

Derivación del Proceso de Explotación de Información desde el Modelado del Negocio

Sebastian Martins

Grupo de Investigación en Sistemas de Información (UNLa GISI)
Departamento de Desarrollo Productivo y Tecnológico
Universidad Nacional de Lanús
Remedios de Escalada, Buenos Aires, Argentina.
martinssebastian@yahoo.com.ar

Resumen—Existen metodologías de explotación de información que destacan la importancia del planeamiento de una elicitación de requerimientos a lo largo de todo el proyecto de una manera ordenada, documentada, consistente y trazable. Sin embargo, dadas las características de este tipo de proyectos, el abordaje que propone la ingeniería de requerimientos clásica no es aplicable al proceso de identificar el problema de explotación de información, ni inferir el proceso que lo resuelve a partir del modelado del dominio del negocio. En este trabajo se propone un procedimiento basado en formalismos de Ingeniería del Conocimiento que permite derivar el proceso de explotación de información a partir de las representaciones del dominio de negocio y del problema de explotación de información.

Palabras Claves — Ingeniería de Explotación de Información, Procesos de explotación de información, Derivación de procesos.

I. INTRODUCCIÓN

En esta sección se presenta el marco del artículo (sección I.A), se da una delimitación del problema (sección I.B), se plantean los elementos de la solución propuesta (sección I.C), y se da una visión general del proyecto (sección I.D).

A. Marco Del Artículo

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 información 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.

B. 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.

C. 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 y del problema de explotación de información, y un procedimiento que permita derivar el proceso de explotación de información a partir del modelado del dominio de negocio y del problema de explotación de información.

D. Visión General

En la sección I (Introducción), se presenta el marco del artículo, se da una delimitación del problema, se plantean los elementos de la solución propuesta, y se da una visión general del proyecto.

En la sección II (Estado de la Cuestión), se presentan los marcos teóricos de la explotación de información y de la ingeniería del conocimiento de interés para esta línea de investigación.

En la sección III (Descripción del Problema), se identifica el problema de investigación, en el cual se resalta la inexistencia de un mecanismo sistematizado para la determinación del proceso de explotación de información a aplicar, se define el problema abierto y se concluye con un resumen de investigación.

En el capítulo IV (Solución), se presenta un proceso de derivación de modelos, en el cual se abordan las cuestiones generales de mayor relevancia, se describe la estructura general del proceso propuesto. Se introducen los estándares definidos para el modelado de los distintos conocimientos, en las que se identifican las notaciones generales y específicas de los formalismos basados en redes semánticas y se presenta un resumen de los mismos. Por último, se presenta una visión detallada del proceso de derivación de modelos propuesto, en la cual se describen los aspectos generales del proceso y el caso de estudio mediante el cual se ilustrará al mismo, se definen las fases que componen al proceso, se identifican las etapas que componen cada una de las fases, se señalan las técnicas y formalismos que se utilizan para el desarrollo de cada fase y se brinda un resumen del proceso.

En el Capítulo V (Conclusiones), se presentan las aportaciones del artículo y se destacan las futuras líneas de investigación que se consideran de interés en base al problema abierto que se presenta en este trabajo.

II. ESTADO DE LA CUESTIÓN

En esta sección se presenta el estado de la cuestión sobre distintas teorías y técnicas que convergen con los objetivos del presente trabajo. Se presenta el marco teórico de la explotación de información de interés para el artículo (sección II.A), y el marco teórico de la ingeniería del conocimiento de interés para esta línea de investigación (sección II.B).

A. Marco Teórico De La Explotación De Información De Interés

El marco teórico de la explotación de información de interés para el artículo, se estructura en base al marco teórico que se presenta en [5], identificando los siguientes conceptos: explotación de información (sección II.A.1), inteligencia de negocios (sección II.A.2), problema de negocio y de explotación de información (sección II.A.3) y procesos de explotación de información (sección II.A.4).

1) Explotación de Información

La explotación de información se ha definido, aunque utilizando el termino Minería de datos en su lugar, como la búsqueda de patrones interesantes y de reglas importantes en grandes masas de información [6;7]. También fue definida como el proceso de descubrir nuevas correlaciones, patrones y tendencias significativas utilizando grandes cantidades de datos almacenados en repositorios, usando tecnologías de reconocimiento de patrones, así como técnicas matemáticas y de estadística [8].

La Ingeniería de Explotación de Información estudia los procesos de extracción de conocimiento no trivial que recibe de manera implícita en los datos que se almacenan en grandes bases de datos. Estos procesos se valen de algoritmos de Minería de Datos (DM, Data Mining) [2]. Dicho conocimiento es previamente desconocido y puede resultar útil para algún proceso [9]. Para un experto, o para el responsable de un sistema, normalmente no son los datos en sí lo más relevante, sino el conocimiento oculto presente en sus relaciones, fluctuaciones y dependencias.

2) Inteligencia de Negocios

Se denomina inteligencia de negocio (Business Intelligence, BI) al conjunto de estrategias y herramientas [10] enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización [11; 12; 13; 14; 15].

Involucra el uso de los datos de una organización para facilitarle a las personas que realizan la toma de decisiones estratégicas del negocio, la comprensión del funcionamiento actual y la anticipación de acciones para dar una dirección bien informada a la organización [16;17; 18].

El sistema de información que se utiliza para la inteligencia de negocios integra diferentes datos extraídos del área de producción, con información relacionada con la organización o sus ámbitos y con datos económicos [19].

Las herramientas de inteligencia de negocios son software de aplicación diseñados para colaborar con la inteligencia de negocios en los procesos de las organizaciones [20].

Específicamente se trata de herramientas que asisten el análisis y la presentación de los datos [21; 22; 23; 24; 25].

Este conjunto de herramientas y metodologías tienen en común las siguientes características [26]:

- Accesibilidad a la información: Los datos son la fuente principal de este concepto. Lo primero que deben garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos, con independencia de la procedencia de éstos.
- Apoyo en la toma de decisiones: Se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- Orientación al usuario final: Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.
- De acuerdo a su nivel de complejidad, las herramientas de inteligencia de negocios se pueden clasificar en [27; 28]:
- Cuadro de Mando Integral (Balance Scorecard / Dashboard): También conocidos como Business Intelligence Dashboards, o Dashboards Ejecutivos. Son resúmenes visuales de información de la organización, que pretende mostrar de una mirada la comprensión global de las condiciones de la organización mediante métricas e indicadores clave de desempeño (KPIs). Esta es una herramienta de inteligencia de negocios en uso desde hace unos pocos años.
- Procesamiento Analítico en Línea (OLAP / On Line Analytic Processing): Es la capacidad de algunos sistemas de soporte de decisiones gerenciales, que permiten examinar de manera interactiva grandes volúmenes de información desde varias perspectivas.

- Aplicaciones de Reportes: Genera vistas de datos agregadas para mantener a la gestión informada sobre el estado de la organización.
- Explotación de Información: Extracción de información de las bases de datos acerca del objeto observado, mediante la utilización de aplicaciones que pueden aislar e identificar patrones o tendencias del objeto observado en un alto volumen de datos. Hay una gran variedad de técnicas de Explotación de Información que revelan distintos tipos de patrones.

3) Problemas de Negocio y de Explotación de Información

Para lograr el éxito de un proyecto de explotación de información es necesario comprender lo que el cliente desea lograr, para ello se deben entender los problemas de negocio que se presentan.

Los problemas de negocio describen los principales objetivos del cliente desde una perspectiva del negocio, además de otras cuestiones relacionadas con el objetivo principal del negocio que el cliente desea abordar [29].

Los problemas de negocio manifiestan los objetivos en la terminología del negocio. Los problemas de explotación de información representan los objetivos del proyecto en términos técnicos. La respuesta esperada obtenida del problema de explotación de información permite lograr los objetivos del negocio [29].

4) Procesos de Explotación de Información

Los procesos de explotación de información definen las técnicas o algoritmos a utilizar en base a las características del problema de explotación. En [5] se presentan los siguientes procesos de explotación de información: descubrimiento de reglas de comportamiento (Sección II.A.4.a), descubrimiento de grupos (Sección II.A.4.b), descubrimiento de atributos significativos (Sección II.A.4.c), descubrimiento de reglas de pertenencia a grupos (Sección II.A.4.d) y ponderación de reglas de comportamiento o de pertenencia (Sección II.A.4.e).

a) Procesos de Explotación de Información

En [30] se define que el proceso de descubrimiento de reglas de comportamiento aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema. Son ejemplos de problemas que requieren este proceso: identificación de características del local más visitado por los clientes, identificación de factores que inciden en el alza las ventas de un producto dado, establecimiento de características o rasgos de los clientes con alto grado de fidelidad a la marca, establecimiento de atributos demográficos y psicográficos que distinguen a los visitantes de un website, entre otros. Para el descubrimiento de reglas de comportamiento definidos a partir de atributos clases en un dominio de problema que representa la masa de información disponible, se propone la utilización de algoritmos de inducción TDIDT para descubrir las reglas de comportamiento de cada atributo clase. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 1. Como resultado de la aplicación del algoritmo de inducción TDIDT al atributo clase se obtiene un conjunto de reglas que definen el comportamiento de dicha clase.

b) Descubrimiento de Grupos

En [30] se define que el proceso de descubrimiento de grupos aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema. Son ejemplos de problemas que requieren este proceso: identificación de segmentos de clientes para bancos y

financieras, identificación de tipos de llamadas de clientes para empresas de telecomunicación, identificación de grupos sociales con las mismas características, identificación de grupos de estudiantes con características homogéneas, entre otros.

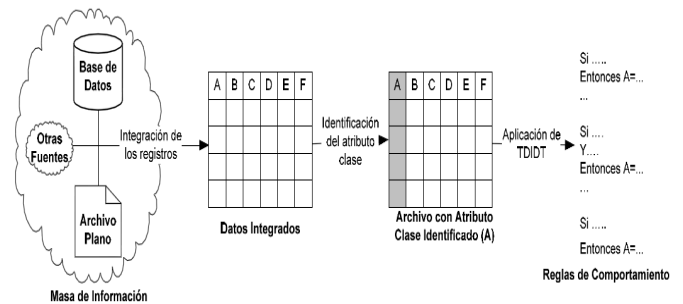


Fig. 1. Proceso de explotación de información: descubrimiento de reglas de comportamiento [30]

Para el descubrimiento de grupos a partir de masas de información del dominio de problema sobre las que no se dispone ningún criterio de agrupamiento “a priori” se propone la utilización de Mapas Auto Organizados de Kohonen o SOM por su sigla en inglés. El uso de esta tecnología busca descubrir si existen grupos que permitan una partición representativa del dominio de problema que la masa de información disponible representa. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 2.

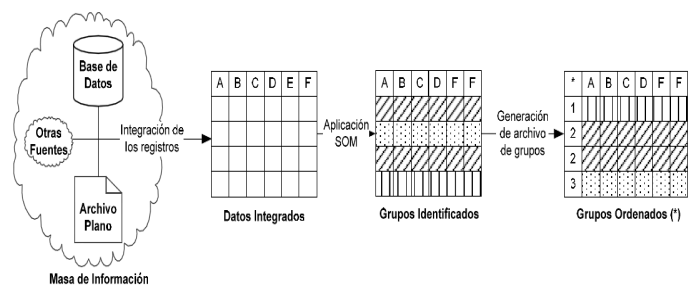


Fig. 2. Proceso de explotación de información: descubrimiento de grupos [30]

En primer lugar se identifican todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se aplican mapas auto-organizados (SOM). Como resultado de la aplicación de SOM se obtiene una partición del conjunto de registros en distintos grupos a los que se llamará grupos identificados. Para cada grupo identificado se generará el archivo correspondiente.

c) Ponderación de Interdependencia de Atributos

En [30] se define que el proceso de ponderación de interdependencia de atributos aplica cuando se requiere identificar cuáles son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema. Son ejemplos de problemas que requieren este proceso: factores con incidencia sobre las ventas, rasgos distintivos de clientes con alto grado de fidelidad a la marca, atributos claves que convierten en vendible a un determinado producto, características sobresalientes que tienen los visitantes de un website, entre otros. Para ponderar en que medida la variación de los valores de un atributo incide sobre la variación del valor de un atributo clase se propone la utilización de Redes Bayesianas. El uso de esta tecnología busca identificar si

existe interdependencia en algún grado entre los atributos que modelan el dominio de problema que la masa de información disponible representa. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 3.

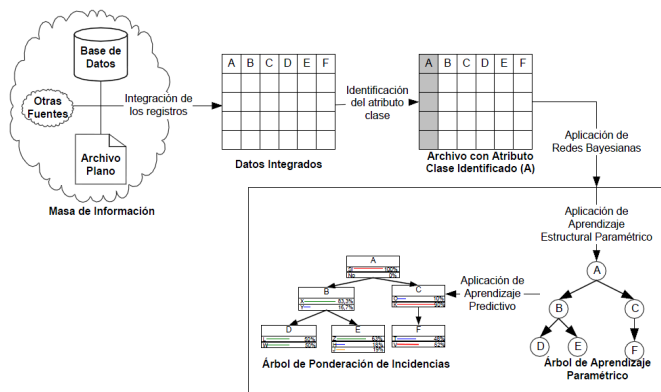


Fig. 3. Proceso de explotación de información: ponderación de interdependencia de atributos [30]

En primer lugar se identifican todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se selecciona el atributo clase (atributo A en la Figura). Como resultado de la aplicación del aprendizaje estructural de las Redes Bayesianas al archivo con atributo clase identificado se obtiene el árbol de aprendizaje; a este se le aplica el aprendizaje predictivo Redes Bayesianas y se obtiene el árbol de ponderación de interdependencias que tiene como raíz al atributo clase y como nodos hojas a los otros atributos con la frecuencia (incidencia) sobre el atributo clase.

d) Procesos de Explotación de Información

En [30] se define que el proceso de descubrimiento de reglas de pertenencia a grupos 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. Son ejemplos de problemas que requieren este proceso: tipología de perfiles de clientes y caracterización de cada tipología, distribución y estructura de los datos de mi website, segmentación etaria de mis estudiantes y comportamiento de cada segmento, clases de llamadas telefónicas en una región y caracterización de cada clase, entre otros. Para el descubrimiento de reglas de pertenencia a grupos se propone la utilización de mapas auto-organizados (SOM) para el hallazgo de los mismos y; una vez identificados los grupos, la utilización de algoritmos de inducción (TDIDT) para establecer las reglas de pertenencia a cada uno. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 4.

En primer lugar se identifican todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se aplican mapas auto-organizados (SOM). Como resultado de la aplicación de SOM se obtiene una partición del conjunto de registros en distintos grupos a los que se llama grupos identificados. Se generan los archivos asociados a cada grupo identificado. A este conjunto de archivos se lo llama grupos ordenados. El atributo “grupo” de cada grupo ordenado se identifica como el atributo clase de dicho grupo,

constituyéndose este en un archivo con atributo clase identificado (GR).

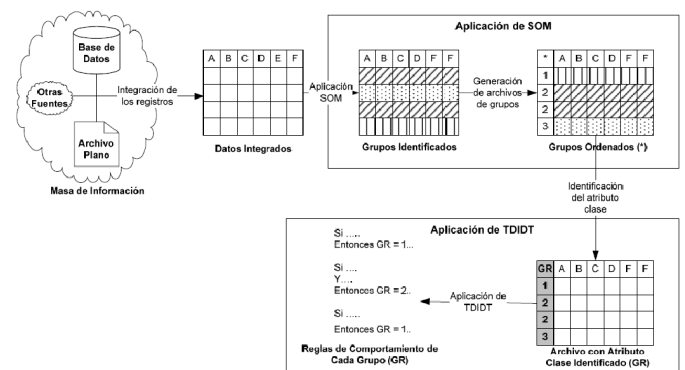


Fig. 4. Proceso de explotación de información: descubrimiento de reglas de pertenencia a grupos [30]

Se aplica el algoritmo de inducción TDIDT al atributo clase de cada grupo GR y se obtiene un conjunto de reglas que definen el comportamiento de cada grupo.

e) Procesos de Explotación de Información

En [30] se define que el proceso de ponderación de reglas de comportamiento o de la pertenencia a grupos 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 estas las que en mayor medida inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo. Son ejemplos de problemas que requieren este proceso: identificación del factor dominante que incide en el alza las ventas de un producto dado, rasgo con mayor presencia en los clientes con alto grado de fidelidad a la marca, frecuencia de ocurrencia de cada perfil de clientes, identificación del tipo de llamada más frecuente en una región, entre otros. Para la ponderación de reglas de comportamiento o de pertenencia a grupos se propone la utilización de redes bayesianas. Esto puede hacerse a partir de dos procedimientos dependiendo de las características del problema a resolver: cuando no hay clases/grupos identificados; o cuando hay clases/grupos identificados. El procedimiento a aplicar cuando hay clases/grupos identificados consiste en la utilización de algoritmos de inducción TDIDT para descubrir las reglas de comportamiento de cada atributo clase y posteriormente se utiliza redes bayesianas para descubrir cuál de los atributos establecidos como antecedentes de las reglas tiene mayor incidencia sobre el atributo establecido como consecuente. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 5.

En primer lugar se identifican todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se selecciona el atributo clase (atributo A en la figura 5).

Como resultado de la aplicación del algoritmo de inducción TDIDT al atributo clase se obtiene un conjunto de reglas que definen el comportamiento de dicha clase. Seguidamente, se construye un archivo con los atributos antecedentes y consecuentes identificados por la aplicación del algoritmo TDIDT. Como resultado de la aplicación del aprendizaje estructural de las Redes Bayesianas al archivo con atributo clase obtenido por la utilización del algoritmo TDIDT (en la

figura 5), se obtiene el árbol de aprendizaje; a este se le aplica aprendizaje predictivo y se obtiene el árbol de ponderación de interdependencias que tiene como raíz al atributo clase (en este caso el atributo consecuente) y como nodos hojas a los atributos antecedentes con la frecuencia (incidencia) sobre el atributo consecuente.

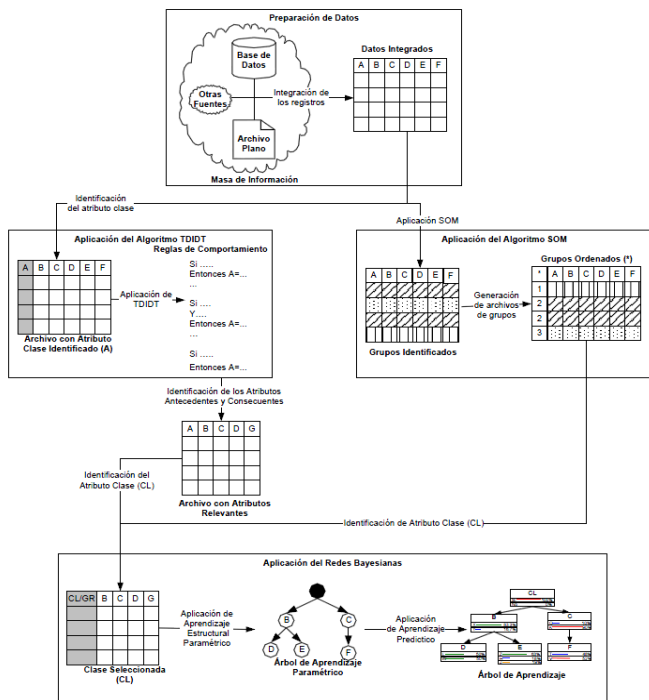


Fig. 5. Proceso de explotación de información: ponderación de reglas de comportamiento o de pertenencia a grupos [30]

El procedimiento a aplicar cuando no hay clases/grupos identificados consiste en identificar todas las fuentes de información (bases de datos, archivos planos, entre otras), se integran entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se aplican mapas auto-organizados (SOM). Como resultado de la aplicación de SOM se obtiene una partición del conjunto de registros en distintos grupos a los que se llamará grupos identificados. Para cada grupo identificado se generará el archivo correspondiente. A este conjunto de archivos se lo llama grupos ordenados. El atributo “grupo” de cada grupo ordenado se identifica como el atributo clase de dicho grupo, constituyéndose este en un archivo con atributo clase identificado (GR).

Como resultado de la aplicación del aprendizaje estructural se obtiene el árbol de aprendizaje; a este se le aplica el aprendizaje predictivo y se obtiene el árbol de ponderación de interdependencias que tiene como raíz al atributo grupo y como nodos hojas a los otros atributos con la frecuencia (incidencia) sobre el atributo grupo.

B. Marco Teórico De Ingeniería Del Conocimiento De Interés

Dentro del marco teórico de ingeniería del conocimiento de interés para el trabajo final de la carrera, se identifican los siguientes conceptos: redes semánticas (sección II.B.1), tabla Concepto-Categoría-Definición (sección II.B.2) y tabla Concepto-Atributo-Valor (sección II.B.3).

1) Redes Semánticas

En [31] se define a las redes semánticas como una notación gráfica para representar conocimiento a través de patrones de

nodos y arcos interconectados. Asimismo se las define como un formalismo o paradigma de representación de conocimiento basado en relaciones, las cuales constan de un grafo orientado formado por nodos etiquetados que representan conceptos o entidades y arcos unidireccionales etiquetados que representan relaciones entre los conceptos o instancias [32]. El formalismo de captura es un grafo orientado como se exhibe en la figura 6.

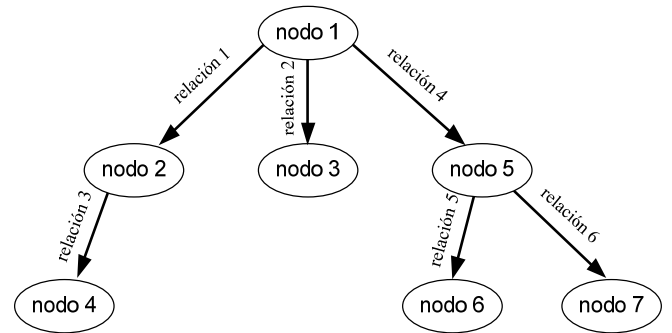


Fig. 6. Red Semántica

2) Tabla Concepto-Categoría-Definición

En [33] se presenta un conjunto de formalismos para modelar las iteraciones en espacios virtuales para trabajos colaborativos. Para este trabajo surge el interés en la tabla Concepto-Categoría-Definición.

La función de la tabla Concepto-Categoría-Definición (CCD) es representar los conocimientos fácticos del modelo conceptual de dinámica grupal. En la tabla CCD se introducen en orden lexicográfico los conceptos que se utilizarán en los otros dos formalismos especificando la categoría y dando la definición. Un concepto puede ser de alguna de las siguientes categorías: actor, objeto y acción. Los actores son los sujetos de la dinámica grupal. Los objetos son la materia o asunto que recibe el ejercicio de las facultades de accionar de los actores. Las acciones definen procesos que los actores ejecutan sobre objetos o sobre otros actores. El formalismo de captura es el de una tabla como se muestra en la tabla I.

TABLA I. TABLA CONCEPTO-CATEGORÍA-DEFINICIÓN

Concepto	Categoría	Definición
Concepto 1	Categoría 1	Definición del Concepto 1
Concepto 2	Categoría 1	Definición del Concepto 2
-----	-----	-----
Concepto N	Categoría Q	Definición del Concepto N

3) Tabla Concepto-Atributo-Valor

En el marco de los modelos para representar conocimientos que propone [34] se presenta la tabla Concepto-Atributo-Valor (CAV) como herramienta para capturar conocimientos.

La tabla Concepto-Atributo-Valor (CAV), es una técnica que modela conocimientos de tipo fácticos, esta proporciona una lista de los conceptos que se manipulan en el dominio de conocimiento relacionado con la familia de problemas que resolverá el Sistema Experto a desarrollar. Cada concepto quedará descrito en términos de los atributos que definen a cada concepto y de los valores que cada atributo puede tomar. La representación genérica de dicho formalismo se ilustra en la tabla II.

TABLA II. TABLA CONCEPTO-ATRIBUTO-VALOR

Concepto	Atributo	Valor
Concepto 1	Atributo 1	Valor 1, valor 2, ... , valor N
	Atributo 2	Valor 1, valor 2, ... , valor N

...	Atributo N	Valor 1, valor 2, ... , valor N
	Atributo 1	Valor 1, valor 2, ... , valor N
Concepto N	Atributo 2	Valor 1, valor 2, ... , valor N
	Atributo 1	Valor 1, valor 2, ... , valor N
	Atributo 2	Valor 1, valor 2, ... , valor N

III. DESCRIPCIÓN DEL PROBLEMA

En este capítulo se presenta el problema de investigación a analizar en el trabajo final de licenciatura en sistemas, comenzando con la identificación del problema de investigación (sección III.A), posteriormente se caracteriza el problema abierto (sección III.B) y se concluye con un resumen de investigación (sección III.C).

A. Identificación Del Problema De Investigación

Se ha definido una metodología para la gestión y desarrollo de sistemas de explotación de información (Modelo de Procesos para Proyectos de Explotación de Información [35]), en dicha metodología se establecen etapas destinadas al entendimiento del negocio, así como a determinar los objetivos del proceso de explotación de información que dichos problemas conllevan.

Dichos procesos son esenciales para el éxito del proyecto, omitir el esfuerzo necesario para la comprensión detallada del problema de negocio conllevará a obtener la respuesta correcta a un problema equivocado o que no satisfaga las demandas del cliente, mientras que evitar una correcta determinación de las metas de la explotación de información producirá una respuesta incorrecta al problema en cuestión.

Además de los riesgos recién mencionados, es sustancial destacar que las etapas de entendimiento del negocio y de modelado son las etapas que mayor tiempo requieren en un proyecto de explotación de información bajo el proceso de desarrollo del proyecto del modelo previamente mencionado.

En [36], se describen los tiempos requeridos por cada fase del proyecto, y detalla las fases y subfases que hacen un mayor uso de dicho recurso. A partir de los resultados obtenidos, se destaca que entre las fases de Comprensión del Negocio y de Modelado se invierte más del 50% de la duración del proyecto. Además, las subfases Determinar los Objetivos de Negocio y Evaluar la Situación, que integran la fase de Comprensión del Negocio, utilizan más del 70% del tiempo pautado. Por otro lado, en la fase de Modelado, la subfase Construcción del Modelo, requiere el 62,97% del plazo concedido a la totalidad de la ejecución de la fase a la que pertenece. En la tabla III se presenta el porcentaje de tiempo que cada fase conlleva:

TABLA III. TABLA DE CARGA DE TRABAJO DE CADA FASE DEL MODELO

FASE	% del TIEMPO
Fase 1 COMPRENSIÓN DEL NEGOCIO	20,70
Fase 2 ENTENDIMIENTO DE LOS DATOS	10,90
Fase 3 PREPARACION DE DATOS	15,61
Fase 4 MODELADO	34,41
Fase 5 EVALUACIÓN	7,45
Fase 6 IMPLANTACION	10,93

A partir de comprender lo crítico de dichas etapas, se resalta que ninguna de las metodologías para proyectos de explotación de información existentes ofrece un método que permita de forma estandarizada definir el proceso de explotación de información a aplicar, en base al dominio del negocio y al problema de explotación de información identificado, debiendo cada ingeniero del conocimiento, representar de la forma que coincidiré oportuna dichos conocimientos, dificultando la comprensión de los mismos.

Mediante la creación de un proceso, y la estandarización en la representación del negocio y del problema de explotación de información, se favorece en la documentación y la comprensión de los proyectos, permitiendo identificar similitudes entre distintos proyectos, reduciendo los tiempos de los mismos y favoreciendo la mejora de los distintos procesos productivos pertenecientes al ciclo de vida del producto.

A partir de la necesidad de utilizar un formalismo para modelar los conocimientos previamente mencionados, surge la necesidad de establecer un conjunto de reglas que estandaricen el modelado de los conocimientos involucrados en el proceso.

Los problemas previamente mencionados pueden ser sintetizados de la siguiente manera:

- Se carece de una forma estandarizada de representar los problemas de negocio y de explotación de información.
- Se carece de una terminología y de una semántica definida para el modelado del dominio del negocio y problema de explotación de información.
- Se carece de un procedimiento que permita establecer el proceso de explotación de información a utilizar, a partir del dominio del negocio, y del problema de explotación de información.

B. Problema Abierto

En la presente sección se identifica el problema abierto, que surge de la necesidad de determinar una forma estandarizada de llevar a cabo las actividades pertenecientes a las etapas críticas de entendimiento del negocio y de modelado, con el fin de garantizar la calidad y el éxito del proyecto.

Además, la ejecución de estas actividades de manera desestructurada y sin estándares, obstaculiza la mejora de las mismas en proyectos futuros. A su vez, surge la necesidad de definir un conjunto de normas para la modelización de los conocimientos dominio de negocio y del problema de explotación de información.

Por consiguiente y en virtud de todo lo expuesto, el problema abierto que se aborda en este artículo, está caracterizado por la brecha conceptual existente entre la definición del modelo de negocio y del problema de explotación de información, y la determinación del proceso de explotación de información a aplicar para dicho contexto (representada gráficamente en la figura 7). Además se destaca la carencia de una forma estándar de representar los conocimientos pertenecientes al modelo de negocio y al problema de explotación de información.

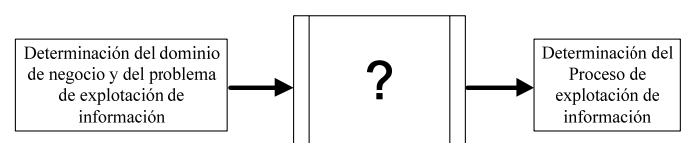


Fig. 7. Representación gráfica de la brecha conceptual entre las actividades de determinación del dominio del negocio y del problema de explotación de información

C. Problema Abierto

De lo anteriormente expuesto surgen las siguientes preguntas de investigación:

Pregunta 1: ¿Existen formalismos que permitan representar el conocimiento del negocio y del problema de explotación de información, mediante los cuales se puedan estandarizar y derivar conclusiones sobre el proceso de explotación de información que provee la solución al problema planteado? En caso afirmativo: ¿Cuáles?

Pregunta 2: ¿Pueden identificarse entre distintos modelos de negocio y problemas de información, patrones que permitan identificar el proceso de explotación de información que satisface al problema analizado? En caso afirmativo: ¿Cuáles?

Pregunta 3: En caso de existir patrones que identifiquen el proceso de explotación de información que satisface al problema analizado, ¿puede definirse un proceso que estandarice la derivación del proceso de explotación de información a partir de los conocimientos previamente mencionados? En caso de existir: ¿Cuáles son sus actividades?

IV. SOLUCIÓN

En el presente capítulo se presenta un modelo de proceso de derivación de modelos (sección IV.A), los estándares definidos para el modelado (sección IV.B) y una visión detallada del proceso de derivación de modelos (sección IV.C).

A. Proceso De Derivación De Modelos

En esta sección se presenta una propuesta de modelo de proceso de derivación de modelos, estructurada en dos partes: generalidades (sección IV.A.1) y la estructura general de la propuesta del modelo de proceso de derivación de modelos (sección IV.A.2).

1) Generalidades

En función del análisis realizado en el sección III correspondiente a la descripción del problema, se considera de interés citar nuevamente el problema abierto que se aborda en este artículo, el cual se focaliza en la brecha conceptual existente en la transición del proceso de explotación de información a aplicar, a partir de la descripción del dominio de negocio, y del problema de explotación de información. Esta brecha conceptual, dificulta la ejecución de la primeras cuatro etapas del Proceso de Desarrollo del Proyecto de Explotación de Información (de acuerdo a la modelo propuesto por [35]), teniendo un impacto crítico en el resultado final del proyecto.

A su vez, la definición de un proceso lógico para la transformación de las características del negocio y del problema de explotación de información al proceso de explotación de información, brinda una serie de actividades bien estructuradas y resultados claramente definidos, que favorecen en la comprensión de las distintas actividades que el proyecto conlleva, así como un mayor entendimiento de la estructura de los datos relevantes para el mismo, favoreciendo el desarrollo de las etapas de entendimiento y preparación de los datos.

En las siguientes citas podemos comprobar la importancia de estandarizar en la ingeniería de software:

“La estandarización de procesos provocará que de manera coordinada los procesos y esfuerzos se diseñen de forma común, es decir, todos los departamentos de la empresa o bien

si se ella cuenta con otras localidades o centros de trabajo entenderán y verán los mismo, este lenguaje único permitirá mejorar la comunicación y dará soporte en todo momento a la toma de decisiones. La organización que pretenda una evolución y crecimiento controlado podrá tomar esta iniciativa.” [37]. Del párrafo de Sánchez, se desprende que un proceso estándar mejora la comunicación entre los miembros del proyecto, reduciendo el esfuerzo y el tiempo necesario para el desarrollo de las distintas actividades, además de producirse una documentación homogénea a través de los distintos proyectos que la empresa u organización lleve a cabo, facilitando su uso.

Los procesos del software se pueden mejorar por la estandarización del proceso donde la diversidad de los procesos del software en una organización sea reducida. Esto conduce a mejorar la comunicación y a una reducción del tiempo de formación, y hace la ayuda al proceso automatizado más económica. La estandarización también es un primer paso importante para introducir nuevos métodos, técnicas y buenas prácticas de ingeniería del software. Esta mejora en los procesos significa entender los procesos existentes y mejorar la calidad del producto y/o reducir los costos y tiempos de desarrollo [38].

A partir de lo recién mencionado, se propone como solución insertar una actividad entre las etapas de determinación del modelo de negocio, y la etapa de determinación del problema de explotación de información y la definición del proceso de explotación de información, la cual denominaremos *Proceso de Derivación de Modelos* (figura 8), y cuya función es, actuar como enlace entre las etapas de determinación del negocio y determinación del problema de explotación de información (PEI), y la etapa de definición del proceso de explotación de información, eliminando la arbitrariedad en las actividades que ellas involucran, favoreciendo el éxito de las mismas.

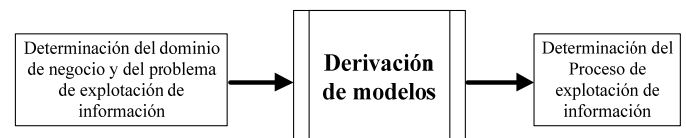


Fig. 8. Representación gráfica de la inserción del proceso de derivación de modelos propuesto entre las etapas de determinación del dominio de negocio y del PEI, y del proceso de explotación de información

El objetivo del proceso de derivación de modelos es proveer de un mecanismo estandarizado para representar la descripción del dominio de negocio y la descripción del problema de explotación de información, en modelos, a partir de los cuales se determina cuál de los procesos de explotación de información satisface o brinda la solución al problema definido, favoreciendo de este modo, al éxito y calidad de los resultados del proyecto, y dejando constancia del proceso implementado a través de la documentación obtenida, proporcionando información relevante para proyectos posteriores.

A partir de la implementación de esta actividad de derivación de modelos es posible la consecución de un conjunto de representaciones gráficas denominadas Modelos de Representación Intermedia (MRI) a partir de las cuales es posible representar la información contenida en la descripción del dominio de negocio y la descripción del problema de explotación de información (por lo general en formato de

“lenguaje natural” y es así como se la supone presentada en este trabajo), a los efectos de simplificar la comprensión del proceso de explotación de información a aplicar, mediante la identificación de patrones o estructuras similares detectables en las representaciones realizadas.

Estos modelos de representación intermedia, hacen uso de distintos formalismos de Ingeniería de Conocimiento, entre ellos: las Redes Semánticas [31], las tablas Concepto-Atributo-Valor [34] y la tabla Concepto-Categoría-Definición propuesta en [33].

Estas representaciones intermedias estarán conformadas, fundamentalmente, por un conjunto de representaciones gráficas basadas en las redes semánticas: Red Semántica del Modelo de Negocio y Red Semántica del Problema de Explotación de Información.

Para la consecución del objetivo principal, mediante la aplicación de las representaciones gráficas basadas en las redes semánticas, surge la necesidad de definir estándares para la representación de dichos formalismos, definiendo un marco de trabajo común, obteniendo así todos los beneficios que esto implica.

En síntesis, se propone el proceso de derivación de modelos, en el cual se define su estructura de forma general (sección IV.A.2) y su estructura de forma detallada (sección IV.C), como herramienta para la derivación de la descripción del dominio de negocio y del problema de explotación de información, en el proceso de explotación de información a aplicar, y propone una serie de reglas aplicadas al modelado de la información del negocio, con el objetivo de estandarizar los modelos de representación intermedia y la documentación que estos producen, explicado en detalle en la sección IV.B.

2) Estructura General del Proceso de Derivación de Modelos

La actividad de derivación de modelos se lleva a cabo por medio de un proceso denominado proceso de derivación de modelos, el cual se estructura en tres fases, de acuerdo al enfoque de las actividades que las componen:

- Una primera fase orientada al análisis del dominio del negocio, cuyo objetivo es el análisis, la comprensión y la conceptualización del modelo de negocio de la organización. Dicha fase se realizará por única vez para los distintos problemas de negocio que la organización desea enfrentar.
- Una segunda fase orientada al análisis del problema de explotación de información, cuyo objetivo abarca el análisis, la comprensión y el modelado del problema de explotación de información. Dicha fase deberá realizarse por cada problema de negocio que se quiera afrontar.
- Una tercera fase orientada a determinar el proceso de explotación de información, cuyo objetivo es identificar el proceso de explotación de información a aplicar, de acuerdo al dominio del negocio y al problema de explotación de información, expresados mediante las representaciones intermedias obtenidas en las fases previas.

Este proceso toma como elementos de partida la descripción del dominio del negocio, dentro de la cual se encuentran definidos los datos del negocio, y la descripción del problema de explotación de información, junto con la descripción de sus datos, y proporciona como salida el proceso de explotación de información a aplicar para dar solución al

problema de negocio representado mediante el problema de explotación de información. La figura 9 ilustra este concepto.

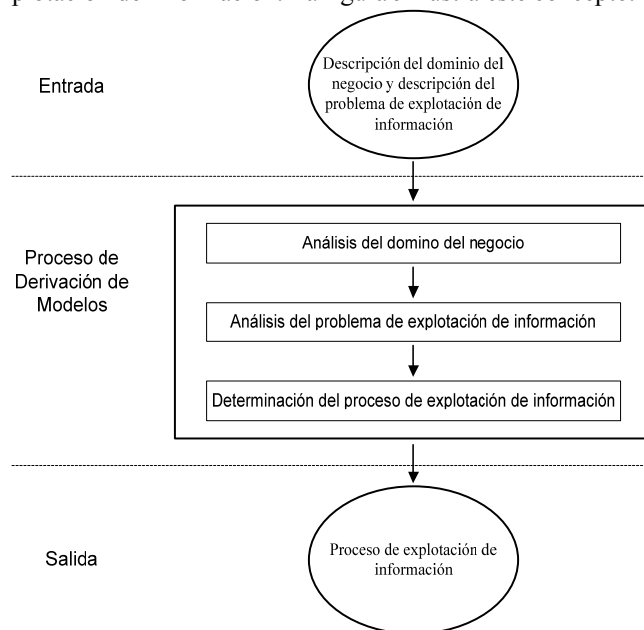


Fig. 9. Estructura general del proceso de derivación de modelos, identificando sus tres fases orientadas al enfoque de las actividades, y sus entradas y salida

El soporte principal del proceso de derivación de modelos está compuesto por tres fases, cuyas primeras dos están conformadas por tres etapas, y la última por una.

Las tres etapas que conforman la primera fase, están orientadas al modelado del dominio del negocio y cuyo resultado final es la Red Semántica del Modelo de Negocio, que brinda una representación de los elementos que componen el dominio del negocio y la interacción entre los mismos.

La segunda fase orientada al modelado del problema de explotación de información, está compuesta por tres actividades cuyo resultado final es la Red Semántica del Problema de Explotación de Información, que brinda una representación de los elementos que componen el dominio del negocio y la interacción entre los mismos.

La tercera fase orientada a determinar el proceso de explotación de información, está conformada por una actividad la cual se compone de ocho subpasos, y su salida es el proceso de explotación de información a aplicar, es decir, el resultado final esperado por el proceso de derivación de modelos.

La distribución de las fases no es de forma arbitraria, sino que definen el orden en el que se deben realizar cada una de las actividades que las conforman. Cada paso o etapa del proceso, tiene elementos de entrada, los cuales se procesan, generando elementos de salida, siendo estos a su vez, elementos de entrada de actividades posteriores.

En la figura 10 se ilustra el flujo de pasos que conlleva la ejecución del proceso de derivación de modelos, las interdependencias entre las fases, sus etapas y sus productos, representadas mediante el ingreso como elemento de entrada de un paso, uno o más elementos de salida de otro. En dicha figura se presentan las siete etapas del proceso de derivación de modelos, distribuidas según las fases a las que pertenecen. Dichas etapas se conectan mediante flechas las cuales indican los elementos de entrada (aquellas que convergen con la parte izquierda del paso) y salida (aquellas que surgen del sector

derecho del paso) correspondientes a cada etapa y como cada elemento de salida retroalimenta otra etapa posterior.

El orden de ejecución de los pasos se define a continuación, identificando las etapas con negrita, los nombres de los pasos entre comillas y los elementos de entrada/salida con cursiva:

El primer paso a ejecutar es el perteneciente a la fase orientada al **análisis del dominio del negocio**, denominado "Identificación de los elementos y estructura del dominio", el cual tiene como elementos de entrada la *descripción del dominio del negocio* y los *datos del negocio*, y produce como salida la *tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD)* y la *tabla Término-Categoría-Definición del Dominio (TCDD)*, los cuales son elementos de entrada al segundo paso denominado "Identificación de relaciones entre conceptos del dominio", y cuyo elemento resultado es la *tabla Concepto-Relación del Dominio (CRD)*, la cual es utilizada por el tercer y último paso de la fase "Conceptualización del dominio", junto con la *tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD)* como elementos de entrada, produciendo como resultado final de la etapa (pero no del proceso), la *Red Semántica del Modelo de Negocio (RSMN)*, la cual será utilizada posteriormente.

Dentro de la fase **orientada al análisis del problema de explotación de información**, podemos observar que los tres pasos que la componen, son independientes a la fase anteriormente descrita, por lo cual su ejecución puede ser realizada en paralelo.

La primera etapa que debe ejecutarse es "Identificación de los elementos y estructura del Problema de Explotación de Información", la cual tiene como elementos de entrada los *datos del problema de negocio* y la *descripción del problema de explotación de información*, produciendo como resultado de

aplicar dicho paso, dos modelos de representación intermedia, la *tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI)* y la *tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI)*.

Ambos elementos mencionados como resultados del paso anterior, ingresan como elementos de entrada a la etapa de "Identificación de relaciones entre conceptos del Problema de Explotación de Información", la cual proporciona como producto de salida la *tabla Concepto-Relación del Problema de Explotación de Información (CRPEI)*, que junto con la *tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI)*, alimentan al último paso de la etapa, denominado "Conceptualización del Problema de Explotación de Información", y que proporciona como resultado la *Red Semántica del Problema de Explotación de Información (RSPEI)*, que será posteriormente utilizada.

A partir de la conceptualización del dominio del negocio y del problema de explotación de información mediante las redes semánticas, obtenidas a partir de la ejecución del tercer paso correspondiente a las dos fases previas respectivamente, se poseen los elementos necesarios para comenzar con la última fase del proceso. Esta fase está compuesta por un único paso denominado "Derivación del Proceso de Explotación de Información", cuyos elementos de entrada son la *Red semántica del modelo de negocio (RSMN)* y la *Red semántica del problema de explotación de información (RSPEI)*, y a partir de los cuales, aplicando una serie de reglas se obtiene como resultado el *Proceso de explotación de información* que satisface el problema de explotación de información analizado.

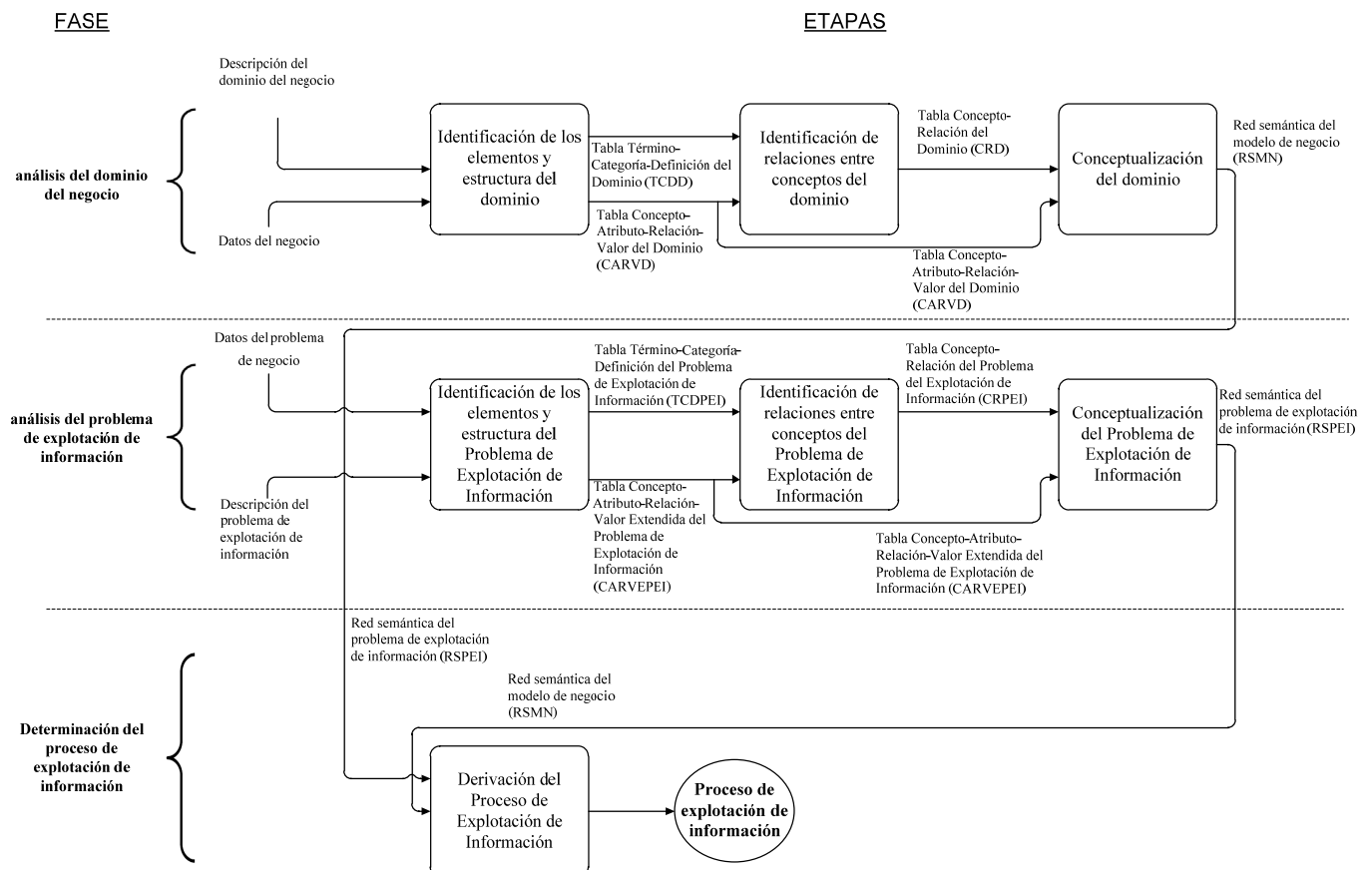


Fig. 10. Flujo e interdependencia entre las fases, pasos y sus productos

B. Estandarización Del Modelado

En esta sección, se definen un conjunto de reglas a aplicar para la realización de los modelos que surgen como producto de las distintas etapas del proceso propuesto, destinados a representar el dominio de negocio y los problemas de explotación de información.

El objetivo de definir una serie de reglas para la representación de dichos problemas es el de establecer un estándar o marco de trabajo, que permita mantener la trazabilidad entre distintos proyectos, facilitando la comprensión de los mismos.

El formalismo definido para la representación de los modelos son las redes semánticas, definidas en la sección II.

Este formalismo se centra en las relaciones existentes entre los conceptos, siendo ampliamente utilizados para representar bases y/o modelos de conocimientos. Entre sus ventajas se puede destacar la simplicidad para graficar el conocimiento, lo que también se ve reflejado en una rápida e intuitiva comprensión del conocimiento que este representa.

La simplicidad para modelar el conocimiento está dada por el reducido número de símbolos y la facilidad visual para interpretar el conocimiento representado.

A pesar de las ventajas descritas, las redes semánticas carecen de una terminología universalmente aceptada y de una semántica uniforme y precisa, dicha afirmación puede verse manifestado para el formalismo de las redes semánticas en el siguiente párrafo perteneciente a [39]:

“Se puede utilizar cualquier etiqueta para dar nombre a cualquier nodo o arco. Los principales problemas de este formalismo de representación son la falta de una terminología adecuada y universalmente aceptada y de una semántica uniforme y precisa”.

Para dar solución a dicho problema, se propone un conjunto de reglas para estandarizar el formalismo aplicado al modelado del negocio y a los problemas de explotación de información, con el objetivo de establecer un marco de trabajo común, para su aplicación en el proceso de derivación de modelos.

De acuerdo a lo presentado en la sección anterior, se distinguen aquellos elementos que hacen uso de las reglas propuestas, dichos modelos son:

- la Red Semántica del Modelo de Negocio, y
- la Red Semántica del Problema de Explotación de Información.

Entre el conjunto de reglas definidas, podemos segmentarlas en dos grupos, el primero que define los elementos de representación comunes en los distintos modelos basados en redes semánticas (sección IV.B.1), y el segundo orientado a aquellos elementos particulares de la Red Semántica del Problema de Explotación de Información, que identifican los datos de entrada y salida del problema (sección IV.B.2). En la sección IV.B.3 se presenta un resumen de los elementos propuestos y los modelos a los que aplican.

1) Notación General de Formalismos Basados en Redes Semánticas

A continuación se definen las reglas de representación que aplican a los ambos modelos indicados en la sección IV.B:

- Los conceptos se representan mediante óvalos con sus líneas continuas (Figura 11),
- los atributos o valores se representan a través de óvalos con sus líneas intermitentes (Figura 12),

- el identificador de una instancia del concepto se define dentro del óvalo concatenándolo al tipo de concepto con un guion “-” entre ambos (Figura 13),
- el valor o referencia de la etiqueta de los arcos dependerá de la relación que esta represente:
 - Si une un atributo con su valor, se etiqueta con la palabra “valor” (Figura 14.a),
 - Si indica una relación de instancia de un concepto, se etiqueta con la palabra “instancia de” (Figura 14.b),
 - Si actúa de enlace entre un concepto y su atributo, se etiqueta con una o más palabras que representen la conexión entre ambos (Figura 14.c),
 - Si indica una relación de herencia de un concepto, se etiqueta con la palabra “subclase de” (Figura 14.d).



Fig. 11. Elemento para representar conceptos en redes semánticas

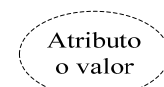


Fig. 12. Elemento para representar atributos o valores en redes semánticas

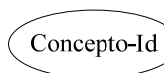


Fig. 13. Identificación de instancias de conceptos en redes semánticas

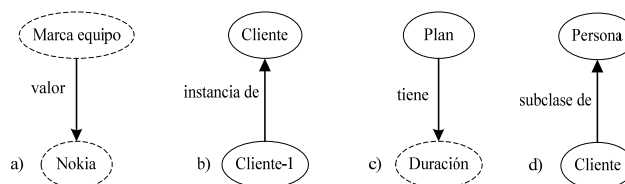


Fig. 14. Representación de las distintas etiquetas que puede tener un arco según los elementos que relaciona. a) Relación entre un atributo y un valor, b) Relación de instanciación entre conceptos, c) Relación entre un concepto y un atributo y d) Relación de herencia entre conceptos

Es primordial destacar de la última figura presentada (Figura 14), que en los casos “a)”, “b)” y “d)”, el valor posible de la etiqueta se encuentra definido, sin aceptarse otro valor para dichas relaciones, mientras que en el caso “c)”, se propone un valor ejemplo, ya que dependiendo del contenido de los conceptos-atributos que se relacionen, el valor de la etiqueta del arco variará. La primera imagen de la figura 14, indica que hay un atributo “Marca equipo”, cuyo valor es Nokia. La segunda imagen, indica que el concepto “cliente-1” (siendo 1 el identificador del cliente como se explicó en la figura 13) pertenece al concepto general cliente por lo tanto obtiene todas las características que a este lo representa. La tercera imagen, indica una relación de pertenencia entre el concepto plan y el atributo duración, indicando que todo plan tiene una duración estimada. La última imagen representa una relación de herencia, en el que la clase hereda todas las características que definen al concepto padre, en este caso el concepto Cliente hereda todas sus características de Persona, ya que un cliente es una persona, de acuerdo a esta relación.

2) Notación Específica de las Redes Semánticas del Problema de Explotación de Información

En la Red Semántica del Problema de Explotación de Información, se debe identificar el flujo del problema descripto, a través de la identificación de los elementos de entrada y de salida, para ello se incorporan elementos adicionales, para poder indicar dicha información. Dichas notaciones adicionales, son:

- Los atributos de entrada se representan con doble ovalo con sus líneas intermitentes (Figura 15),
- Los atributos de salida se representan con doble ovalo, teniendo el ovalo interno su línea continua y el ovalo externo su línea intermitente (Figura 16),
- Los identificadores de instancias de conceptos, que sean elemento de entrada se representan mediante doble ovalo, teniendo el ovalo externo su línea continua, y el ovalo interno su línea intermitente (Figura 17).
- Los identificadores de instancias de conceptos, que sean elemento de salida se representan mediante doble ovalo cuyas líneas son continuas (Figura 18).

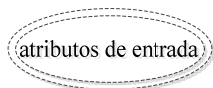


Fig. 15. Elemento para representar atributos de entrada en redes semánticas

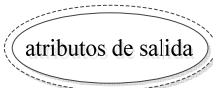


Fig. 16. Elemento para representar atributos de salida en redes semánticas

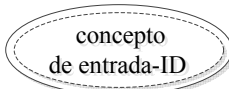


Fig. 17. Elemento para representar la variable identificadora del concepto como elemento de entrada en redes semánticas

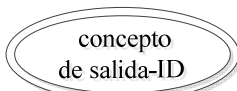


Fig. 18. Elemento para representar la variable identificadora del concepto como elemento de salida en redes semánticas

3) Resumen de Notación de Formalismos Basados en Redes Semánticas

La tabla IV que sintetiza los estándares definidos para los modelos basados en redes semánticas, y a que modelos corresponden.

C. Proceso De Derivación De Modelos

En esta sección, se describe en detalle al proceso de derivación de modelos propuesto, brindando una visión ampliada de la provista en la sección IV.A, y haciendo uso de los formalismos definidos en la sección IV.B.

Se presentan los aspectos generales del proceso propuesto y se describe el caso de estudio con el que se ilustrará el mismo (sección IV.C.1), se describe detalladamente la fase orientada al análisis del dominio del negocio (sección IV.C.2), y se definen los pasos que componen dicha fase: “Identificación de los elementos y estructura del dominio” (sección IV.C.2.a), “Identificación de relaciones entre conceptos del dominio”

(sección IV.C.2.b) y “Conceptualización del dominio” (sección IV.C.2.c). Se explica de forma semejante la fase orientada al análisis del problema de explotación de información (sección IV.C.3), la que se describe a partir de las etapas que la componen: “Identificación de los elementos y estructura del Problema de Explotación de Información” (sección IV.C.3.a), “Identificación de relaciones entre conceptos del Problema de Explotación de Información” (sección IV.C.3.b) y “Conceptualización del Problema de Explotación de Información” (sección IV.C.3.c).

TABLA IV. ELEMENTOS PROPUESTOS Y LOS MODELOS BASADOS EN REDES SEMÁNTICAS A LOS QUE CORRESPONDEN

Elemento notacional	Formalismo donde es utilizado	
	Red Semántica del Modelo de Negocio (RSMN)	Red Semántica del problema de explotación de Información (RSPEI)
Concepto	✓	✓
Atributo o valor	✓	✓
Concepto-Id	✓	✓
valor ↓	✓	✓
instancia de ↑	✓	✓
tiene ↓	✓	✓
subclase de ↑	✓	✓
atributos de entrada		✓
atributos de salida		✓
concepto de entrada-ID		✓
concepto de salida-ID		✓

Se describe detalladamente la última fase determinación del proceso de explotación de información (sección IV.C.4), conformada por el paso: “Derivación del Proceso de Explotación de Información” (sección IV.C.4.a). En la sección IV.C.5 se brinda un cuadro que sintetiza los pasos a realizar en el proceso de derivación de modelos, indicando sus entradas y salidas.

1) Aspectos Generales

Luego de finalizado las primeras tres actividades del modelo de procesos para proyectos de explotación de información [35] (entendimiento del negocio, entendimiento de los datos y preparación de los datos), la primera destinada al entendimiento de los objetivos y requerimientos del proyecto desde una perspectiva del negocio, y las restantes a identificar,

seleccionar y preparar los datos que son de utilidad para el proyecto, se poseen los elementos suficientes para dar comienzo al proceso de derivación de modelos, estos son: la descripción del dominio de negocio, la descripción de los datos del negocio, la descripción del problema de explotación de información y la descripción de los datos del problema de negocio. A partir de estos elementos es posible realizar las primeras dos fases del proceso propuesto (análisis del dominio del negocio y análisis del problema de explotación de información) debido a su independencia de conceptos, como se explicó previamente en el capítulo.

A los fines de favorecer la comprensión del proceso de derivación de modelos, las actividades serán desarrolladas en el orden propuesto, divididas en tres secciones según la fase a la que pertenecen (fase orientada al análisis del dominio del negocio (sección IV.C.2), fase orientada al análisis del problema de explotación de información (sección IV.C.3) y la fase orientada a determinar el proceso de explotación de información (sección IV.C.4)).

A efectos de ilustrar los entregables producidos en las distintas etapas del proceso propuesto, se presenta el siguiente caso de estudio [40], con el fin de facilitar la comprensión de las tareas a realizar en cada uno de los pasos y la aplicación de las distintas técnicas utilizadas en ellos.

“...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 a 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 acorde a 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 a través del 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.

Problema de explotación de información: Determinar las reglas, en base a las variables situación, el tipo de plan y su duración, la marca del equipo y el tipo de cliente, que identifican el comportamiento de un conjunto de clientes de acuerdo a su ubicación geográfica (región, provincia, localidad).

Descripción de los datos: Los siguientes son los datos relevantes almacenados por la empresa durante el ejercicio de sus actividades:

Código del Cliente: Código de cliente que la empresa utiliza para identificarlo. Ejemplo: 99861598.

Situación del Cliente: Situación del Cliente. Ejemplo: AAA (cliente activo), BA (cliente de baja).

Fecha de Alta: Fecha de alta del cliente. Ejemplo: 10/03/2003.

Código de la localidad: Código de localidad de residencia del cliente. Ejemplo: 379.

Código de la provincia: Código de la provincia a la cual pertenece la localidad. Ejemplo: 3.

Plan Contratado: Código de plan tarifario contratado por el cliente. Ejemplo: U21.

Duración: Cantidad de meses de contratación del cliente. Ejemplo: 12.

Tipo de Cliente: Tipo de cliente. Ejemplo: PR/PO (cliente prepago), FF (cliente con facturación).

Marca del Equipo: Marca del celular adquirido por el cliente. Ejemplo: NOKIA.

El conjunto de clientes será identificado mediante un valor numérico.

Región geográfica: Indica la región a la que pertenece cada provincia (tabla 4.2)...”

2) Fase Orientada al Análisis del Dominio del Negocio

En esta sección se aborda detalladamente la primera fase del proceso propuesto, la cual abarca las actividades que solo se ejecutarán por única, independientemente de que se requiera del análisis de más de un problema de explotación de información. Esta fase tienen como objetivo la comprensión, definición y modelado del dominio del negocio, analizado en la primera etapa del ciclo de vida del proyecto, generándose mediante la aplicación de los pasos propuestos una serie de resultados intermedios (modelos de representación intermedia), a través de los cuales se logra obtener la representación final de la fase (pero intermedia desde el punto de vista del proceso) la Red Semántica del Modelo de Negocio. Los pasos que componen dicha fase son: identificación de los elementos y estructura del dominio (sección IV.C.2.a), identificación de relaciones entre conceptos del dominio (sección IV.C.2.b) y conceptualización del dominio (sección IV.C.2.c).

a) Paso 1: “Identificación de los elementos y estructura del dominio”

El paso “Identificación de los elementos y estructura del dominio”, tiene como objetivo generar a partir de la descripción del dominio de negocio y de los datos del negocio (elementos de entrada del paso), una estructura que defina los elementos que integran el modelo de negocio, en la que se identifiquen los conceptos que forman parte del mismo, junto con las características que lo definen.

En base a lo establecido previamente, es sustancial definir los siguientes términos: “conceptos” son aquellos elementos que poseen características las cuales permiten distinguirse de otros, que pueden interactuar con otros elementos, o que pueden ser agrupados o generalizados. Las características de los conceptos se las denomina “atributos”, los cuales pueden tener distintos “valores”, entendiéndose por estos el estado en el que se encuentra un atributo, y los que definen los distintos estados posibles que un atributo puede tener. Las “relaciones” son las interacciones entre los distintos elementos.

Los elementos de entrada de dicho paso son la descripción del dominio del negocio y los datos del negocio.

El desarrollo del paso consta de aplicar dos técnicas, la primera de ellas la tabla Término-Categoría-Definición del Dominio (TCDD), cuyo objetivo es identificar y clasificar todos los elementos relevantes pertenecientes al dominio del negocio (sección IV.C.2.a.i) y la segunda, la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD), mediante la

cual se define la estructura de los elementos del negocio (sección IV.C.2.a.ii). Los productos resultados de aplicar ambas técnicas son elementos de salida del paso, los cuales serán utilizados posteriormente.

i. Técnica Tabla Término-Categoría-Definición del Dominio

La tabla Término-Categoría-Definición del Dominio del negocio, se basa en el formalismo tabla CCD propuesto en [33].

Por medio de esta técnica, se identifican los elementos relevantes pertenecientes al modelo de negocio, posteriormente utilizados en otros formalismos orientados a dicho dominio. Los términos detectados serán introducidos en la tabla en orden alfabético. Un término pertenece a alguna de las siguientes categorías: Concepto, Atributo ó Relación (definidas en la sección previa). Para la ejecución de la tabla TCDD, se requiere como elemento de entrada a la descripción del dominio del negocio.

Para su aplicación, se procede a identificar en la descripción del dominio del negocio, aquellos elementos relevantes para la comprensión del mismo, a partir de las definiciones anteriormente provistas. Por cada término detectado, se procederá con la identificación de la categoría a la que pertenece, y se definirá el mismo en base al dominio del negocio analizado.

La tabla V resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

TABLA V. TÉCNICA TABLA TÉRMINO-CATEGORÍA-DEFINICIÓN DEL DOMINIO

Técnica Tabla Término-Categoría-Definición del Dominio	
Entrada	Descripción del dominio del negocio
Salida	Tabla Término-Categoría-Definición del Dominio
Procedimiento	Paso 1. Identificar en el dominio del negocio aquellos términos relevantes para la comprensión del mismo Paso 2. Identificar la categoría a la que pertenece cada término identificado Paso 3. Definir cada término en base al dominio del negocio analizado

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla VI. En esta tabla se observan todos los términos identificados, a partir de la descripción del dominio del negocio descrita en el caso de estudio brindado en la sección IV.C.1, luego se define la categoría a la que pertenecen, de acuerdo a las definiciones provistas en la sección IV.C.2.a, y se brinda una descripción, la cual define al termino en base al dominio de negocio analizado.

ii. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio

Por medio de esta técnica, se define la estructura del dominio de negocio, proporcionando un listado con los conceptos que se manipulan en el mismo. Cada concepto quedará descrito a través de los atributos que lo componen, el identificador que describe su relación con el concepto y los valores posibles de cada atributo. Para la ejecución de la tabla CARVD, se dispone como productos de entrada la tabla TCDD y la descripción de los datos del negocio.

TABLA VI. TABLA TÉRMINO-CATEGORÍA-DEFINICIÓN DEL DOMINIO (TCDD) A PARTIR DEL CASO DE ESTUDIO

Término	Categoría	Definición
Cliente	Concepto	Persona que contrata el servicio
Código Cliente	Atributo	Identificación univoca de cada cliente
Código Plan	Atributo	Identificación univoca de cada plan
Contrata	Relación	El cliente contrata un plan determinado
Duración	Atributo	Cantidad de meses de contratación
es	Relación	El cliente es de un cierto tipo de clientes
Fecha alta	Atributo	Fecha en que se dio de alta en la empresa
identifica	Relación	El código de cliente identifica un cliente particular
identifica	Relación	El código de plan identifica a un plan particular
Localidad	Atributo	Localidad en la que reside el cliente
Marca equipo	Atributo	Marca de celular que el cliente posee.
pertenece	Relación	El cliente pertenece a una región
pertenece	Relación	El cliente pertenece a una provincia
pertenece	Relación	El cliente pertenece a una localidad
Plan	Concepto	Tipos de contratos que definen el servicio prestado
Provincia	Atributo	Provincia en la que reside el cliente
Región	Atributo	Región en la que reside el cliente
se registro	Relación	El cliente se registró en una fecha particular
Situación	Atributo	Estado en el que pertenece el cliente con la empresa
tiene	Relación	El cliente tiene una situación con respecto a la empresa
tiene	Relación	El cliente tiene un equipo telefónico de cierta marca
tiene	Relación	El plan tiene una duración determinada
Tipo cliente	Atributo	Clasificación del cliente según su facturación

Para su aplicación, se procede a identificar en la tabla TCDD, aquellos términos cuya categoría es “Concepto”, introduciéndolos en la columna etiquetada con el mismo nombre. Posteriormente se agrupan en la columna atributo, aquellos términos presentes en la tabla TCDD cuya categoría es “Atributo”, asignándolos en la fila correspondiente al concepto con el cual dicho atributo se relaciona, es decir, existe un término en la tabla TCDD, cuya categoría es “Relación”, que vincula al concepto con el atributo a asignar. El identificador del término de categoría “Relación” previamente identificado, será asignado en la columna relación, obteniéndose la estructura general del negocio. A partir de la descripción de los datos, se procede a identificar los posibles valores de cada atributo en la columna valor.

La tabla VII resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

La tabla VIII se obtiene al asignar en la columna concepto aquellos términos identificados en la tabla VI, cuya categoría es “Concepto”. Luego se asigna en la columna atributo aquellos términos cuya categoría es homónima, en la fila correspondiente al concepto con el cual se relaciona, y se registra la descripción del término de categoría “Relación” que vincula al concepto con el atributo en la columna relación.

Finalmente, en la columna valor se asignan los posibles valores del atributo, de acuerdo a lo definido en la descripción de los datos (sección IV.C.1).

TABLA VII. TÉCNICA TABLA CONCEPTO-ATRIBUTO-RELACIÓN-VALOR DEL DOMINIO

Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio	
Entrada	Tabla Término-Categoría-Definición del Dominio Descripción de los datos del negocio
Salida	Tabla Concepto-Atributo-Relación-Valor del Dominio
Procedimiento	<p>Paso 1. identificar en la tabla TCDD, aquellos términos cuya categoría es "Concepto" y registrarlos en la columna homónima</p> <p>Paso 2. identificar en la tabla TCDD, aquellos términos cuya categoría es "Atributo" y se relacionen con uno de los conceptos previamente identificados, registrándolos en la columna <i>atributo</i> de la fila perteneciente al concepto asociado</p> <p>Paso 3. identificar en la tabla TCDD, el término "Relación" que vincula al concepto con el atributo que lo compone y registrar su nombre en la columna <i>relación</i></p> <p>Paso 4. registrar en base a la descripción de los datos del negocio, los valores posibles de cada atributo identificado</p>

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla VIII:

TABLA VIII. TABLA CONCEPTO-ATRIBUTO-RELACIÓN-VALOR DEL DOMINIO (CARVD) A PARTIR DEL CASO DE ESTUDIO

Concepto	Atributo	Relación	Valor
Cliente	Código	identifica	Númérico
	situación	tiene	Alfabetico EJ: AAA, BA.
	fecha Alta	se registró	dd/mm/AAAA
	provincia	pertenece	Númérico EJ: 379
	localidad	pertenece	Númérico EJ: 3
	región	pertenece	1 al 5
	tipo cliente	es	Alfabetico EJ: PR/PO, FF.
Plan	marca equipo	tiene	Alfabetico EJ: Nokia
	Código	identifica	Alfanumérico EJ: U21.
	duración	tiene	Númérico

b) Paso 2: "Identificación de Relaciones entre Conceptos del Dominio"

El objetivo del paso "Identificación de Relaciones entre Conceptos del Dominio", es detectar a partir de los conceptos identificados en el paso anterior, las relaciones existentes en el modelo del negocio, entre dichos conceptos, estableciendo los vínculos y modos de interacción que tienen los distintos elementos que componen al dominio del negocio. Los elementos de entrada de dicho paso son las tablas TCDD y CARVD generadas en el paso anterior.

Para la aplicación de dicho paso, se implementa la técnica tabla Concepto-Relación del Dominio, mediante la cual se identifican las relaciones entre los conceptos detectados según el dominio del negocio (sección IV.C.2.b.i). El elemento resultante de aplicar la técnica recién mencionada, será la salida del paso actual.

i. Técnica Tabla Concepto-Relación del Dominio

Por medio de esta técnica, se definen las interacciones entre los conceptos del dominio de negocio. Dicha tabla proporciona una lista de las relaciones entre los conceptos que se definen en el dominio. Cada relación quedará definida a través de los

conceptos que la conforman, el nombre de la relación y una descripción de la misma. Para la aplicación de la tabla CRD, se dispone como productos de entrada las tablas TCDD y CARVD.

Para comenzar a aplicar la tabla CRD, se procede a identificar aquellos términos cuya categoría sea "Relación", presentes en la tabla TCDD que indiquen un vínculo entre dos conceptos del problema identificados en la tabla CARVD. Por cada elemento identificado se registrará en la columna concepto, aquel concepto que genera dicha relación, en la columna concepto asociado aquel concepto vinculado, en la columna relación el nombre que describe el tipo de vínculo y la descripción que defina en base al dominio del negocio la relación identificada, en la columna homónima. Estos últimos dos elementos presentes en la tabla TCDD.

La tabla IX resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

TABLA IX. TÉCNICA TABLA CONCEPTO-RELACIÓN DEL DOMINIO

Técnica Tabla Concepto-Relación del Dominio	
Entrada	Tabla Término-Categoría-Definición del Dominio Tabla Concepto-Atributo-Relación-Valor del Dominio
Salida	Tabla Concepto-Relación del Dominio
Procedimiento	<p>Paso 1. identificar en la tabla TCDD, aquellos términos cuya categoría es "Relación" que vinculen dos conceptos identificados en la tabla CARVD y registrar su nombre identificativo y su descripción en las columnas correspondientes</p> <p>Paso 2. identificar los conceptos que intervienen en cada relación registrada, y registrar al concepto que origina dicha relación y al concepto vinculado, en las columnas concepto y concepto asociado respectivamente</p>

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla X.

TABLA X. TABLA CONCEPTO-RELACIÓN DEL DOMINIO (CRD) A PARTIR DEL CASO DE ESTUDIO

Concepto	Concepto asociado	Relación	Descripción
cliente	plan	contrata	El cliente contrata un plan por un tiempo determinado

En la tabla X, se asignaron aquellos términos identificados en la tabla TCDD cuya categoría es "Relación" y en su descripción se vinculan dos conceptos, identificando en la columna "concepto", aquel que genera la relación y en la columna "concepto asociado", el concepto vinculado.

c) Paso 3: "Conceptualización Del Dominio"

A través del paso "Conceptualización Del Dominio", se obtiene un modelo que representa gráficamente toda la información relevante del dominio del negocio, enfocado en las relaciones entre los elementos que componen al negocio. Los elementos de entrada de dicho paso son las tablas CARVD y CRD, esta última obtenida del paso anterior.

Para su aplicación, se implementa la Técnica Red Semántica del Modelo de Negocio (sección IV.C.2.c.i) mediante la cual se obtiene la representación del dominio del negocio, y cuyo formalismo resultante será el elemento de salida del paso.

i. Técnica Red Semántica del Modelo de Negocio

Por medio de esta técnica, se expresan las interacciones entre los elementos del modelo de negocio, las cuales representan el conocimiento del negocio, desde un enfoque orientado a las relaciones de sus componentes. La conceptualización obtenida de aplicar dicha técnica, proporciona una visión general, simple y precisa de los elementos del negocio, sus características y sus relaciones. La notación a utilizar es la definida en la sección IV.B. Para la aplicación de la RSMN, se dispone como productos de entrada la tabla CARVD, que contiene la estructura del negocio, y CRD, que comprende las relaciones entre los conceptos del negocio.

Para comenzar a aplicar la RSMN, se procede a asociar los elementos (conceptos, atributos, relaciones y valores) identificados en la tabla CARVD obteniendo las estructuras de los conceptos implicados. A partir de las estructuras obtenidas, se asocian aquellas que se relacionen, en base a la tabla CRD, constituyendo la red semántica del modelo de negocio.

La tabla XI resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

TABLA XI. TABLA CONCEPTO-RELACIÓN DEL DOMINIO (CRD) A PARTIR DEL CASO DE ESTUDIO

Técnica Red Semántica del Modelo de Negocio	
Entrada	Tabla Concepto-Atributo-Relación-Valor del Dominio Tabla Concepto-Relación del Dominio
Salida	Red Semántica del Modelo de Negocio
Procedimiento	<p>Paso 1. Asociar los elementos (conceptos, atributos, relaciones y valores) identificados en la tabla CARVD obteniendo las estructuras de los conceptos implicados</p> <p>Paso 2. A partir de las estructuras, se asocian aquellas que se identifica una relación en la tabla CRD, constituyendo la red semántica del modelo de negocio</p>

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la figura 19. Para realizar dicha figura, se grafica mediante las notaciones propuestas en la sección IV.B, los elementos identificados en la tabla CARVD. Se representan los conceptos y posteriormente se los relaciona con los atributos que lo componen, en base a las relaciones definidas. Una vez definida la estructura, se vinculan aquellos conceptos relacionados en la tabla CRD.

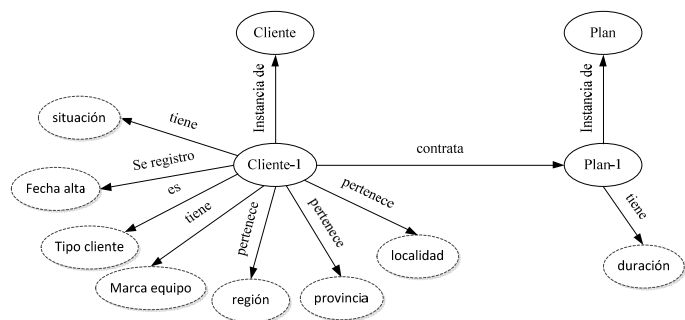


Fig. 19. Red Semántica del Modelo de Negocio (RSMN) a partir del caso de estudio

3) Fase Orientada al Análisis del Dominio del Negocio

En esta sección se aborda detalladamente la segunda fase del proceso propuesto, la cual abarca aquellas actividades dedicadas a la comprensión y representación del problema de explotación de información, definido en la primera etapa del ciclo de vida del proyecto previamente mencionado.

La aplicación de esta fase puede realizarse en paralelo a la fase orientada al análisis del dominio del negocio, ya que esta tiene independencia con respecto a las salidas de las actividades de la fase previa. Esta fase deberá realizarse por cada problema de explotación de información a analizar.

A través de aplicar los pasos propuestos, se obtienen una serie de resultados intermedios (modelos de representación intermedia), mediante de los cuales se logra obtener la representación final de la fase (pero intermedia desde el punto de vista del proceso) la Red Semántica del Problema de Explotación de Información. Las etapas que la componen son: identificación de los elementos y estructura del problema de explotación de información (sección IV.C.3.a), identificación de relaciones entre conceptos del problema de explotación de información (sección IV.C.3.b) y conceptualización del problema de explotación de información (sección IV.C.3.c).

a) Paso 4: "Identificación de los elementos y estructura del Problema de Explotación de Información"

El objetivo del paso "Identificación de los elementos y estructura del Problema de Explotación de Información", es detectar a partir de la descripción del problema de explotación de información los distintos elementos que lo componen. A partir de los elementos detectados, junto con la descripción de los datos del problema de explotación de información se define la estructura de los conceptos identificados, se determinan las relaciones existentes entre conceptos y se define el flujo del problema de explotación de información mediante los elementos de entrada y salida del mismo, entendiéndose como los atributos a partir de los cuales se define el resultado y los atributos esperados como resultados respectivamente. Los elementos de entrada de dicho paso son la descripción del problema de explotación de información y los datos del problema de explotación de información.

El desarrollo del paso consta de aplicar dos técnicas, la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI), cuyo objetivo es identificar y clasificar todos los elementos relevantes pertenecientes al problema de explotación de información (sección IV.C.3.a.i), y la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), mediante la cual se define la estructura de los elementos que integran al problema y se define el flujo del problema, a partir de la identificación de sus elementos de entrada y salida (sección IV.C.3.a.ii). Los productos resultados de aplicar estas técnicas son elementos de salida del paso, los cuales serán utilizados posteriormente.

i. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información

La tabla Término-Categoría-Definición del Problema de Explotación de Información, se basa en el formalismo CCD propuesto en [33]. Por medio de esta técnica, se identifican las ideas pertenecientes al problema de explotación de información, posteriormente utilizadas en formalismos orientados al problema de explotación. Los términos detectados serán introducidos en la tabla en orden alfabético. Un término puede ser categorizado en "Concepto", "Atributo"

o “Relación”, de acuerdo a lo definido en la sección IV.C.2.a. Para la ejecución de la tabla TCDPEI, se dispone como productos de entrada la descripción del problema de explotación de información y de los datos del problema de negocio.

Para comenzar a aplicar la tabla TCDPEI, se procede a identificar en el problema de explotación de información los elementos relevantes para la comprensión del mismo, a partir de las definiciones previamente descriptas, por cada término detectado, se procederá con la identificación de la categoría a la que pertenece, y se definirá el mismo en base al problema de explotación de información analizado.

La tabla XII resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

TABLA XII. TÉCNICA TABLA TÉRMINO-CATEGORÍA-DEFINICIÓN DEL PROBLEMA DE EXPLOTACIÓN DE INFORMACIÓN

Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información		
Entrada	Descripción del problema de explotación de información	
	Descripción de los datos del problema de negocio	
Salida	Tabla Término-Categoría-Definición del Problema de Explotación de Información	
Procedimiento	Paso 1.	Identificar en el problema de explotación de información los elementos relevantes para la comprensión del mismo
	Paso 2.	Por cada término detectado, identificar la categoría a la que pertenece cada elemento
	Paso 3.	Definir cada elemento en base al problema de explotación de información analizado

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla XIII. En esta tabla se observan todos los términos identificados, a partir de la descripción del problema de explotación de información, descriptos en el caso de estudio brindado en la sección IV.C.1, luego se define la categoría a la que pertenecen, de acuerdo a las definiciones provistas en la sección IV.C.2.a, y se brinda una descripción que define a los mismos.

i. Técnica Red Semántica del Modelo de Negocio

Por medio de esta técnica, se define la estructura del problema de explotación de información, proporcionando un listado con los conceptos que se manipulan en dicho problema. Cada concepto quedará descripto a través de los atributos que lo componen, el identificador que describe su relación con el concepto y los valores posibles de cada atributo. La tabla Concepto-Atributo-Relación-Valor Extendida, presenta una columna adicional denominada entrada/salida, en la cual se identifican los atributos que actúan como factores iniciales y resultantes del problema de explotación de información identificándolos con la palabra “entrada” y “salida” respectivamente. Para la ejecución de la tabla CARVEPEI, se dispone como productos de entrada la tabla TCDPEI y la descripción de los datos y del problema de explotación de información.

Para la aplicación la tabla CARVEPEI, se identifican en la tabla TCDPEI, aquellos términos cuya categoría es “concepto”, introduciéndolos en la columna etiquetada con el mismo nombre. Posteriormente se agrupan en la columna atributo, aquellos términos presentes en la tabla TCDPEI cuya categoría es “Atributo”, asignándolos en la fila correspondiente al concepto con el cual dicho atributo se relaciona, es decir, existe

un término en la tabla TCDPEI, cuya categoría es “Relación”, que vincula al concepto con el atributo a asignar.

TABLA XIII. TABLA TÉRMINO-CATEGORÍA-DEFINICIÓN DEL PROBLEMA DE EXPLOTACIÓN DE INFORMACIÓN (TCDPEI) A PARTIR DEL CASO DE ESTUDIO

Término	Categoría	Definición
Cliente	Concepto	Persona que contrata el servicio
Código de grupo	Atributo	Identificación univoca de cada grupo
Código Plan	Atributo	Identificación univoca de cada plan
Código regla	Atributo	Identificación univoca de la regla
define	Relación	La regla define la asignación del cliente a un grupo determinado
duración	Atributo	Cantidad de meses de contratación
Grupo	Concepto	Conjunto de clientes
identifica	Relación	El código de plan identifica a un plan particular
identifica	Relación	El código de grupo identifica un grupo particular
identifica	Relación	El código regla identifica a una regla particular
integrado por	Relación	El grupo está integrado por clientes
Localidad	Atributo	Localidad en la que reside el cliente
Marca equipo	Atributo	Marca de celular que el cliente posee.
Plan	Concepto	Tipos de contratos que definen el servicio prestado
Provincia	Atributo	Provincia en la que reside el cliente
Región	Atributo	Región en la que reside el cliente
Regla	Concepto	Variables que definen la asignación de un cliente a un determinado grupo
según	Relación	El cliente pertenece a un grupo según su localidad
según	Relación	El cliente pertenece a un grupo según su provincia
según	Relación	El cliente pertenece a un grupo según su región
Situación	Atributo	Estado en el que se encuentra el cliente con la empresa
Subconjunto	Relación	La situación puede ser una variable que defina la asignación del cliente en un grupo
Subconjunto	Relación	El plan puede ser una variable que defina la asignación del cliente en un grupo
Subconjunto	Relación	La marca del celular puede ser una variable que defina la asignación del cliente en un grupo
Subconjunto	Relación	El tipo de cliente puede ser una variable que defina la asignación del cliente en un grupo
tiene	Relación	El plan tiene una duración determinada
Tipo cliente	Atributo	Clasificación del cliente según su facturación

El identificador del término de categoría “Relación” previamente identificado, será asignado en la columna relación, obteniéndose la estructura general del negocio. A partir de la descripción de los datos, se procede a identificar los posibles valores de cada atributo en la columna valor. Finalmente, se identifica en la columna entrada/salida, los atributos que actúan como entrada y salida del problema de explotación de información.

La tabla XIV resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

TABLA XIV. TÉCNICA TABLA TÉRMINO-CATEGORÍA-DEFINICIÓN DEL PROBLEMA DE EXPLOTACIÓN DE INFORMACIÓN

Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información	
Entrada	Descripción del problema de explotación de información Descripción de los datos del problema de negocio
Salida	Tabla Término-Categoría-Definición del Problema de Explotación de Información Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información
Procedimiento	<p>Paso 1. identificar en la tabla TCDPEI, aquellos términos cuya categoría es "Concepto" y registrarlos en la columna homónima</p> <p>Paso 2. identificar en la tabla TCDPEI, aquellos términos cuya categoría es "Atributo" y se relacionen con uno de los conceptos previamente identificados, registrándolos en la columna atributo de la fila perteneciente al concepto asociado</p> <p>Paso 3. identificar en la tabla TCDPEI, el término "Relación" que vincula al concepto con el atributo que lo compone y registrar su nombre en la columna relación</p> <p>Paso 4. registrar en base a la descripción de los datos del negocio, los valores posibles de cada atributo identificado</p> <p>Paso 5. registrar los atributos de entrada y salida del problema analizado</p>

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla XV.

TABLA XV. TÉCNICA TABLA TÉRMINO-CATEGORÍA-DEFINICIÓN DEL PROBLEMA DE EXPLOTACIÓN DE INFORMACIÓN

Concepto	Atributo	Relación	Entrada/Salida	Valor
Regla	Código regla	identifica	Salida	Númérico
	Situación	subconjunto		Alfabético EJ: AAA, BA.
	Marca equipo	subconjunto		Alfabético EJ: Nokia
	Tipo cliente	subconjunto		Alfabético EJ: PR/PO, FF.
Plan	Código Plan	identifica		Alfanumérico EJ: U21.
	duración	tiene		Númérico
Grupo	Código de grupo	identifica	Salida	Númérico
Cliente	provincia	según	Entrada	Númérico EJ: 379
	localidad	según	Entrada	Númérico EJ: 3
	región	según	Entrada	1 al 5

La tabla XV, se obtiene al asignar en la columna concepto aquellos términos identificados en la tabla XIII, cuya categoría es "Concepto". Luego se asigna en la columna atributo aquellos términos cuya categoría es homónima, en la fila correspondiente al concepto con el cual se relaciona, y se registra la descripción del término de categoría "Relación" que vincula al concepto con el atributo en la columna relación. Posteriormente, en la columna valor se asignan los posibles valores del atributo, de acuerdo a lo definido en la descripción de los datos (sección IV.C.1). Finalmente, se identifican en la columna entrada/salida, aquellos atributos que actúan como elementos de ingreso y resultado en el problema analizado.

b) Paso 5: "Identificación de relaciones entre conceptos del Problema de Explotación de Información"

El objetivo del paso "Identificación de relaciones entre conceptos del Problema de Explotación de Información", es detectar a partir de los conceptos identificados en el paso anterior, las relaciones existentes en el problema de explotación de información, entre dichos conceptos, estableciendo los vínculos y modos de interacción que tienen los distintos elementos que componen al problema. Los elementos de entrada de dicho paso son las tablas TCDPEI y CARVEPEI generadas en el paso anterior.

Para la aplicación de dicho paso, se implementa la técnica tabla Concepto-Relación del Problema de Explotación de Información, mediante la cual se identifican las relaciones entre los conceptos detectados según el problema de explotación de información (sección IV.C.3.b.i). El elemento resultante de aplicar la técnica recién mencionada, será el elemento salida del paso actual.

i. Técnica Tabla Concepto-Relación del Problema de Explotación de Información

Por medio de esta técnica, se definen las interacciones entre los conceptos del problema de explotación de información. Dicha tabla proporciona un listado de las relaciones entre los conceptos que se definen en el problema tratado. Cada relación quedará definida a través de los conceptos que la conforman, el nombre de la relación y una descripción de la misma. Para la aplicación de la tabla CRPEI, se dispone como productos de entrada las tablas TCDPEI y CARVEPEI.

Para comenzar a aplicar la tabla CRPEI, se procede a identificar aquellos términos cuya categoría sea "Relación", presentes en la tabla TCDPEI que indiquen un vínculo entre dos conceptos del problema identificados en la tabla CARVEPEI. Por cada elemento identificado se registrará en la columna concepto, aquel concepto que genera dicha relación, en la columna concepto asociado aquel concepto vinculado, en las columnas relación y descripción, se registra el nombre que describe el tipo de vínculo y la definición del mismo en base al problema de explotación de información, presentes en la tabla TCDPEI.

La tabla XVI resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

TABLA XVI. TÉCNICA TABLA CONCEPTO-RELACIÓN DEL PROBLEMA DE EXPLOTACIÓN DE INFORMACIÓN

Técnica Tabla Concepto-Relación del Problema de Explotación de Información	
Entrada	Tabla Término-Categoría-Definición del Problema de Explotación de Información Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información
Salida	Tabla Concepto-Relación del Problema de Explotación de Información
Procedimiento	<p>Paso 1. identificar en la tabla TCDPEI, aquellos términos cuya categoría es "Relación" que vinculen dos conceptos identificados en la tabla CARVEPEI y registrar su nombre identificativo y su descripción en las columnas correspondientes</p> <p>Paso 2. identificar los conceptos que intervienen en cada relación registrada, y registrar al concepto que origina dicha relación y al concepto vinculado, en las columnas <i>concepto</i> y <i>concepto asociado</i> respectivamente</p>

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la tabla XVII.

En la tabla XVII, se asignaron aquellos términos identificados en la tabla TCDPEI cuya categoría es “Relación” y en su descripción se vinculan dos conceptos, identificando en la columna “concepto”, aquel que genera la relación y en la columna “concepto asociado”, el concepto vinculado,

TABLA XVII. TABLA CONCEPTO-RELACIÓN DEL PROBLEMA DE EXPLOTACIÓN DE INFORMACIÓN (CRPEI) A PARTIR DEL CASO DE ESTUDIO

Concepto	Concepto asociado	Relación	Descripción
Plan	Regla	Subconjunto	El plan puede ser una variable que defina la asignación del cliente en un grupo
Regla	Grupo	define	La regla define la asignación del cliente a un grupo determinado
Grupo	Cliente	integrado por	El grupo está integrado por clientes

asignando en la columna “Relación” un nombre que identifique a la misma.

c) Paso 6: “Conceptualización del Problema de Explotación de Información”

A través del paso “Conceptualización del Problema de Explotación de Información”, se obtiene un modelo que representa gráficamente toda la información relevante del problema de explotación de información, enfocado en las relaciones entre los elementos que lo componen. Los elementos de entrada de dicho paso son las tablas CARVEPEI y CRPEI, esta última obtenida en el paso anterior.

Para su aplicación, se implementa la Técnica Red Semántica del Problema de Explotación de Información (sección IV.C.3.c.i) mediante la cual se obtiene la representación de la estructura del problema de explotación de información, su flujo y las variables de inferencia definidas a través de los elementos de entrada y salida, y cuyo formalismo resultante será el elemento de salida del paso.

i. Técnica Red Semántica del Problema de Explotación de Información

Por medio de esta técnica, se expresan las interacciones entre los elementos del problema de explotación de información, las cuales representan el conocimiento del problema, desde un enfoque orientado a las relaciones de sus componentes. La notación a utilizar es la definida en la sección IV.B. Para la aplicación de la RSPEI, se dispone como productos de entrada la tabla CARVEPEI, que contiene la estructura del problema de explotación de información, y CRPEI, que comprende las relaciones entre los conceptos del problema.

Para su aplicación, se procede a asociar los elementos (conceptos, atributos, relaciones y valores) identificados en la tabla CARVEPEI obteniendo las estructuras de los conceptos implicados. A partir de las estructuras obtenidas, se asocian aquellas relacionadas, en base a la tabla CRPEI. Finalmente, se identifican los elementos de entrada, salida y las variables de inferencia en base a las notaciones específicas de las redes semánticas del problema de explotación de información definidas en la sección IV.B.2, constituyendo la RSPEI.

La tabla XVIII resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

TABLA XVIII. TÉCNICA RED SEMÁNTICA DEL PROBLEMA DE EXPLOTACIÓN DE INFORMACIÓN

Técnica Red Semántica del Problema de Explotación de Información	
Entrada	Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información Tabla Concepto-Relación del Problema de Explotación de Información
Salida	Red Semántica del Problema de Explotación de Información
Procedimiento	Paso 1. Asociar los elementos (conceptos, atributos, relaciones y valores) identificados en la tabla CARVEPEI obteniendo las estructuras de los conceptos implicados Paso 2. Asociar aquellas estructuras que se relacionen, en base a tabla CRPEI Paso 3. Identificar los nodos de entrada, salida y variables

Como resultado de aplicar el procedimiento definido al caso de estudio propuesto, se obtiene la figura 20.

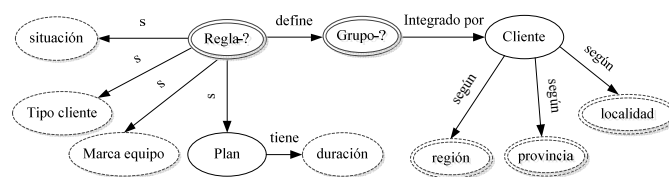


Fig. 20. Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio

Para realizar la figura 20, se grafica mediante las notaciones propuestas en la sección IV.B, los elementos identificados en la tabla CARVEPEI. Se representan los conceptos y posteriormente se los relaciona con los atributos que lo componen. Una vez definida la estructura, se vinculan aquellos conceptos relacionados en la tabla CRPEI. Los nodos variables identifican la incógnita definida en el problema de explotación, estos se ligan a los nodos constantes de la red (aquellos nodos que no son variables) para encontrar una equiparación que responda al problema planteado.

La interpretación del problema planteado es: “determinar una regla compuesta por el subconjunto de los elementos situación, tipo de cliente, marca equipo, plan y la duración que tiene el plan, que definan al grupo (mediante su identificador) que está compuesto por clientes según su región, provincia y localidad.”.

El significado de la interpretación debe implicar el mismo sentido que el problema de explotación definido a partir del problema de negocio presentado en la sección IV.C.1.

4) Fase de Determinación del Proceso de Explotación de Información

En esta sección se aborda detalladamente la última fase del proceso propuesto, cuyo objetivo es derivar a partir de las representaciones obtenidas en las fases anteriores, el proceso de explotación de información correspondiente.

En base a lo mencionado en el párrafo anterior, se puede inferir a partir de las dependencias que posee la fase, que para su ejecución es necesario haber finalizado las fases previas. Esta fase deberá realizarse por cada problema de explotación de información a analizar.

Esta fase se compone de la etapa derivación del proceso de explotación de información detallada a continuación (sección IV.C.4.a).

a) Paso 7: “Derivación del Proceso de Explotación de Información”

El objetivo del paso “derivación del proceso de explotación de información”, es brindar un mecanismo que permita deducir el proceso de explotación de información a aplicar para dar solución al problema de explotación de información definido. Para lograr dicho objetivo, se busca identificar en los modelos de representación intermedia, previamente realizados, un conjunto de patrones o estructuras comunes que permitan definir el proceso de explotación de información que provee la solución al problema. Los elementos de entrada son los resultados finales de cada fase previa la RSMN y la RSPEI.

Para su aplicación, se implementa la Técnica Algoritmo de Derivación del Proceso de Explotación de Información (sección IV.C.4.a.i), la cual define una serie de sub pasos que analizan la estructura de las redes semánticas obtenidas como elemento de entrada, con el objetivo de definir el proceso de explotación de información, el cual es el resultado final del paso y del proceso propuesto.

i. Técnica Algoritmo de Derivación del Proceso de Explotación de Información

Por medio de esta técnica, se obtienen las técnicas y algoritmos de explotación de información que deben aplicarse para dar solución al problema de negocio planteado, pudiendo procesar la información disponible una vez que los datos se encuentren transformados y listos para ser utilizados, en otras palabras, el proceso de explotación de información. Este elemento resultante es la salida esperada del proceso de derivación de modelos. Para la aplicación del ADPEI, se dispone como productos de entrada la RSMN y la RSPEI.

Para comenzar a aplicar el ADPEI, se procede a identificar en las distintas redes semánticas ingresadas como elemento de entrada, la estructura descrita en el primer subpaso, variando la secuencia posterior de subpasos a ejecutar, a partir del resultado obtenido en cada subpaso. Como resultado de realizar los distintos caminos posibles (en base a las respuestas obtenidas en el caso analizado), se pueden obtener dos posibles respuestas:

- Representación errónea, ó
- El proceso de explotación de información correspondiente al problema definido.

La primera de las respuestas posibles, indica que alguna de las etapas previamente desarrolladas, se ha realizado de manera errónea. La segunda es la respuesta esperada y el resultado final del proceso. La tabla XIX resume los pasos necesarios para la implementación de esta técnica y sus elementos de entrada y salida.

El paso general indicado en la tabla 4.17, que define el procedimiento está compuesto de 8 subpasos condicionales, que a partir del resultado obtenido al ejecutar el primer subpaso, se define el flujo de pasos mediante los cuales

TABLA XIX. TÉCNICA ALGORITMO DE DERIVACIÓN DEL PROCESO DE EXPLOTACIÓN DE INFORMACIÓN

Técnica Algoritmo de Derivación del Proceso de Explotación de Información	
Entrada	Red Semántica del Modelo de Negocio
	Red Semántica del Problema de Explotación de Información
Salida	Proceso de explotación de información
Procedimiento	Paso 1. Identificar en las redes semánticas de entrada, aquellos patrones definidos en el Algoritmo de Derivación del Proceso de Explotación de Información

se obtiene el proceso de explotación de información a aplicar.

Para comprender cada uno de los subpasos que componen al ADPEI, es necesario definir los siguientes conceptos:

- **Nodo variable** son aquellos elementos interrogantes definidos en el problema de explotación de información, identificados en la red semántica con el símbolo “?”.
- Al menos debe existir un nodo variable en la red semántica del problema de explotación de información, de otra forma, la representación de la misma es incorrecta.

La tabla XX define cada uno de los pasos y su destino de acuerdo al resultado obtenido. En dicha tabla, se puede observar que el primer paso busca identificar en la Red Semántica del Problema de Explotación de Información (RSPEI), la cantidad de nodos variables existentes.

Si solo se identifica un nodo variable, se procede con el subpaso 7.2, el cual busca identificar si el único nodo variable en la RSPEI, tiene un arista la cual se identifica con las palabras pertenece, integra, es, es miembro de, u otras palabras de similar significado. Si la respuesta al paso 7.2 es afirmativa, entonces el proceso a aplicar es el proceso de descubrimiento de grupos, sino se procede con el subpaso 7.3, el cual busca identificar si el nodo variable previamente reconocido, se identifica con algunas de las siguientes palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado. Si la respuesta es afirmativa, se define que el proceso a aplicar es el proceso de ponderación de interdependencia de atributos, sino se procede con el paso 7.4, el cual busca identificar si el nodo variable previamente reconocido, se identifica con algunas de las siguientes palabras: reglas, normas, criterios u otras palabras de similar significado. Si la respuesta es afirmativa, entonces se debe aplicar el proceso de descubrimiento de reglas de comportamiento, sino se determina que la representación no es correcta.

Si la respuesta del subpaso 7.1 fuese mayor a uno, se procede a ejecutar el subpaso 7.5 el cual busca identificar en la RSPEI si alguno de los nodos variables posee una arista (originada en este), identificada con alguna de las siguientes palabras: integrado por, compuesto por, formado por, u otras palabras de similar significado y dicho nodo variable, se relaciona con otro nodo variable siendo este (el primer nodo) el nodo destino de la relación. Si la respuesta es negativa, se determina que la representación no es correcta, sino se prosigue con el subpaso 7.6 identificando si el nodo variable de origen (aquel que surge la arista que lo relaciona con el nodo variable identificado en el subpaso previo), se identifica con alguna de las siguientes palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado.

Si la respuesta al subpaso 7.6 es afirmativa, se procede con el subpaso 7.7, que busca identificar si el nodo variable de destino, identificado en el subpaso 7.5, existe en la Red Semántica del Modelo del Negocio (RSMN). Si es un nodo existente en la conceptualización del dominio, el proceso a aplicar es el Proceso Ponderación de Reglas de Pertenencia a Grupos (Grupos definidos), si este no existiese, entonces se aplica el Proceso Ponderación de Reglas de Pertenencia a Grupos (Grupos no definidos).

Si la respuesta al subpaso 7.6 es negativa, se procede con el subpaso 7.8, el cual busca identificar si el nodo variable de origen (aquel nodo de donde surge la arista que lo relaciona con otro nodo variable) identificado en el paso 7.5, esta

etiquetado con alguna de las siguientes palabras: reglas, normas, criterios u otras palabras de similar significado.

TABLA XX. ALGORITMO DE DERIVACIÓN DEL PROCESO DE ALGORITMO DE DERIVACIÓN DEL PROCESO DE EXPLOTACIÓN DE INFORMACIÓN

Subpaso 7.1:	SI	en la Red Semántica del Problema de Explotación de Información (RSPEI) puede identificarse un único nodo variable
	ENTONCES	Ir a Subpaso 7.2
	SINO	Ir a Subpaso 7.5
Subpaso 7.2:	SI	el objetivo de la Red Semántica del Problema de Explotación de Información (RSPEI) es el identificador del nodo variable cuya arista (la cual surge a partir del nodo variable) se identifica con las palabras: pertenece, integra, es, es miembro de, u otras palabras de similar significado
	ENTONCES	Aplicar Proceso de Descubrimiento de Grupos
	SINO	Ir a Subpaso 7.3
Subpaso 7.3:	SI	el nodo variable identificado en la Red Semántica del Problema de Explotación de Información (RSPEI) se identifica con alguna de las siguientes palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado
	ENTONCES	Aplicar Proceso de Ponderación de Interdependencia de atributos
	SINO	Ir a Subpaso 7.4
Subpaso 7.4:	SI	el nodo variable identificado en la Red Semántica del Problema de Explotación de Información (RSPEI) se identifica con las palabras: reglas, normas, criterios u otras palabras de similar significado
	ENTONCES	Aplicar Proceso de Descubrimiento de Reglas de Comportamiento.
	SINO	Representación errónea.
Subpaso 7.5:	SI	en la Red Semántica del Problema de Explotación de Información (RSPEI) se identifica un nodo variable el cual posee una arista, la cual surge a partir de este, identificada con alguna de las siguientes palabras: integrado por, compuesto por, formado por, u otras palabras de similar significado y se relaciona con otro nodo variable siendo este el nodo destino de la relación
	ENTONCES	Ir a Subpaso 7.6.
	SINO	Representación errónea.
Subpaso 7.6:	SI	el nodo variable de origen que integra la relación identificada en el Subpaso anterior, se identifica con las palabras: elementos claves, factores destacados, elementos relevantes u otras palabras de similar significado
	ENTONCES	Ir a Subpaso 7.7.
	SINO	Ir a Subpaso 7.8.
Subpaso 7.7:	SI	el nodo variable de destino que integra la relación, entre los nodos variables, referida en el Subpaso 7.5, es un nodo existente en la Red Semántica del Modelo del Negocio (RSMN)
	ENTONCES	Aplicar Proceso de Ponderación de Reglas de Pertenencia a Grupos (Grupos definidos).
	SINO	Aplicar Proceso de Ponderación de Reglas de Pertenencia a Grupos (Grupos no definidos).
Subpaso 7.8:	SI	el nodo variable de origen que integra la relación, entre los nodos variables, referida en el Subpaso 7.5, se identifica con las palabras: reglas, normas, criterios u otras palabras de similar significado
	ENTONCES	Aplicar Proceso de Descubrimiento de Reglas de Pertenencia a Grupos.
	SINO	Representación errónea.

EXPLOTACIÓN DE INFORMACIÓN (ADPEI)

En caso positivo, el proceso a aplicar es el Proceso Descubrimiento de Reglas de Pertenencia a Grupos, de otra forma, las representaciones previamente realizadas no son correctas.

Como resultado de aplicar la técnica definida al caso de estudio propuesto, se obtiene la siguiente secuencia de pasos, en la cual se presentan las redes implicadas con los elementos que comprueban el resultado obtenido de cada etapa, resaltados mediante líneas punteadas (tanto nodos como arcos) y mediante subrayado para textos:

Subpaso 7.1: Se identifican dos nodos variables en la Red Semántica del Problema de Explotación de Información (figura 21).

- Resultado obtenido: Ir a Subpaso 7.5

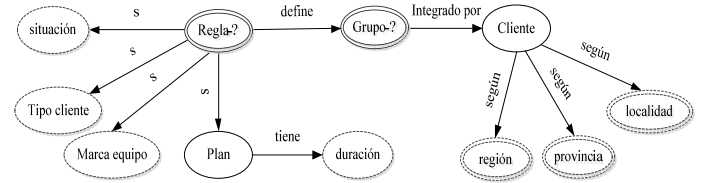


Fig. 21. Identificación de subpaso 7.1 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio

Subpaso 7.5: En la Red Semántica del Problema de Explotación de Información, uno de los nodos variable (etiquetado como “Grupo”), posee una arista identificada con las palabras “integrado por” y es el nodo destino de la relación “define” con el nodo variable “Regla” (figura 22).

- Resultado obtenido: Ir a Subpaso 7.6

Subpaso 7.6: En la Red Semántica del Problema de Explotación de Información, no se identifica un nodo variable etiquetado con las palabras mencionadas.

- Resultado obtenido: Ir a Subpaso 7.8

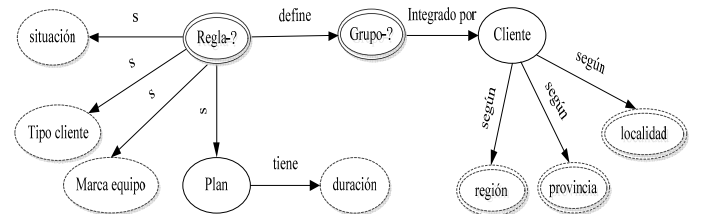


Fig. 22. Identificación de subpaso 7.5 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio

Subpaso 7.8: En la Red Semántica del Problema de Explotación de Información, el nodo variable origen de la relación detectada en el subpaso 7.5, se identifica con la palabra “Regla” (figura 23).

- Resultado obtenido: Aplicar Proceso de Descubrimiento de Reglas de Pertenencia a Grupos

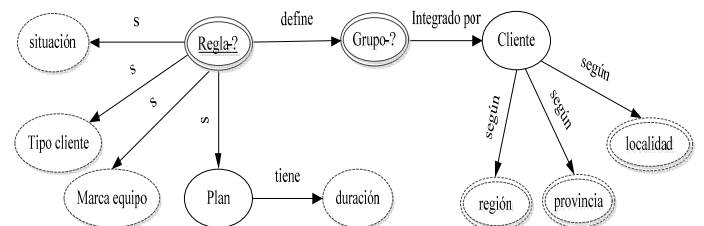


Fig. 23. Identificación de subpaso 7.8 en la Red Semántica del Problema de Explotación de Información (RSPEI) a partir del caso de estudio

Retornando como respuesta del paso y del proceso de derivación de modelos aplicado al caso de estudio, el texto:

“Proceso de Descubrimiento de Reglas de Pertenencia a Grupos”.

5) Resumen del Proceso de Derivación de Modelos

En esta sección se presenta un resumen del proceso de derivación de modelos propuesto (tabla XXI), que describe la secuencia de pasos que componen dicho proceso, junto con sus entradas y salidas correspondientes.

TABLA XXI. RESUMEN DEL PROCESO DE DERIVACIÓN DE MODELOS

		INPUT		OUTPUT	
		ENTRADA	REPRESENTACIÓN	SALIDA	REPRESENTACIÓN
Paso 1:	Identificación de los elementos y estructura del dominio	<ul style="list-style-type: none"> Descripción del dominio del negocio Datos del negocio 	<ul style="list-style-type: none"> Texto plano 	<ul style="list-style-type: none"> Representación sistemática de los elementos relevantes del dominio Representación sistemática de la estructura general del dominio 	<ul style="list-style-type: none"> Tabla Término-Categoría-Definición del Dominio (TCDD) Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD)
Paso 2:	Identificación de relaciones entre conceptos del dominio	<ul style="list-style-type: none"> Representación sistemática de los elementos relevantes del dominio Representación sistemática de la estructura general del dominio 	<ul style="list-style-type: none"> Tabla Término-Categoría-Definición del Dominio (TCDD) Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) 	<ul style="list-style-type: none"> Representación sistemática de las relaciones entre los Conceptos del dominio 	<ul style="list-style-type: none"> Tabla Concepto-Relación del Dominio (CRD)
Paso 3:	Conceptualización del dominio	<ul style="list-style-type: none"> Representación sistemática de la estructura general del dominio Representación sistemática de las relaciones entre los Conceptos del dominio 	<ul style="list-style-type: none"> Tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) Tabla Concepto-Relación del Dominio (CRD) 	<ul style="list-style-type: none"> Representación sistemática del modelo de negocio 	<ul style="list-style-type: none"> Red semántica del modelo de negocio (RSMN)
Paso 4:	Identificación de los elementos y estructura del Problema de Explotación de Información	<ul style="list-style-type: none"> Problema de explotación de información Datos del problema de negocio 	<ul style="list-style-type: none"> Texto plano 	<ul style="list-style-type: none"> Representación sistemática de los elementos relevantes del dominio Representación sistemática de la estructura del Problema de Explotación de Información y sus elementos de entrada y salida 	<ul style="list-style-type: none"> Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI)
Paso 5:	Identificación de relaciones entre conceptos del Problema de Explotación de Información	<ul style="list-style-type: none"> Representación sistemática de los elementos relevantes del dominio Representación sistemática de la estructura del Problema de Explotación de Información y sus elementos de entrada y salida 	<ul style="list-style-type: none"> Tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) 	<ul style="list-style-type: none"> Representación sistemática de las relaciones entre los Conceptos del Problema de Explotación de Información 	<ul style="list-style-type: none"> Tabla Concepto-Relación del Problema del Explotación de Información (CRPEI)
Paso 6:	Conceptualización del Problema de Explotación de Información	<ul style="list-style-type: none"> Representación sistemática de la estructura del Problema de Explotación de Información y sus elementos de entrada y salida Representación sistemática de las relaciones entre los Conceptos del Problema de Explotación de Información 	<ul style="list-style-type: none"> Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) Tabla Concepto-Relación del Problema del Explotación de Información (CRPEI) 	<ul style="list-style-type: none"> Representación sistemática del problema de explotación de información 	<ul style="list-style-type: none"> Red semántica del problema de explotación de información (RSPEI)
Paso 7:	Derivación del Proceso de Explotación de Información	<ul style="list-style-type: none"> Representación sistemática del modelo de negocio Representación sistemática del problema de explotación de información 	<ul style="list-style-type: none"> Red semántica del modelo de negocio (RSMN) Red semántica del problema de explotación de información (RSPEI) 	<ul style="list-style-type: none"> Proceso de explotación de información 	<ul style="list-style-type: none"> Texto plano

V. CONCLUSIONES

En esta sección se presentan las aportaciones de este artículo (sección V.A) y se destacan las futuras líneas de investigación que se consideran de interés en base al problema abierto presentado (sección V.B).

A. Aportes Del Artículo

En este artículo se ha definido un proceso que permite partiendo de la descripción del dominio del negocio, del problema de explotación de información, y la descripción de sus datos, derivar el proceso de explotación de información a aplicar que responda al problema de negocio planteado por el cliente, permitiendo realizar dicho proceso de manera controlada y documentada. Además se han propuesto una serie de formalismos de representación de conocimiento que permiten documentar la información necesaria para llevar a cabo la ejecución del Proceso de Derivación de Modelos propuesto, a partir del cual se identifica el proceso de explotación de información a aplicar.

En este contexto, en este TFL se ha propuesto:

I. Un modelo de proceso de derivación del proceso de explotación de información que se desarrolla en tres fases: orientada al análisis del dominio del negocio, orientada al análisis del problema de explotación de información y orientada a determinar el proceso de explotación de información.

II. Para la Fase Orientada al Análisis del Dominio del Negocio, se han propuesto las siguientes etapas: [i] Identificación de los elementos y estructura del dominio, la cual necesita la descripción del dominio del negocio y los datos del negocio como producto de entrada y proporciona como producto de salida la tabla Término-Categoría-Definición del Dominio (TCDD) y la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD); [ii] Identificación de relaciones entre conceptos del dominio, que toma como producto de entrada a la tabla Término-Categoría-Definición del Dominio (TCDD) y la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) y proporciona como producto de salida la tabla Concepto-Relación del Dominio (CRD); y [iii] Conceptualización del dominio, cuyos elementos de entrada son la tabla Concepto-Atributo-Relación-Valor del Dominio (CARVD) y la tabla Concepto-Relación del Dominio (CRD), y proporciona como producto de salida la Red Semántica del Modelo de Negocio (RSMN).

III. Para la Fase Orientada al Análisis del Problema de Explotación de Información, se han propuesto las siguientes etapas: [iv] Identificación de los elementos y estructura del Problema de Explotación de Información, la cual necesita como productos de entrada la descripción del problema de explotación de información y los datos del problema de negocio y produce como producto de salida la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) y la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI); [v] Identificación de relaciones entre conceptos del Problema de Explotación de Información, que tiene como elementos de entrada la tabla Término-Categoría-Definición del Problema de Explotación de Información (TCDPEI) y la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI), y produce como salida la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI); [vi]

Conceptualización del Problema de Explotación de Información que tiene como elementos de entrada la tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (CARVEPEI) y la tabla Concepto-Relación del Problema de Explotación de Información (CRPEI), y cuyo elemento de salida es la Red Semántica del Problema de Explotación de Información (RSPEI).

IV. Para la Fase de Determinación del Proceso de Explotación de Información, se ha propuesto la siguiente etapa: [vii] Derivación del Proceso de Explotación de Información, la cual tiene como elementos de entrada la Red Semántica del Modelo de Negocio (RSMN) y la Red Semántica del Problema de Explotación de Información (RSPEI), y produce como elemento de salida el proceso de explotación de información a aplicar.

V. Para la Fase Orientada al Análisis del Dominio del Negocio, se han desarrollado las siguientes técnicas: Técnica Tabla Término-Categoría-Definición del Dominio, Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio, Técnica Tabla Concepto-Relación del Dominio y Técnica Red Semántica del Modelo de Negocio.

VI. Para la Fase Orientada al Análisis del Problema de Explotación de Información, se han desarrollado las siguientes técnicas: Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información, Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información, Técnica Tabla Concepto-Relación del Problema de Explotación de Información y Técnica Red Semántica del Problema de Explotación de Información.

VII. Para la Fase de Determinación del Proceso de Explotación de Información, se ha desarrollado la Técnica Algoritmo de Derivación del Proceso de Explotación de Información.

B. Futuras Líneas De Investigación

Durante el desarrollo de este artículo han surgido cuestiones que si bien no son centrales al tema abordado en la misma, constituyen temas concomitantes que (en opinión del autor) darían lugar a las siguientes futuras líneas de investigación:

- En relación al proceso propuesto, quedan como líneas de investigación la ampliación de la validación empírica mediante técnicas de muestras apareadas de control.
- Si bien el algoritmo de derivación propuesto prevé la existencia de errores como una posibilidad en las representaciones realizadas, queda como línea de trabajo abierta la identificación de un proceso que permita a partir de la obtención de "Representación errónea" como resultado de aplicar la Técnica Algoritmo de Derivación del Proceso de Explotación de Información, detectar el o los errores en los formalismos de representación realizados.

FINANCIAMIENTO

Esta investigación ha sido financiada 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 Secretaría de Ciencia y Técnica de la Universidad Nacional de Lanús (Argentina).

REFERENCIAS

- [1] Negash, S., Gray, P. 2008. Business Intelligence. En Handbook on Decision Support Systems 2, ed. F. Burstein y C. Holsapple (Heidelberg, Springer), Pág. 175-193.
- [2] García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. 2011. 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.
- [3] García-Martínez, R., Britos, P., Rodríguez, D. 2013. Information Mining Processes Based on Intelligent Systems. Lecture Notes on Artificial Intelligence, 7906: 402-410. ISBN 978-3- 642-38576-6.
- [4] García Martínez, R., Servente, M. y Pasquini, D. 2003. Sistemas Inteligentes. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.
- [5] Britos, P. 2008. Procesos de Explotación de Información Basados en Sistemas Inteligentes. Tesis doctoral. Universidad Nacional de La Plata.
- [6] Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uhturudsamy, R. 1996. Advances in Knowledge Discovery and Data Mining, (editors). AAAI Press.
- [7] Grossman, R., Kasif, S., Moore, R., Rocke, D., Ullman, J. 1999. Data Mining Research: Opportunities and Challenges.
- [8] Larose, D. T. 2005. Discovering Knowledge in Data, an introduction to DataMining. John Wiley & Sons, inc. Estados Unidos.
- [9] Stefanovic, N., Majstorovic, V., Stefanovic, D. 2006. Supply Chain Business Intelligence Model. Proceedings 13th International Conference on Life Cycle Engineering. Pág. 613-618.
- [10] Reinschmidt, J., Allison F. 2000. Business Intelligence Certification Guide. IBM International Technical Support Organization. SG24-5747-00.
- [11] Koubarakis, M., Plexousakis, D. 2000. A Formal Model for Business Process Modeling and Design. Lecture Notes in Computer Science, 1789: 142-156.
- [12] Grigori, D., Casati, F., Dayal, U., Shan, M. 2001. Improving Business Process Quality through Exception Understanding, Prediction, and Prevention. Proceedings 27th International Conference on Very Large Data Bases. Pág. 159-168.
- [13] Grigori, D., Casati, F., Castellanos, M., Dayal, u., Sayal, M., Shan, M. 2004. Business Process Intelligence. Computers in Industry 53(3): 321-343.
- [14] Mäkipää, M. 2004. The Role and Types of Business Information in Different "Schools of Thought" of Strategic Management. Proceedings Frontiers of E-Business Research. Pág. 414-427.
- [15] Negash, S., Gray, P. 2008. Business Intelligence. En Handbook on Decision Support Systems 2, ed. F. Burstein y C. Holsapple (Heidelberg, Springer), Pág. 175-193.
- [16] Thomas, J. 2001. Business Intelligence - Why?. AI Journal, Julio: 47-49.
- [17] Thomsen, E. 2003. BI's Promised Land. Intelligent Enterprise, 6(4): 21-25. Lönnqvist, A., Pirttimäki, V. 2006. The Measurement of Business Intelligence. Information Systems Management, 23(1): 32-40.
- [18] Lönnqvist, A., Pirttimäki, V. 2006. The Measurement of Business Intelligence. Information Systems Management, 23(1): 32-40.
- [19] Schiefer, J., Jeng, J., Kapoor, S., Chowdhary, P. 2004. Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pág. 162-169.
- [20] Pirttimäki, V., Lönnqvist, A., Karjaluo, A. 2006. Measurement of Business Intelligence in a Finnish Telecommunications Company. Electronic Journal of Knowledge Management, 4(1): 83-90.
- [21] Reiterer, H., Mußler, G., Mann, T., Handschuh, S. 2000. INSYDER: An Information Assistant for Business Intelligence. Proceedings 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Pág.112-119.
- [22] Morik, K., Rüping, S. 2002. A Multistrategy Approach to the Classification of Phases in Business Cycles. Lecture Notes in Computer Science, 2430: 307-318.
- [23] Moss, L. 2003. Nontechnical Infrastructure of BI Applications. DM Review 13(1): 42-45.
- [24] Rudin, K., Cressy, D. 2003. Will the Real Analytic Application Please Stand Up? DM Review 13(3): 30-34.
- [25] Nguyen, T., Schiefer, J., Tjoa, M. 2005. Sense & Response Service Architecture (SARESA): An Approach Towards a Real-Time Business Intelligence Solution and Its Use for a Fraud Detection Application. Proceedings of the 8th ACM International Workshop on Data Warehousing and OLAP. Pág. 77-86.
- [26] Osterwalder, A. 2004. The Business Model Ontology a Proposition in a Design Science Approach. Tesis Doctoral. Universidad de Lausanne.
- [27] Langseth, J., Vivatrat, N. 2003. Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound. Intelligent Enterprise 5(18): 34-41.
- [28] Dasgupta, S., Vankayala, K. 2007. Developing Realtime Business Intelligence Systems. The Agile Way. Proceedings 1st Annual 2007 IEEE Systems Conference. Pág. 63-69.
- [29] Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 1999. CRISP-DM 1.0 Step by step data mining guide.
- [30] Britos, P. y García Martínez, R. 2009. Propuesta de Procesos de Explotación de Información. Proceedings XV Congreso Argentino de Ciencias de la Computación Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1
- [31] Sowa, J. 1992. Semantic Networks. En Encyclopedia of Artificial intelligence (Editor S. Shapiro, 2da edición). Wiley & Sons. ISBN 978-0471503071.
- [32] Winston, P. 1994. Inteligencia Artificial (3ra edición). Addison Wesley Iberoamericana. ISBN 0-201-51876-7.
- [33] Rodríguez, D. 2012. Espacios virtuales para la formación de investigadores elementos de análisis y diseño. Tesis de Maestría. Universidad Nacional de La Plata.
- [34] García Martínez, R. y Britos, P. 2004. Ingeniería de Sistemas Expertos. Editorial Nueva Librería. ISBN 987-1104-15-4.
- [35] Vanrell, J. A. 2011. Un Modelo de Procesos para Proyectos de Explotación de Información. Tesis de Maestría en Ingeniería en Sistemas de Información. Escuela de Postgrado FRBA-UTN.
- [36] Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2010. Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información. Anales del XVI Congreso Argentino de Ciencias de la Computación. Pág. 664-673. ISBN 978-950-9474-49-9.
- [37] Sánchez, J. P. 2012. Propuesta de Estandarización de Procesos. Tesis de Maestro en Administración.
- [38] Sommerville, I. 2005. Ingeniería del software séptima edición.
- [39] Suárez-Figueroa, M. y Gómez-Pérez, A. 2008. Redes Semánticas y Marcos. En: Inteligencia Artificial: Técnicas, Métodos y Aplicaciones. McGraw Hill. Pág. 131-170. ISBN 978-84-481-5618-3.

[40] Basso, D. y Vegega, C. 2013. 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.



Sebastian Martins. Licenciado en Sistemas por la Universidad Nacional de Lanús (UNLa). Su investigación se centra en “Derivación de Procesos de Explotación de Información”. Es Investigador Asistente del Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información del Grupo de Investigación en Sistemas de Información (UNLa GISI) y Docente de la Asignaturas Tecnologías de Explotación de Información y Fundamentos de la Teoría de la Computación de la Licenciatura en Sistemas del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús.