tabla concepto-atributo-valor del problema de explotación de información (CAVPEI) es una representación sistemática de los conceptos del problema de explotación de información, con detalles de los atributos y valores.

Tabla concepto-relación del problema de explotación de información (CRPEI) busca modelar las relaciones entre los conceptos del problema de explotación de información.

Tabla de entradas/salidas del problema de explotación de información (E/S-PEI) es un listado de entradas/salidas del problema de explotación de información.

Red semántica del problema de explotación de información (RSPEI) es una representación sistemática del problema de explotación de información.

Red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI-DE/S) modela el problema de explotación de información con los datos identificados de entrada y salida.

Tabla concepto-atributo-valor Integrada (CAVI) es una representación sistemática integrada de los conceptos del dominio y del problema de explotación de información, con detalles de los atributos y valores.

Tabla concepto-relación Integrada (CRI) modela de manera integrada las relaciones entre los conceptos del dominio y del problema de explotación de información.

Red semántica integrada modelo negocio-explotación de información (RSI-MN-EI) es una representación sistemática integrada del modelo de negocio y del problema de explotación de información.

Red semántica integrada modelo negocio-explotación de información con datos de entrada y salida identificados (RSI-MN-EI-DE/S) es una representación sistemática integrada del modelo de negocio y del problema de explotación de información con los datos identificados de entrada y salida.

En el Capítulo se usa la siguiente notación: los conceptos se representan con óvalos, las relaciones entre conceptos con flechas marcadas (con el rótulo de la relación), los atributos y valores de conceptos se representan con óvalos segmentados, los atributos y valores de entrada se representa con doble óvalo segmentado, los atributos y valores de salida se representan con doble óvalo mixto (entero por dentro y segmentado por fuera) y los conceptos de salida con doble óvalo entero.

10.2.2 Procedimiento de Derivación del Proceso de Explotación de Información a Partir del Modelado del Dominio de Negocio

En esta Sección se propone un procedimiento para derivar el proceso de explotación de información a partir del modelado del dominio de negocio

con los formalismos introducidos en la Sección 3.1. El procedimiento consta de diez pasos y se presenta en la Tabla 1.

Tabla 1. Procedimiento de derivación del proceso de explotación de información a partir del modelado del dominio de negocio

<i>Procedimiento de derivacion</i>	
Paso 1:	Identificación de conceptos del dominio.
Paso 2:	Identificación de relaciones entre conceptos.
Paso 3:	Conceptualización del dominio.
Paso 4:	Identificación de conceptos y relaciones del problema de explotación de información.
Paso 5:	Conceptualización del problema de explotación de información.
Paso 6:	Identificación de los datos de entradas y resultados del problema de explotación de información.
Paso 7:	Integración de las novaciones que introduce el problema de explotación de información al dominio.
Paso 8:	Integración de los conceptos del dominio y del problema de explotación de información.
Paso 9:	Identificación Integrada de los datos de entrada y objetivo.
Paso 10:	Derivación del proceso de explotación de información.

El procedimiento conceptualiza el dominio a partir de la identificación de los conceptos del dominio y las relaciones entre estos conceptos (pasos 1 a 3). Después conceptualiza el problema de explotación de información con base en la identificación de conceptos y relaciones del problema de explotación de información y la identificación de los datos de entradas y resultados del problema de explotación de información (pasos 4 a 6). En la representación lograda hasta el momento se producen novaciones que introduce el problema de explotación de información al dominio, las cuales se incorporan a la representación del dominio. En este punto se pueden integrar los conceptos del dominio y del problema de explotación de información; e identificar de manera integral los datos de entrada y objetivo (pasos 7 a 9). A partir de este punto, y con base en los subpasos del paso 10 que se presentan en la Tabla 2, se deriva el proceso de explotación de información.

Tabla 2. Detalle de los subpasos de derivación del proceso de explotación de información

<i>Detalle del paso 10: derivación del proceso de explotación de información.</i>		
Subpaso 10.1:	SI	En la Red semántica integrada modelo negocio-explotación de información con datos de entrada y salida identificados (RSI-MN-EI-DE/S) se puede identificar un resultado intermedio.
	ENTONCES	Ir a Subpaso 10.5.
	SINO	Ir a Subpaso 10.2.
Subpaso 10.2:	SI	El objetivo de la red semántica integrada modelo negocio-explotación de información con datos de entrada y salida identificados (RSI-MN-EI-DE/S) es un único nodo cuya arista se identifica con las palabras: pertenece, integra, es, es miembro de, u otras palabras de similar significado.
	ENTONCES	Aplicar proceso descubrimiento de grupos.
	SINO	Ir a Subpaso 10.3.
Subpaso 10.3:	SI	En la red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI-DE/S) se busca detectar los valores que más se destacan o que aparecen con mayor frecuencia del conjunto de datos de salida
	ENTONCES	Aplicar proceso ponderación de interdependencia de atributos.
	SINO	Ir a Subpaso 10.4.
Subpaso 10.4:	SI	En la red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI-DE/S) se busca identificar, a partir del conjunto de datos de salida, el valor de un atributo determinado
	ENTONCES	Aplicar proceso descubrimiento de reglas de comportamiento.
	SINO	Representación errónea.
Subpaso 10.5:	SI	La Red semántica integrada modelo negocio-explotación de información con datos de entrada y salida identificados (RSI-MN-EI-DE/S) tiene un concepto resultado intermedio, para el cual el arco que llega se rotula con las palabras: "pertenece", "integra", "es", "es miembro de" u otras palabras de similar significado
	ENTONCES	Ir a Subpaso 10.6.
	SINO	Representación errónea.

<i>Detalle del paso 10: derivación del proceso de explotación de información.</i>		
Subpaso 10.6:	SI	El resultado intermedio de la red semántica integrada modelo negocio-explotación de información con datos de entrada y salida identificados (RSI-MN-EI-DE/S), cuya relación se define con las palabras pertenece, integra, es, es miembro de, u otras palabras de similar significado, es un nodo existente en la red semántica del modelo de negocio (RSMN).
	ENTONCES	Aplicar proceso ponderación de reglas de pertenencia a grupos (grupos definidos).
	SINO	Ir a Subpaso 10.7.
Subpaso 10.7:	SI	En la Red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI-DE/S) se busca detectar aquellos valores de los atributos pertenecientes al concepto identificado como resultado intermedio, que más se destacan o que aparecen con mayor frecuencia en el conjunto de datos de salida.
	ENTONCES	Aplicar proceso ponderación de reglas de pertenencia a grupos (grupos no definidos).
	SINO	Ir a Subpaso 10.8.
Subpaso 10.8:	SI	En la red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI-DE/S) se busca definir el identificador del resultado intermedio, de acuerdo con los atributos de salida
	ENTONCES	Aplicar proceso descubrimiento de reglas de pertenencia a grupos.
	SINO	Representación errónea.

En la Tabla 3 se presenta un cuadro resumen del procedimiento con mención de las entradas y salidas, y un detalle de los formalismos utilizados en cada caso.

Tabla 3.a. Entradas/Salidas y formalismos de representación del procedimiento propuesto

	Input		Output	
	Entrada	Representación	Salida	Representación
Paso 1:	Identificación de conceptos del dominio	Descripción del dominio del negocio Datos del negocio	Texto plano	Tabla Concepto - Atributo - Valor del Dominio (CAVD)
Paso 2:	Identificación de relaciones entre conceptos	Descripción del dominio del negocio Representación sistemática de los Conceptos del dominio, con detalles de los Atributos y Valores	Texto plano o Tabla CAVD	Tabla Concepto - Relación del Dominio (CRD)
Paso 3:	Conceptualización del dominio	Representación sistemática de los Conceptos del dominio, con detalles de los Atributos y Valores Representación sistemática de las relaciones entre los Conceptos del dominio	Tabla CAVD Tabla CRD	Red semántica del modelo de negocio (RSMN)
Paso 4:	Identificación de relaciones del Problema de Explotación de Información	Problema de explotación de información Datos del negocio	Texto plano	Tabla Concepto - Atributo - Valor del Problema de Explotación de Información (CAVPEI) Tabla Concepto - Relación del Problema de Explotación de Información (CRPEI) Tabla E/S-PEI
Paso 5:	Conceptualización del Problema de Explotación de Información	Problema de explotación de los Conceptos del dominio, con detalles de los Atributos y Valores Representación sistemática de las relaciones entre los Conceptos del Problema de Explotación de Información	Tabla CAVPEI Tabla CRPEI	Red semántica del problema de explotación de información (RSPEI)
Paso 6:	Identificación de los datos de entradas y resultados del problema de explotación de información	Representación sistemática del problema de explotación de información Listado de entradas/salidas del problema de explotación de información	R S P E I - T a b l a E/S-PEI	Red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI-DE/S)

Tabla 3.b. Entradas/Salidas y formalismos de representación del procedimiento propuesto

		Input		Output	
		Entrada	Representación	Salida	Representación
paso 7:	integración de las innovaciones introducidas por el problema de explotación al dominio	Representación sistemática de los Conceptos del dominio, con detalles de los Atributos y Valores Representación sistemática de las relaciones entre los Conceptos del dominio Representación sistemática de los Conceptos del Problema de Explotación de Información, con detalles de los Atributos y Valores Representación sistemática de las relaciones entre los Conceptos del Problema de Explotación de Información	Tabla CAVD Tabla CRD Tabla CAVPEI Tabla CRPEI	Representación sistemática integrada de los Conceptos del dominio y del Problema de Explotación de Información, con detalles de los Atributos y Valores Representación sistemática integrada de las relaciones entre los Conceptos del dominio y del Problema de Explotación de Información	Tabla Concepto -Atributo - Valor Integrada (CAVI) Tabla concepto-relación Integrada (CRI)
Paso 8:	Integración de los conceptos del dominio y del problema de explotación de información	Representación sistemática integrada de los Conceptos del dominio y del Problema de Explotación de Información, con detalles de los Atributos y Valores Representación sistemática integrada de las relaciones entre los Conceptos del dominio y del Problema de Explotación de Información	Tabla CAVI Tabla CRI	Representación sistemática integrada del modelo de negocio y del problema de explotación de información	Red Semántica Integrada Modelo Negocio – Explotación de Información (RSI-MN-EI)
Paso 9:	Identificación Integrada de los datos de entrada y objetivo	Representación sistemática integrada del modelo de negocio y del problema de explotación de información Representación sistemática del problema de explotación de información con sus datos de entrada y salida identificados	RSI-MN-EI RSPEI-DE/S	Representación sistemática integrada del modelo de negocio y del problema de explotación de información con sus datos de entrada y salida identificados	Red Semántica Integrada Problema Negocio – Explotación de Información con datos de entrada y salida identificados (RSI-PN-EI-DE/S)
Paso 10:	Derivación del proceso de explotación de información	Representación sistemática integrada del modelo de negocio y del problema de explotación de información con sus datos de entrada y salida identificados Representación sistemática del modelo de negocio Representación sistemática del problema de explotación de información con sus datos de entrada y salida identificados	RSI-PN-EI-DE/S RSMN RSPEI-DE/S	Proceso de explotación de información	Texto plano

10.3 PRUEBA DE CONCEPTO

En esta sección se presenta una prueba de concepto sobre el procedimiento de derivación del proceso de explotación de información a partir del modelado del dominio de negocio. Se toma el siguiente caso [8]:

"...Descripción del Negocio y Problema de Negocio:

"Servicios Móviles Argentina S.A." es una empresa perteneciente al sector de las telecomunicaciones que opera en todo el país. Ofrece productos de diversas marcas comerciales y genéricas, servicios y una amplia variedad de planes tarifarios (de uso personal o corporativo) de acuerdo con las necesidades de los clientes. Actualmente, está impulsando campañas de retención de clientes personalizadas en todo el país con el objetivo de ofrecer nuevos productos y servicios acordes con las características de cada cliente, para adelantarse a sus competidores, que también están creciendo a nivel nacional. El objetivo de negocio es caracterizar a los clientes en las diferentes regiones geográficas de cobertura de la empresa, con el fin de facilitar la definición de campañas de marketing orientadas al mantenimiento de los clientes actuales y a la obtención de nuevos clientes según las preferencias comunes de cada región, lo que ayudará también a mejorar la rentabilidad de la empresa. Con esto, se busca también determinar el comportamiento de sus clientes y mejorar el conocimiento que se tiene de ellos. "Servicios Móviles Argentina S.A." tiene como meta la satisfacción de sus clientes, lo cual realiza mediante el ofrecimiento de productos y servicios de calidad, dividiendo todas las operaciones de telefonía móvil del país en cinco regiones: Litoral, Cuyo, Pampeana/Centro, Patagonia y Norte.

Descripción de los datos:

Los siguientes son los datos relevantes que almacena la empresa durante el ejercicio de sus actividades:

- *Código del Cliente: que la empresa utiliza para identificarlo. Ejemplo: 99861598.*
- *Situación del Cliente: ejemplo: AAA (cliente activo), BA (cliente de baja).*
- *Fecha de Alta: pertenece al cliente. Ejemplo: 10/03/2003.*
- *Código de la localidad: para la residencia del cliente. Ejemplo: 379.*
- *Código de la provincia: al cual pertenece la localidad. Ejemplo: 3.*
- *Plan Contratado: código de plan tarifario que contrata el cliente. Ejemplo: U21.*

- *Duración: cantidad de meses de contratación del cliente. Ejemplo: 12.*
- *Tipo de Cliente: ejemplo: PR/PO (cliente prepago), FF (cliente con facturación).*
- *Marca del Equipo: perteneciente al cliente. Ejemplo: NOKIA.*
- *Región geográfica: indica la región a la que pertenece cada provincia (véase la Tabla 4)...”*

Tabla 4. Código de regiones por provincias

<i>Región</i>	<i>Descripción</i>	<i>Provincias integrantes</i>
1	Litoral	Misiones, Corrientes, Formosa, Chaco, Santa Fe
2	Cuyo	Mendoza, San Luis, San Juan
3	Pampeana/ Centro	Buenos Aires, Capital Federal, Córdoba, La Pampa
4	Patagonia	Neuquén, Río Negro, Chubut, Santa Cruz, Tierra del Fuego
5	Norte	Jujuy, Salta, Tucumán, Catamarca, La Rioja, Santiago del Estero

10.3.1 Paso 1: Identificación de Conceptos del Dominio

Input:	Descripción del dominio del negocio. Datos del negocio.
Output:	Representación sistemática de los conceptos del dominio, con detalle de los atributos y valores (véase la Tabla 5).

10.3.2 Paso 2: Identificación de Relaciones entre Conceptos

Input: Descripción del dominio del negocio.

Representación sistemática de los conceptos del dominio, con detalle de los atributos y valores (véase la Tabla 5).

Tabla 5. Tabla concepto-atributo-valor del dominio (CAVD)

<i>Concepto</i>	<i>Atributo</i>	<i>Valor</i>
Cliente	código	Numérico
	situación	Alfabético EJ: AAA, BA.
	fecha alta	dd/mm/AAAA
	provincia	Numérico EJ: 379
	localidad	Numérico EJ: 3

Concepto	Atributo	Valor
Cliente	región	1 al 5
	tipo	Alfabético EJ: PR/PO, FF.
	marca equipo	Alfabético EJ: Nokia
Plan	código	Alfanumérico EJ: U21.
	duración	Numérico

Output: Representación sistemática de las relaciones entre los conceptos del dominio (véase la Tabla 6).

Tabla 6. Tabla concepto-relación del dominio (CRD)

Conceptos	Conceptos Asociados	Relación	Descripción
cliente	plan	contrata	El cliente contrata un plan por un tiempo determinado.

10.3.3 Paso 3: Conceptualización del Dominio

Representación sistemática de los conceptos del dominio, con detalle de los atributos y valores (véase la Tabla 5).

Input:

Representación sistemática de las relaciones entre los conceptos del dominio (véase la Tabla 6).

Output:

Representación sistemática del modelo de negocio (véase la Figura 1).

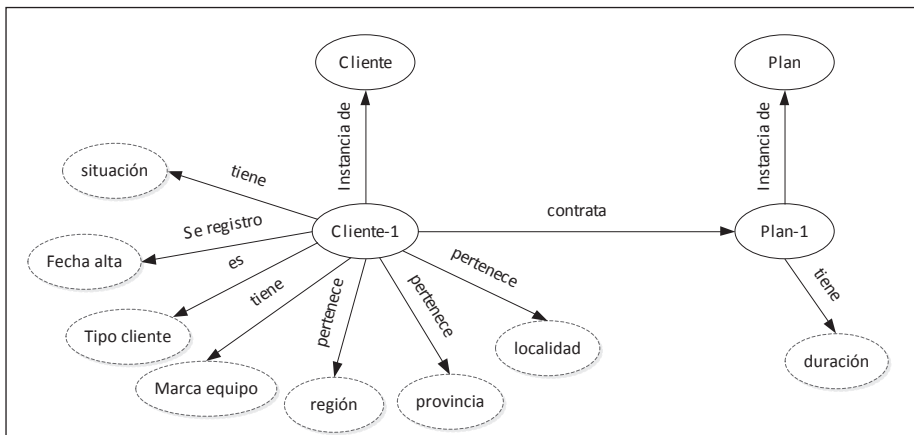


Fig. 1. Red semántica del modelo de negocio (RSMN)

10.3.4 Paso 4: Identificación de Conceptos y Relaciones del Problema de Explotación de Información

El objetivo correspondiente al problema de explotación de información es: “determinar reglas que identifiquen el comportamiento de un conjunto de clientes de acuerdo con su ubicación geográfica (región, provincia, localidad)”.

- Input: Problema de explotación de información.
 Datos del negocio.
 Representación sistemática de los conceptos del problema de explotación de información, con detalle de los atributos y valores (véase la Tabla 7).
- Output: Representación sistemática de las relaciones entre los conceptos del problema de explotación de información (véase la Tabla 8).
 Listado de entradas/salidas del problema de explotación de información (véase la Tabla 9).

Tabla 7. Tabla concepto-atributo-valor del problema de explotación de información (CAVPEI)

<i>Concepto</i>	<i>Atributo</i>	<i>Valor</i>
Cliente	provincia	Numérico EJ: 379
	localidad	Numérico EJ: 3
	región	1 al 5
Plan	código	Alfanumérico EJ: U21.
	duración	Numérico
Grupo	situación	Alfabético EJ: AAA, BA.
	tipo	Alfabético EJ: PR/PO, FF.
	marca equipo	Alfabético EJ: Nokia

Tabla 8. Tabla E/S-PEI

	<i>Tipo</i>	<i>Nombre</i>
ENTRADA	Atributo	Cliente.provincia
	Atributo	Cliente.localidad
	Atributo	Cliente.región
SALIDA	Concepto	Grupo
	Atributo	Grupo.situación
	Atributo	Grupo.tipo

	<i>Tipo</i>	<i>Nombre</i>
	Atributo	Grupo.Marca_equipo
	Concepto	Grupo.Plan
	Atributo	Grupo.Plan.duración

Tabla 9. Tabla Concepto-Relación del Problema del Explotación de Información (CRPEI)

<i>Concepto</i>	<i>Conceptos asociados</i>	<i>Relación</i>	<i>Descripción</i>
cliente	plan	contrata	el cliente contrata un plan por un tiempo determinado.
Cliente	grupo	pertenece	el cliente pertenece a un grupo de acuerdo con sus características.
Grupo	plan	tiene	el grupo tiene un plan dentro de sus características comunes.

10.3.4 Paso 5: Conceptualización del Problema de Explotación de Información

Representación sistemática de los conceptos del problema de explotación de información, con detalle de los atributos y valores (véase la Tabla 7).

Input:

Representación sistemática de las relaciones entre los conceptos del problema de explotación de información (véase la Tabla 8).

Output:

Representación sistemática del problema de explotación de información (véase la Figura 2).

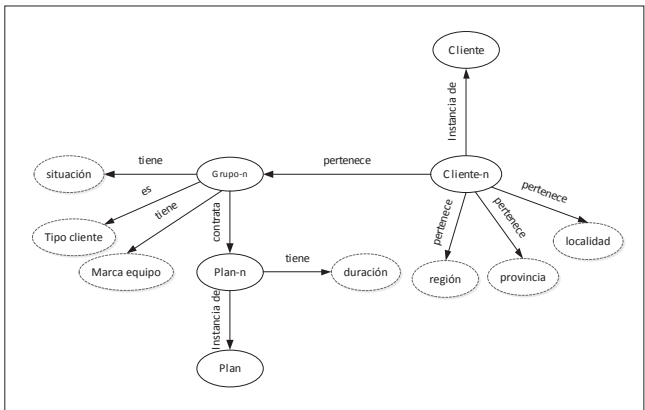


Fig. 2. Red semántica del problema de explotación de información (RSPEI)

10.3.5 Paso 6: Identificación de Datos de Entrada y Resultados del Problema de Explotación de Información

Input:	Representación sistemática del problema de explotación de información (véase la Figura 2). Listado de entradas/salidas del problema de explotación de información (véase la Tabla 9).
Output:	Representación sistemática del problema de explotación de información con sus datos de entrada y salida identificados (véase la Figura 3).
Nota:	Obsérvese que la Figura 3 difiere de la Figura 2, ya que se marcaron en ella los conceptos, atributos y valores de entrada/salida.

10.3.6 Paso 7: Integración de las Novaciones Introducidas por el Problema de Explotación de Información al Dominio

Input:	Representación sistemática de los conceptos del dominio, con detalle de los atributos y valores (véase la Tabla 5). Representación sistemática de las relaciones entre los conceptos del dominio (véase la Tabla 6). Representación sistemática de los conceptos del problema de explotación de información, con detalle de los atributos y valores (véase la Tabla 7). Representación sistemática de las relaciones entre los conceptos del problema de explotación de información (véase la Tabla 8).
Output:	Representación sistemática integrada de los conceptos del dominio y del problema de explotación de información, con detalle de los atributos y valores (véase la Tabla 10). Representación sistemática integrada de las relaciones entre los conceptos del dominio y del problema de explotación de información (véase la Tabla 11).

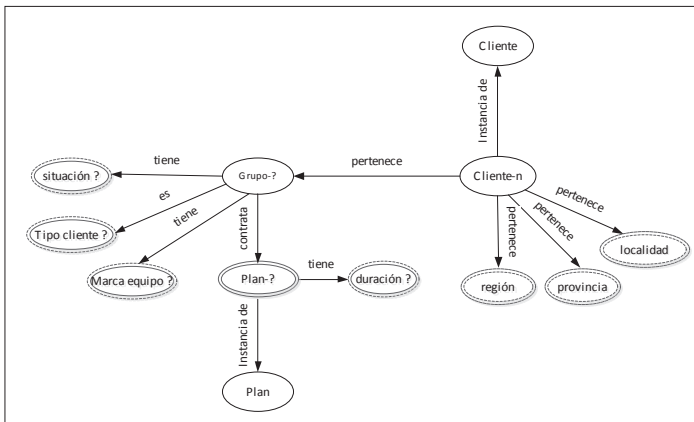


Fig. 3. Red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI-DE/S)

Tabla 10. Tabla Concepto-Atributo-Valor Integrada (CAVI)

<i>Concepto</i>	<i>Atributo</i>	<i>Valor</i>
Cliente	Código	Numérico
	situación	Alfabético EJ: AAA, BA.
	fecha Alta	dd/mm/AAAA
	provincia	Numérico EJ: 379
	localidad	Numérico EJ: 3
	región	1 al 5
	tipo	Alfabético EJ: PR/PO, FF.
	Marca equipo	Alfabético EJ: Nokia
Plan	Código	Alfanumérico EJ: U21.
	duración	Numérico
Grupo	situación	Alfabético EJ: AAA, BA.
	tipo	Alfabético EJ: PR/PO,FF.
	Marca equipo	Alfabético EJ: Nokia

Tabla 11. Tabla concepto-relación Integrada (CRI)

<i>Conceptos</i>	<i>Conceptos Asociados</i>	<i>Relación</i>	<i>Descripción</i>
cliente	plan	contrata	El cliente contrata un plan por un tiempo determinado.
cliente	grupo	pertenece	El cliente pertenece a un grupo de acuerdo a sus características.
grupo	plan	tiene	El grupo tiene un plan dentro de sus características comunes.

10.3.7 Paso 8: Integración de Conceptos del dominio y del Problema de Explotación de Información

Representación sistemática integrada de los conceptos del dominio y del problema de explotación de información, con detalle de los atributos y valores (véase la Tabla 10).

Input:

Representación sistemática integrada de las relaciones entre los concepto del dominio y del problema de explotación de información (véase la Tabla 11).

Output:

Representación sistemática integrada del modelo de negocio y del problema de explotación de información (véase la Figura 4).

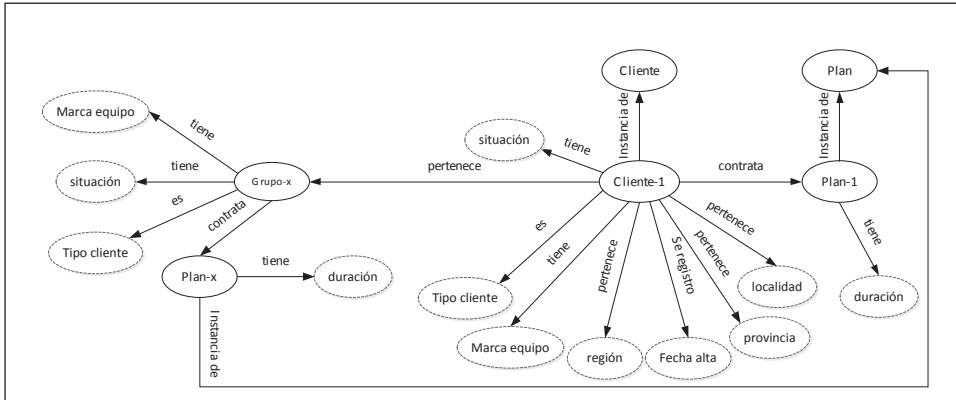


Fig. 4. Red Semántica Integrada modelo negocio-explotación de información. (RSI-MN-EI)

10.3.8 Paso 9: Identificación Integrada de los Datos de Entrada y Objetivo.

Representación sistemática integrada del modelo de negocio y del problema de explotación de información (véase la Figura 4).

Input:

Representación sistemática del problema de explotación de información con sus datos de entrada y salida identificados (véase la Figura 3).

Output:

Representación sistemática integrada del modelo de negocio y del problema de explotación de información con sus datos de entrada y salida identificados (véase la Figura 5).

Nota:

Observar que la Figura 5 difiere de la Figura 4 ya que se marcaron en ella los conceptos, atributos y valores de entrada/salida.

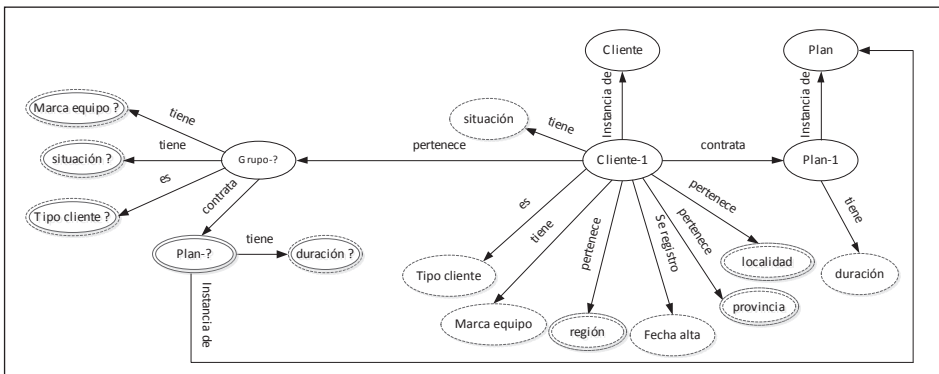


Fig. 5. Red semántica integrada modelo negocio-explotación de información con datos de entrada y salida identificados (RSI-MN-EI-DE/S)

10.3.9 Paso 10: Derivación del Proceso de Explotación de Información.

Input:	<p>Representación sistemática integrada del modelo de negocio y del problema de explotación de información con sus datos de entrada y salida identificados (véase la Figura 5).</p> <p>Representación sistemática del modelo de negocio (véase la Figura 1).</p> <p>Representación sistemática del problema de explotación de información con sus datos de entrada y salida identificados (véase la Figura 3).</p>
Output:	<p>Determinación del proceso de explotación de información:</p> <p>Por ejecución del subpaso 10.1 del procedimiento propuesto dado que en la red semántica integrada modelo negocio-explotación de información con datos de entrada y salida identificados (RSI-MN-EI-DE/S) se puede identificar un resultado intermedio, se ejecuta el subpaso 10.5; dado que en la red semántica integrada modelo negocio-explotación de información con datos de entrada y salida identificados (RSI-MN-EI-DE/S) tiene un concepto resultado intermedio, para el que el arco que llega se rotula con las palabras: "pertenece", "integra", "es", "es miembro de" u otras palabras de similar significado, se ejecuta el subpaso 10.6; como no aplica el proceso de ponderación de reglas de pertenencia a Grupos (definidos) se ejecuta el subpaso 10.7; como en la red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI-DE/S) no se detectan aquellos valores de los atributos pertenecientes al concepto identificado como resultado intermedio, que más se destacan o que aparecen con mayor frecuencia en el conjunto de datos de salida se ejecuta el Subpaso 10.8; este paso propone utilizar el proceso descubrimiento de reglas de pertenencia a grupos.</p>

10.4 CONCLUSIONES

Este Capítulo partió de la premisa de que las herramientas involucradas en los procesos de ingeniería de software no se aplican a proyectos de explotación de información. Los autores pertenecen a un grupo de investigación que desarrolla desde hace diez años un cuerpo de conocimientos para una Ingeniería de Explotación de Información, tratando de contribuir a mejorar la productividad de las PYMEs (pequeñas y medianas empresas) del área.

En este contexto, se introdujeron en el capítulo formalismos basados en técnicas de representación de conocimiento (redes semánticas y tablas concepto-atributo-valor) para modelar el dominio de negocio y de explotación de información. En línea con la tradición de Ingeniería de Conocimiento, el proceso de transformación que conlleva la formalización, permite hacer explícitos conceptos (y relaciones entre estos) que se encuentran en la descripción del dominio y del problema de negocio. Con base en la explicitación de los conceptos y de las relaciones, se presentó un procedimiento

que permite derivar el proceso de explotación de información a partir de los formalismos propuestos. Se desarrolló una prueba de concepto que permite ilustrar la aplicación del procedimiento propuesto. Como futura línea de trabajo, se prevé desarrollar la validación del procedimiento propuesto en casos de estudio de proyectos de explotación de información seleccionados.

10.5 FINANCIAMIENTO

Las investigaciones que se reportan en este capítulo han sido financiadas por Beca BENTR13 de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (Argentina), Beca UNLa-SCyT-33A105 y parcialmente por el Proyecto de Investigación 33A167 de la Secretaria de Ciencia y Técnica de la Universidad Nacional de Lanús (Argentina).

REFERENCIAS BIBLIOGRÁFICAS

1. Negash, S., Gray, P.: Business Intelligence. En Handbook on Decision Support Systems 2, ed. F. Burstein y C. Holsapple Heidelberg, Springer, Pág. 175-193 (2008)
2. García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J.: Towards an Information Mining Engineering. En Software Engineering, Methods, Modeling and Teaching. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Páginas 83-99 (2011)
3. García-Martínez, R., Britos, P., Rodríguez, D.: *Information Mining Processes Based on Intelligent Systems*. Lecture Notes on Artificial Intelligence, 7906: 402-410. ISBN 978-3-642-38576-6 (2013)
4. García Martínez, R., Servente, M. y Pasquini, D.: *Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7. (2003)
5. Britos, P. *Procesos de Explotación de Información Basados en Sistemas Inteligentes*. Tesis de Doctorado en Ciencias Informáticas. Universidad Nacional de La Plata (2008)
6. Sowa, J.: *Semantic Networks*. En Encyclopedia of Artificial intelligence (Editor S. Shapiro, 2da edición). Wiley & Sons. ISBN 978-0471503071 (1992)
7. García Martínez, R. y Britos, P.: *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4 (2004)
8. Basso, D. y Vegega, C.: *Caso de Estudio: Telefonía Celular*. Trabajo Final del Seminario de Inteligencia de Negocio. Maestría en Ingeniería de Sistemas de Información. Escuela de Posgrado. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina (2013)