

# Índice de Contenidos

1.	Introducción.....	10
1.1	Descripción y justificación del proyecto .....	10
1.2	Objetivos.....	12
1.2.1	Objetivo General.....	12
1.2.2	Objetivos Específicos .....	12
1.3	Contribuciones.....	12
1.4	Metodología.....	12
1.4.1	Definir los requerimientos de información.....	12
1.4.2	Establecer indicadores .....	13
1.4.3	Proceso ETL .....	13
1.4.4	Análisis y diseño del modelo de datos.....	13
1.4.5	Web Mining.....	13
1.4.6	Implementación y visualización de indicadores .....	13
1.4.7	Metodología de web mining Crips-DM.....	14
1.5	Alcances.....	16
1.6	Descripción del trabajo.....	17
2.	Marco Conceptual.....	18
2.1	Introducción a la Web.....	18
2.1.1	La Web .....	18
2.1.2	Funcionamiento de la Web.....	19
2.2	Sistemas de almacenamiento, procesamiento y análisis de datos .....	20
2.2.1	Base de datos .....	21
2.2.2	Sistemas de información.....	21
2.2.3	Modelo Relacional.....	23
2.2.4	Data Warehouse.....	24
2.2.5	Modelo Multidimensional .....	28
2.2.6	Arquitectura para Data Warehouse.....	31
2.3	Repositorio de información para la web.....	36
2.4	Web Mining.....	38
2.4.1	Tipos de datos originados en la Web.....	38
2.4.2	Web Logs.....	39
2.4.3	Procesamientos de los contenidos de documentos .....	42
2.4.4	Proceso KDD .....	44
2.5	Análisis multidimensional .....	45
2.5.1	Operación Drill-Down.....	52
2.5.2	Operación Drill-up.....	52
2.5.3	Operación Dill-across .....	53
2.5.4	Roll-across .....	53
2.5.5	Pivot.....	54
2.6	Análisis de comportamiento de los usuarios .....	54
3.	Metodología y diseño del webhouse .....	58
3.1	Línea base .....	58
3.2	Requerimientos.....	63
3.2.1	Requerimientos de información.....	63

3.2.2	Conociendo el portal.....	63
3.3	Indicadores.....	65
3.4	Fuentes de datos.....	66
3.5	Modelo del Data Staging Area (DSA).....	66
3.6	El repositorio de información.....	68
3.7	Arquitectura tecnológica.....	73
3.7.1	Capa de datos.....	74
3.7.2	Capa media.....	75
3.7.3	Capa de presentación.....	77
4.	Construcción del data mart.....	79
4.1	Descripción del sitio.....	79
4.2	Proceso ETL.....	81
4.2.1	Extracción y transformación de datos.....	81
4.3	Carga a la DSA.....	86
4.4	Creación de cubos de información.....	93
4.5	Análisis de comportamiento de los usuarios.....	100
4.5.1	Sesiones consideradas.....	100
4.5.2	Contenido de las páginas.....	100
4.5.3	Transformación y agrupamiento de contenidos.....	103
4.6	Interfaz de usuario.....	107
5.	Resultados.....	110
5.1	Indicadores obtenidos.....	110
5.2	Resultado de la aplicación de filtros a los web logs.....	118
5.3	Resultados del uso del algoritmo K-means.....	119
6.	Conclusiones.....	124
	Bibliografía.....	127
	Anexos.....	130
	A: Tablas de resultado del clustering de contenidos.....	130
	B: Resultado Bajo Interfaz Rapidminer.....	131
	C: Interfaz de aplicación.....	134
	D: Indicadores de mes de septiembre.....	136