

EXPLOTACIÓN DE INFORMACIÓN

INGENIERÍA DE PROYECTOS

Ramón García-Martínez - Paola Britos
Sebastian Martins - Ezequiel Baldizzoni



García-Martínez, Ramón - Britos, Paola - Martins, Sebastian - Baldizzoni, Ezequiel
Ingeniería de proyectos de explotación de información. - 1a ed. - Buenos
Aires : Nueva Librería, 2015.
292 p. ; 23.5 x 17 cm

ISBN 978-987-1871-34-6
1. AAA. I. Título
CDD 000

© 2015

Nueva Librería S.R.L.

Estados Unidos 301

(1101) Buenos Aires, Argentina

Tel.: (54 11) 4362.9266

libros@nuevalibreria.com.ar

www.nuevalibreria.com.ar

ISBN 10: 987-1871-34-1

ISBN 13: 978-987-1871-34-6

Reservados todos los derechos. Ninguna parte del material cubierto por esta obra puede ser reproducida, almacenada en un sistema de informática o transmitida de cualquier forma o por cualquier método electrónico, mecánico, fotocopia, grabación u otros métodos sin el expreso permiso escrito del autor.

Hecho el depósito que marca la Ley 11.723.

Impreso en Argentina - Printed in Argentina

ÍNDICE

CAPITULO 1: INTRODUCCIÓN AL A INGENIERÍA DE PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN	17
1.1. Convenciones	17
1.2. Proyectos de Ingeniería de Explotación de Información	18
1.3. Equipos de Investigación y Desarrollo en Ingeniería de Explotación de Información en Argentina	20
1.4. Estructura del Libro	23
CAPITULO 2: MODELO DE PROCESO DE ADMINISTRACIÓN Y DESARROLLO DE PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN	25
2.1. Introducción	25
2.2. Notación Utilizada en la Solución	25
2.3. Modelo de Proceso: MoProPEI	27
2.3.1. Aspectos Generales	28
2.3.2. Estructura General	29
2.3.3. Sub-Proceso: Gestión	31
2.3.3.1. Iniciación del Proyecto	36
2.3.3.2. Planificación del proyecto	36
2.3.3.3. Soporte	37
2.3.3.4. Gestión del control y calidad	38
2.3.3.5. Gestión de la Entrega	38
2.3.4. Sub-Proceso: Desarrollo	49
2.3.4.1. Entendimiento Del Dominio	53
2.3.4.2. Entendimiento de los datos	53
2.3.4.3. Modelado	54
2.3.4.4. Preparación de los datos	54
2.3.4.5. Implementación	54
2.3.4.6. Evaluación y Presentación	55
CAPITULO 3: EDUCCIÓN DE REQUISITOS PARA EXPLOTACIÓN DE INFORMACIÓN	63
3.1. Introducción	63
3.2. Estructura del Proceso de Educción de Requisitos	66
CAPÍTULO 4: VIABILIDAD DE PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN	73
4.1. Contexto y Generalidades del Modelo para Evaluación de la Viabilidad	73
4.2. Modelo para la Evaluación de la Viabilidad	76

4.2.1. Condiciones a considerar para la Evaluación de la Viabilidad	77
4.2.2. Proceso para Evaluación de la Viabilidad	78
4.3. Análisis Preliminar del Modelo para Evaluación de la Viabilidad	83
4.3.1. Prueba de Concepto del Modelo para Evaluación de la Viabilidad	83
4.3.2. Análisis Estadístico del Modelo para Evaluación de la Viabilidad	86
4.3.2.1. Análisis Estadístico por Dimensión	87
4.3.2.2. Análisis Estadístico por Categoría	91
4.3.2.2. Análisis Estadístico por Categoría ???	--
4.3.2.3. Análisis Estadístico por Viabilidad	92
4.3.2.4. Conclusiones del Análisis Estadístico para el Modelo de Viabilidad	93
CAPÍTULO 5: ESTIMACIÓN DEL ESFUERZO DE PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN	95
5.1. Contexto y Aspectos generales del Modelo de Estimación del Esfuerzo para Proyectos de Explotación de Información en el contexto de las PyMEs	95
5.2. Modelo para la Estimación de Esfuerzo	96
5.2.1. Factores de Costo para la Estimación de Esfuerzo	97
5.2.2. Especificación de la Fórmula Lineal para la Estimación de Esfuerzo	102
5.2.3. Especificación del Método Empírico para la Estimación de Esfuerzo	117
5.3. Análisis Preliminar del Modelo para la Estimación de Esfuerzo	119
5.3.1. Prueba de Concepto del Modelo para la Estimación de Esfuerzo	119
5.3.2. Análisis Estadístico del Modelo para la Estimación de Esfuerzo	121
5.3.2.1. Análisis Estadístico General	122
5.3.2.2. Análisis Estadístico de la Fórmula Lineal para la Estimación de Esfuerzo	124
5.3.2.3. Análisis Estadístico del método Empírico para la Estimación de Esfuerzo	128
5.3.2.4. Conclusiones del Análisis Estadístico para el Modelo de Estimación de Esfuerzo	131
CAPÍTULO 6: PROCESOS DE EXPLOTACIÓN DE INFORMACIÓN	133
6.1. Introducción	133
6.2. Procesos de Explotación de Información	133
6.2.1. Descubrimiento de Reglas de Comportamiento	134
6.2.2. Descubrimiento de Grupos	134
6.2.3. Ponderación de Interdependencia de Atributos	134
6.2.4. Descubrimiento de Reglas de Pertenencia a Grupos	134
6.2.5. Ponderación de Reglas de Comportamiento o de la Pertenencia a Grupos	135
6.3. Procesos de Explotación de Información Basados en Tecnologías de Sistemas Inteligentes	135
6.3.1. Procesos Basados en Tecnologías de Sistemas Inteligentes Unitarias	135

6.3.1.1. Algoritmos TDIDT Aplicados al Descubrimiento de Reglas de Comportamiento	135
6.3.1.2. Mapas Auto Organizados (SOM) Aplicados al Descubrimiento de Grupos	138
6.3.1.3. Redes Bayesianas Aplicadas a la Ponderación de Interdependencia entre Atributos	142
6.3.2. Procesos Basados en Tecnologías de Sistemas Inteligentes Combinadas	144
6.3.2.1. SOM Y TDIDT Aplicados al Descubrimiento de Reglas de Pertenencia a Grupos	144
6.3.2.2. Redes Bayesianas Aplicadas a la Ponderación de Reglas de Comportamiento o a la Pertenencia a Grupos	146
6.3.2.2.1. Ejemplo de aplicación del procedimiento cuando hay clases/grupos identificados	149
6.3.2.2.2. Ejemplo de aplicación del procedimiento cuando no hay clases/grupos identificados	152
6.4. Datos Utilizados en los Ejemplos	154
CAPITULO 7: DERIVACIÓN DEL PROCESO DE EXPLOTACIÓN DE INFORMACIÓN	157
7.1. Introducción	157
7.2. Proceso de Derivación de Modelos	157
7.2.1. Generalidades	157
7.2.2. Estructura General del Proceso de Derivación de Modelos	159
7.3. Estandarización del Modelado	164
7.3.1. Notación General de Formalismos Basados en Redes Semánticas	165
7.3.2. Notación Específica de las Redes Semánticas del Problema de Explotación de Información	167
7.3.3. Resumen de Notación de Formalismos Basados en Redes Semánticas	167
7.4. Proceso de Derivación de Modelos	168
7.4.1. Aspectos Generales	169
7.4.2. Fase Orientada al Análisis del Dominio del Negocio	171
7.4.2.1. Paso 1: "Identificación de los elementos y estructura del dominio"	171
7.4.2.1.1. Técnica Tabla Término-Categoría-Definición del Dominio	172
7.4.2.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor del Dominio	174
7.4.2.2. Paso 2: "Identificación de Relaciones entre Conceptos del Dominio"	175
7.4.2.2.1. Técnica Tabla Concepto-Relación del Dominio	176
7.4.2.3. Paso 3: "Conceptualización Del Dominio"	177
7.4.2.3.1. Técnica Red Semántica del Modelo de Negocio	177
7.4.3. Fase Orientada al Análisis del Problema de Explotación de Información	178
7.4.3.1. Paso 4: "Identificación de los elementos y estructura del Problema de Explotación de Información"	179
7.4.3.1.1. Técnica Tabla Término-Categoría-Definición del Problema de Explotación de Información	180
7.4.3.1.2. Técnica Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información	181

7.4.3.2. Paso 5: “Identificación de relaciones entre conceptos del Problema de Explotación de Información”	184
7.4.3.2.1. Técnica Tabla Concepto-Relación del Problema de Explotación de Información	185
7.4.3.3. Paso 6: “Conceptualización del Problema de Explotación de Información”	186
7.4.3.3.1. Técnica Red Semántica del Problema de Explotación de Información	186
7.4.4. Fase de Determinación del Proceso de Explotación de Información	188
7.4.4.1. Paso 7: “Derivación del Proceso de Explotación de Información”	188
7.4.4.1.1. Técnica Algoritmo de Derivación del Proceso de Explotación de Información	189
7.4.5. Resumen del Proceso de Derivación de Modelos	193
CAPÍTULO 8: TRANSFORMACIÓN DE LOS DATOS	195
8.1. Introducción	195
8.2. Descripción Del Modelo De Transformación De Datos	195
8.3. Estructura General Del Modelo	196
8.4. Actividades	197
8.4.1. Enriquecer Los Datos	197
8.4.2. Obtención Y Ejecución De Los Casos Testigo	199
8.4.3. Determinar Y Aplicar La Estructura De Los Datos	200
8.4.4. Construir El Modelo De Entrada De Datos	202
8.4.5. Inspeccionar Los Datos	203
8.5. Ejemplo: Predicción De Potenciales Clientes De Depósitos A Largo Plazo	205
8.5.1. Aplicación De La Actividad De Enriquecimiento De Los Datos	205
8.5.1.1. Paso 1: Conocer El Problema A Resolver	205
8.5.1.2. Paso 2: Analizar Solución A Obtener	206
8.5.1.3. Paso 3: Generación De Documento De Solución	206
8.5.1.4. Paso 4: Analizar Técnicas De Modelado A Utilizar	206
8.5.1.5. Paso 5: Generación De Documento De Técnicas De Modelado	207
8.5.2. Aplicación De La Actividad De Obtención Y Ejecución De Los Casos Testigo	207
8.5.2.1. Paso 1: Planteo De Los Casos Testigo	207
8.5.2.2. Paso 2: Generar Lista De Chequeo	209
8.5.2.3. Paso 3: Test De Los Datos	210
8.5.2.4. Paso 4: Documentar Conclusiones	210
8.5.3. Aplicación De La Actividad De Determinar Y Aplicar La Estructura De Los Datos	211
8.5.3.1. Paso 1: Determinar Las Fuentes De Los Datos	212
8.5.3.2. Paso 2: Determinar Las Relaciones	212
8.5.3.3. Paso 3: Unificar Tipos De Datos	213
8.5.3.4. Paso 4: Unificar Rangos De Variables	213
8.5.3.5. Paso 5: Generar Documento De Integración	213

8.5.4. Aplicación De La Actividad De Construir El Modelo De Entrada De Datos	213
8.5.4.1. Paso 1: Efectuar Análisis Iniciales	214
8.5.4.2. Paso 2: Ejecutar Las Distintas Fases De Transformación	215
8.5.4.3. Paso 3: Generar Documento De Estructuración	216
8.5.5. Aplicación De La Actividad De Inspección De Los Datos	216
8.5.5.1. Paso 1: Efectuar La Inspección De Los Datos	217
8.5.5.2. Paso 2: Actualizar El Repositorio Del Conocimiento De La Compañía	217
8.5.5.3. Paso 3: Preparar Los Datos Para El Siguiendo Paso Del Proyecto De Explotación De Información	217
CAPÍTULO 9: METRICAS PARA PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN	219
9.1. Introducción	219
9.2. Clasificación de Métricas para Proyectos de Explotación de Información	220
9.3. Métricas de Datos	220
9.3.1. Métrica de Densidad de Valores Nulos y Nivel de Compleción de la Tabla	223
9.3.1.1. Resultado del Comportamiento	223
9.3.1.2. Interpretación de las métricas	224
9.3.2. Métrica de Grado de Corrección de Datos de la Tabla	224
9.3.2.1. Resultado del Comportamiento	224
9.3.2.2. Interpretación de la métrica	225
9.3.3. Métrica de Grado de Utilidad de Atributos	225
9.3.3.1. Resultado del Comportamiento	227
9.3.3.2. Interpretación de la métrica	228
9.4. Métricas de Modelos	228
9.4.1. Modelo de Descubrimiento de Grupos	229
9.4.2. Modelo de Descubrimiento de Reglas	230
9.4.3. Modelo de Descubrimiento de Dependencias Significativas	232
9.4.4. Métrica de Exactitud del Modelo	235
9.4.4.1. Resultado del Comportamiento	236
9.4.4.2. Interpretación de la métrica	236
9.4.5. Métrica de Precisión del Modelo	237
9.4.5.1. Resultado del Comportamiento	238
9.4.5.2. Interpretación de la métrica	239
9.4.6. Métrica de Tasa de Aciertos y Tasa de Errores del Modelo	239
9.4.6.1. Resultado del Comportamiento	240
9.4.6.2. Interpretación de las métricas	241
9.4.7. Métrica de Cobertura de una Regla	242
9.4.7.1. Resultado del Comportamiento	242

9.4.7.2. Interpretación de la métrica	243
9.4.8. Métrica de Precisión de una Regla	243
9.4.8.1. Resultado del Comportamiento	244
9.4.8.2. Interpretación de la métrica	245
9.4.9. Métrica de Usabilidad de los Atributos del Modelo	245
9.4.9.1. Resultado del Comportamiento	246
9.4.9.2. Interpretación de la métrica	246
9.4.10. Métrica de Grado de Incidencia de Atributos	248
9.4.10.1. Resultado del Comportamiento	249
9.4.10.2. Interpretación de la métrica	250
9.5. Métricas de Proyectos	251
9.5.1. Métrica de Éxito de Resultados del Proceso de Desarrollo	252
9.5.1.1. Resultado del Comportamiento	255
9.5.1.2. Interpretación de la métrica	256
9.5.2. Métrica de Desvío del Esfuerzo del Proyecto	256
9.5.2.1. Resultado del Comportamiento	257
9.5.2.2. Interpretación de la métrica	258
APÉNDICE: INTRODUCCIÓN A BIG DATA	261
A.1. Introducción	261
A.2. Ciclo de vida del Big Data	262
A.3. Arquitectura Lambda	264
A.4. Propiedades de un Modelo de Big Data	265
A.5. Otras consideraciones relacionadas al modelo de Big Data	266
A.6. Tecnologías Asociadas al Modelo de Big Data	267
REFERENCIAS	269