

**CMIN 1.0 – Herramienta CASE para el soporte de proyectos de
Minería de Datos basado en CRISP-DM**

ANEXOS



Anteproyecto de Trabajo de Grado

**Jhon Emmanuel Zuñiga Paredes
Juan Carlos Guarín Walteros**

Director: MSc. Carlos Alberto Cobos Lozada

**Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Sistemas
Grupo de I+D en Tecnologías de la Información
Gestión de la Información – Minería de Datos
Popayán, Abril de 2009**

Tabla de Contenido

LISTA DE FIGURAS.....	3
LISTA DE TABLAS.....	3
ANEXO A. CASOS DE USO.....	4
1 CASOS DE USO ALTO NIVEL	5
1.1 CASO DE USO GENERAL DEL USUARIO	5
1.2 CASOS DE USO DE USUARIOS EXPERTOS	5
2 CASOS DE USO REALES DEL USUARIO.....	6
ANEXO B. DIAGRAMAS	19
3 DIAGRAMA FISICO DEL SISTEMA.....	21
3.1 LISTA DE TABLAS.....	22
4 DIAGRAMAS DE CLASES DEL SISTEMA	24
4.1 LISTA DE CLASES.....	25
5 DIAGRAMA DE CLASES DEL MODULO DE WORK FLOW	26
5.1 LISTA DE CLASES DEL MODULO DE WORK FLOW	26
ANEXO C. POSTER CIENTÍFICO PRESENTADO EN: DEMOFEST MICROSOFT RESEARCH ACADEMIC SUMMIT.....	28
6 POSTER CIENTÍFICO	29
ANEXO D. REPORTES: EVALUACIÓN DE PROTOTIPO CMIN POR ESTUDIANTES DE LA ELECTIVA MINERÍA DE DATOS.....	30
7 RESULTADOS DE LA PRUEBA.....	31
7.1 FASE 1	31
7.2 FASE 2	38
7.3 FASE 3	41
7.4 FASE 4	45
7.5 FASE 5	54
7.6 FASE 6	56
8 ANEXO E. DISEÑO DE LAS EVALUCIONES UTILIZADAS EN LA REALIZACION DE LA PRUEBA BETA.....	59
9 ARTICULO	64

Lista de Figuras

FIGURA 1. CASO DE USO GENERAL DEL SISTEMA.	5
FIGURA 2. CASOS DE USO DE USUARIOS EXPERTOS.	6
FIGURA 3. DIAGRAMA FÍSICO DE LA BASE DE DATOS.	21
FIGURA 4. DIAGRAMA DE CLASES DE LOS MÓDULOS DE GESTIÓN DE PROCESOS, PROYECTOS Y PLANTILLAS.	24
FIGURA 5. DIAGRAMA DE CLASES DEL MÓDULO DE WORK FLOW.	26
FIGURE 6. PÓSTER CIENTÍFICO	29

Lista de Tablas

TABLA 1. CASO DE USO REAL ENTRAR AL SISTEMA.....	8
TABLA 2. CASO DE USO REAL CREAR PROCESOS	10
TABLA 3. CASO DE USO REAL ADICIONAR PASOS.....	11
TABLA 4. CASO DE USO REAL CREAR CAMPOS.....	14
TABLA 5. CASO DE USO REAL CREAR PROYECTOS.	17
TABLA 6. CASO DE USO REAL DESARROLLAR PROYECTO.	18

ANEXO A. CASOS DE USO

1 CASOS DE USO ALTO NIVEL

1.1 Caso de Uso General del Usuario

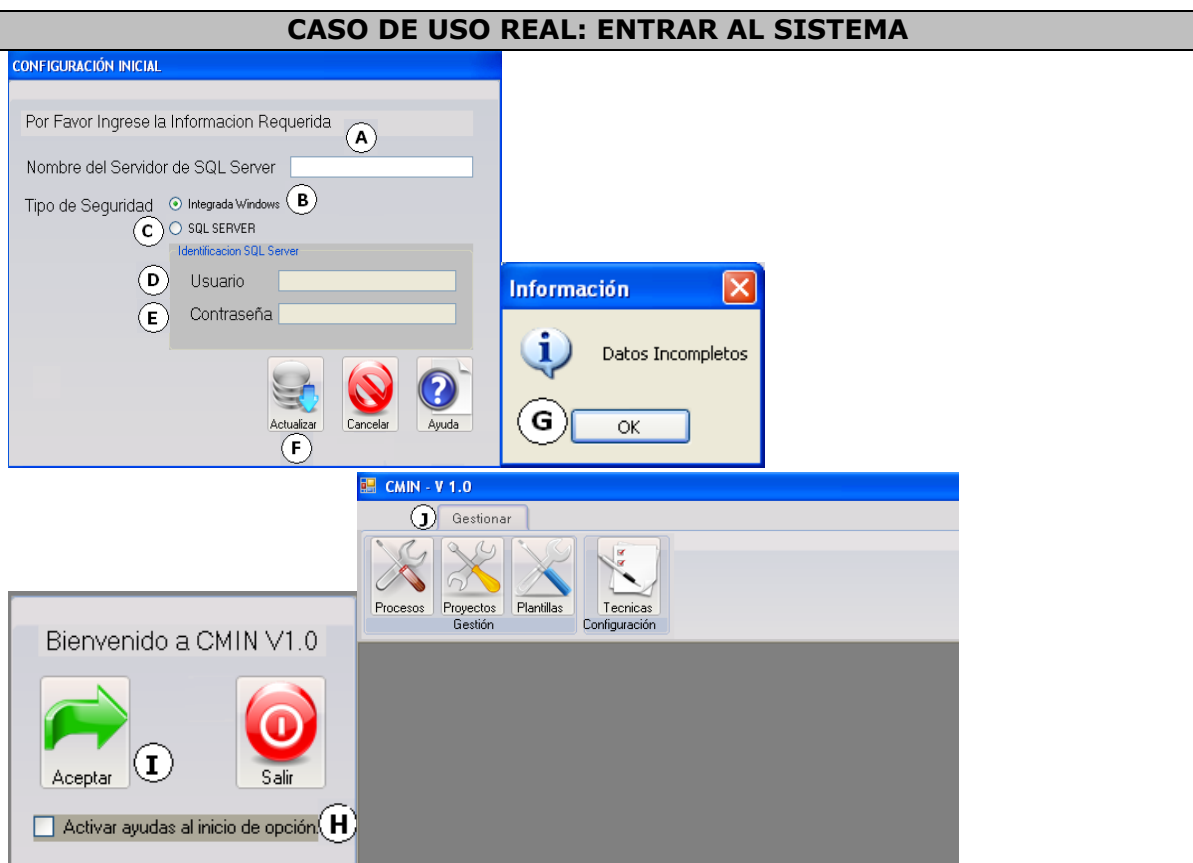
La Figura 1. Caso de Uso General del Sistema.

1.2 Casos de Uso de Usuarios Expertos

La Figura 2. Casos de Uso de Usuarios Expertos.

2 CASOS DE USO REALES DEL USUARIO

CASO DE USO REAL: ENTRAR AL SISTEMA



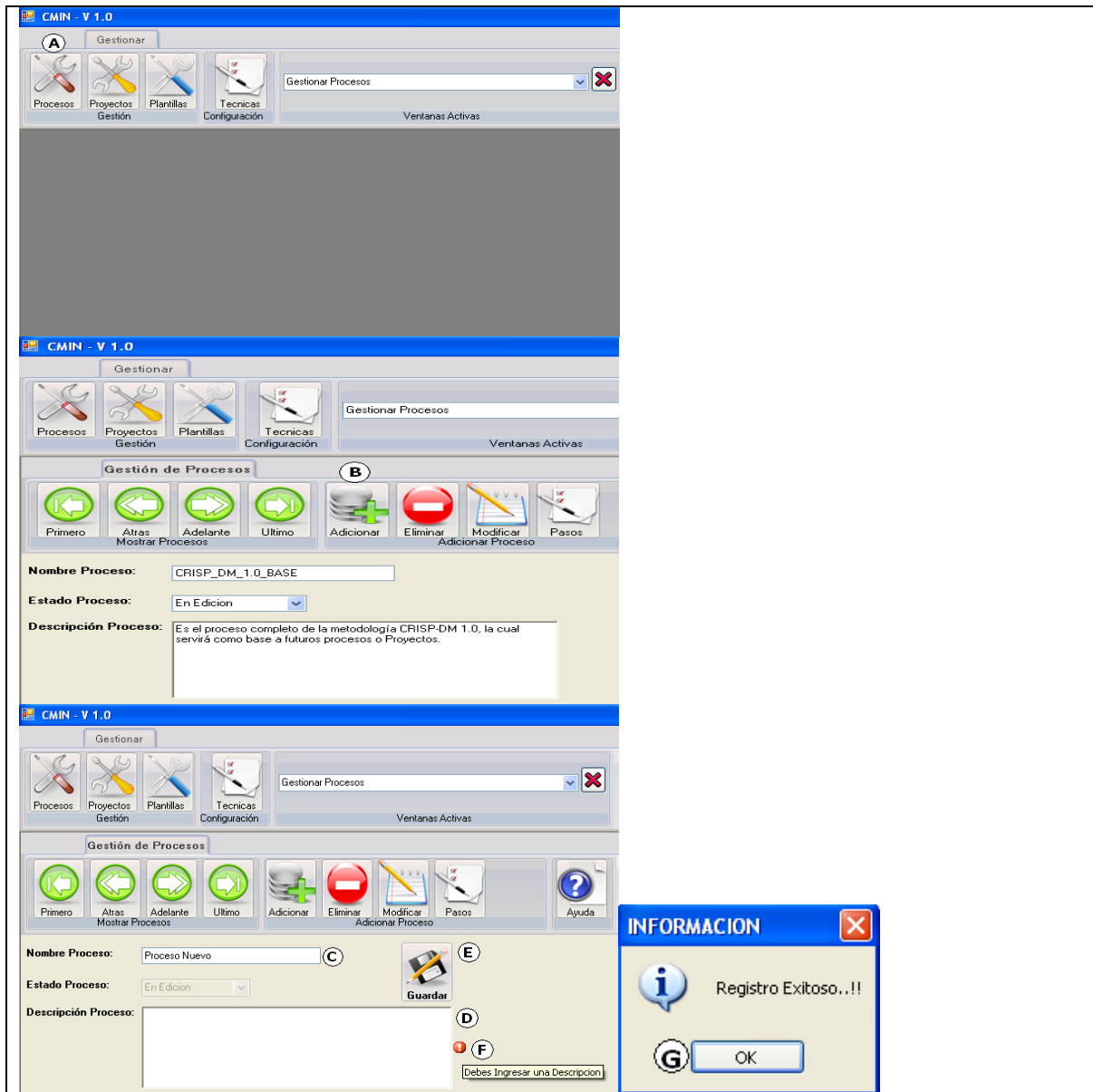
CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario digita el nombre del servidor de SQL Server [A].	
2. El usuario selecciona la opción de seguridad Integrada de Windows [B].	3. El sistema deshabilita los campos para usuario [D] y contraseña [E] de seguridad SQL Server.

4. El usuario selecciona la opción de seguridad SQL Server [C].	5. El sistema habilita los campos para usuario [D] y contraseña [E] de seguridad SQL Server.
6. El usuario digita el usuario [D].	
7. El usuario digita la contraseña [E].	
8. El usuario da clic en el botón actualizar [F].	9. El sistema valida que con la información suministrada se pueda ingresar al sistema.
	10. El sistema despliega la ventana de bienvenida de CMIN 1.0.
11. El usuario debe seleccionar o no la opción "Activar ayudas al inicio de opción." [H].	12. Si es seleccionado el sistema mostrara automáticamente las ayudas del sistema cuando se ejecute una opción de lo contrario no.
13. El usuario da clic en el botón Aceptar [I].	14. El sistema muestra la ventana principal del sistema con una pestaña con las acciones disponibles para todos los usuarios [J].
CURSO ALTERNO	
Acción del actor	Respuesta del sistema
	9. El usuario no digito la información obligatoria en los campos [D]y [E]. Se le informa al usuario que los datos están incompletos [G].
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 1. Caso de uso real Entrar al sistema.

CASO DE USO REAL: CREAR PROCESOS

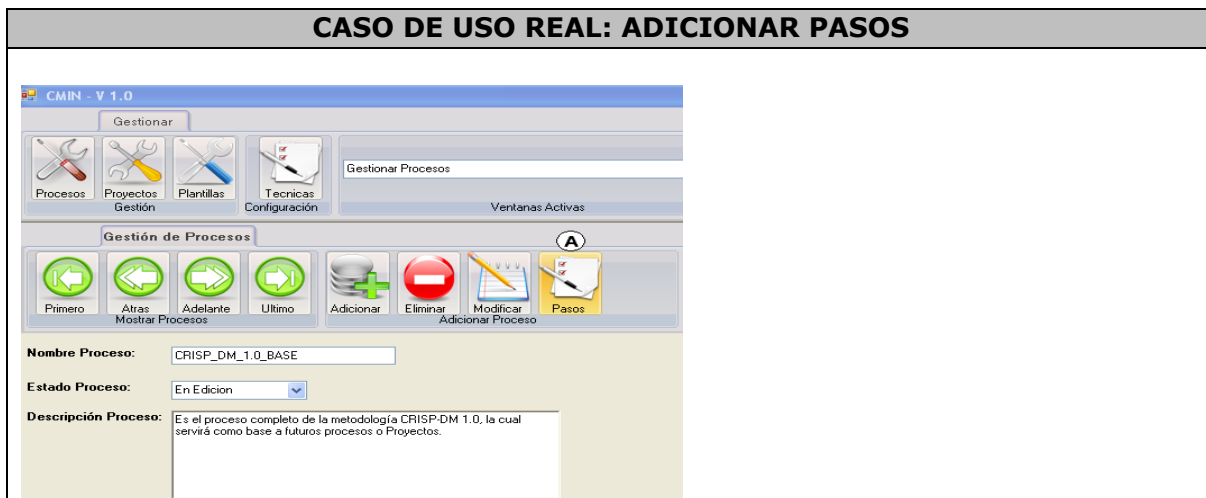


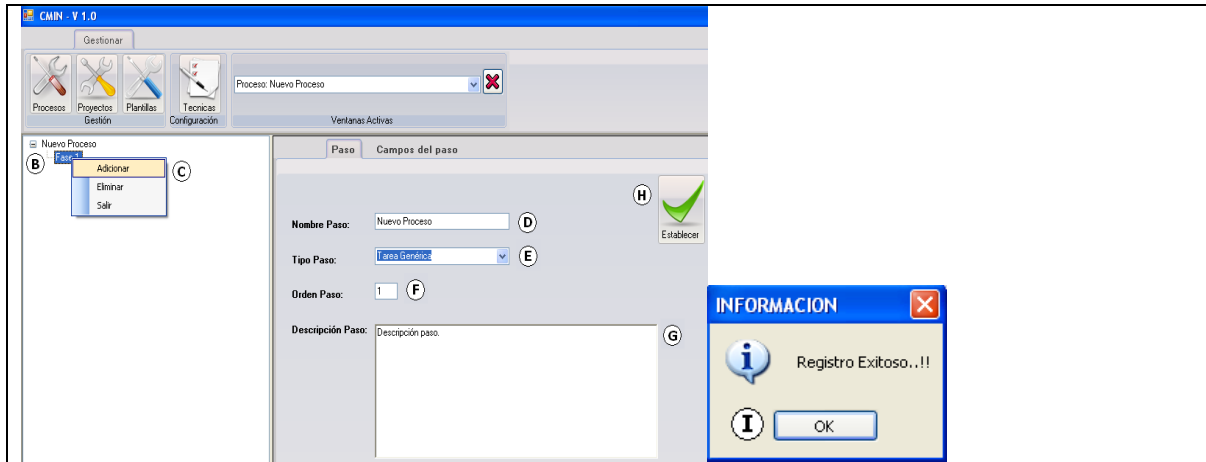
CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Procesos [A].	2. El sistema muestra la ventana de gestión de procesos.
3. El usuario da clic en el botón Adicionar [B].	4. El sistema limpia los controles para crear un nuevo proceso [C] y [D], además muestra el botón Guardar [E].
5. El usuario digita el nombre del proceso nuevo	

[C].	
6. El usuario digita la descripción del proceso nuevo [D].	
7. El usuario da clic en el botón Guardar [E].	8. El sistema valida que los campos estén digitados, después guarda la información general del nuevo proceso en la base de datos y muestra la ventana de éxito.
9. El usuario da clic en el botón OK [G].	10. El sistema termina la rutina.
CURSO ALTERNO	
Acción del actor	Respuesta del sistema
	8. El usuario no digito la información obligatoria en los campos [C]y [D]. Se le informa al usuario que los datos están incompletos [F].
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 2. Caso de so Real Crear Procesos





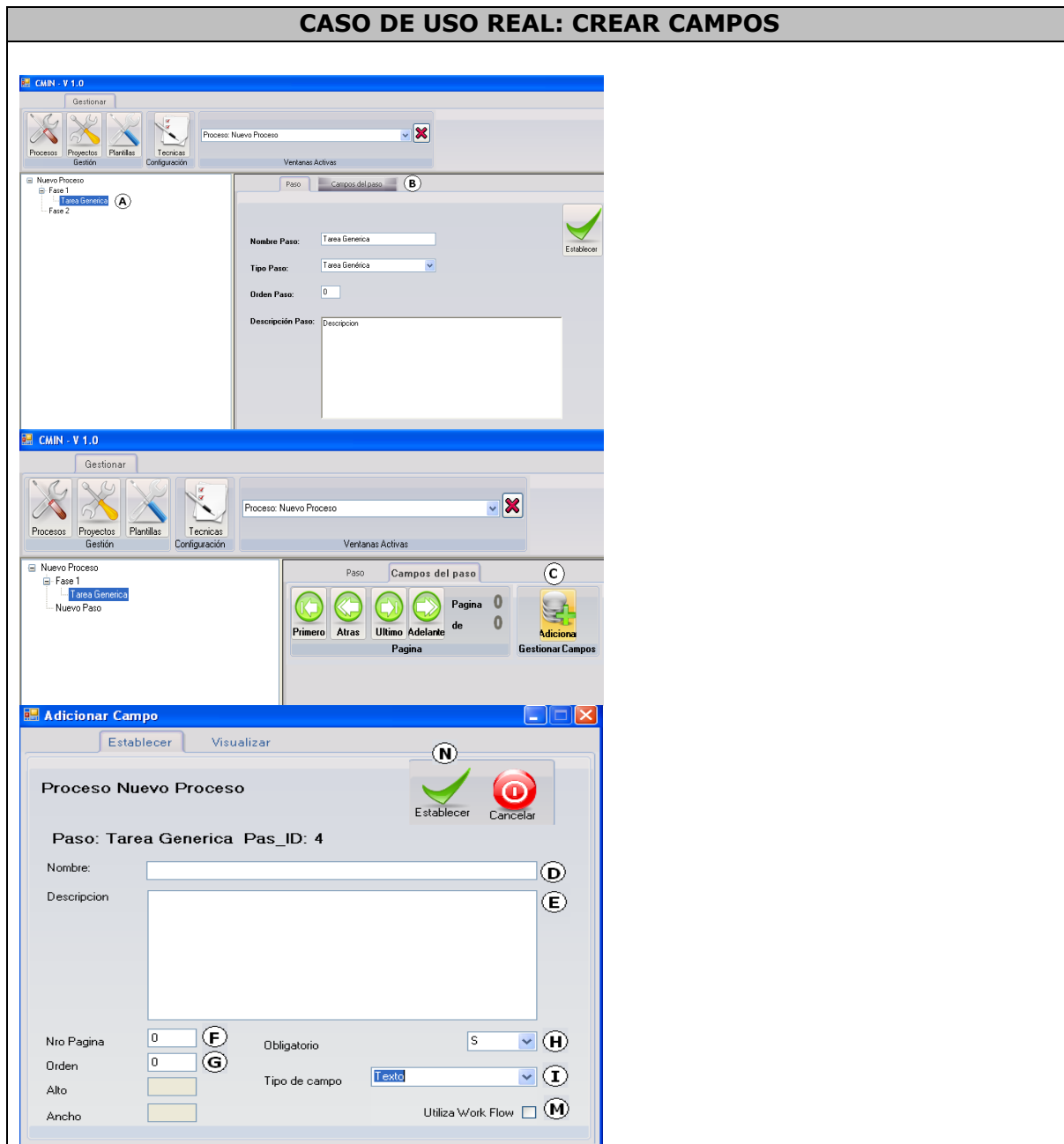
CURSO NORMAL DE LOS EVENTOS

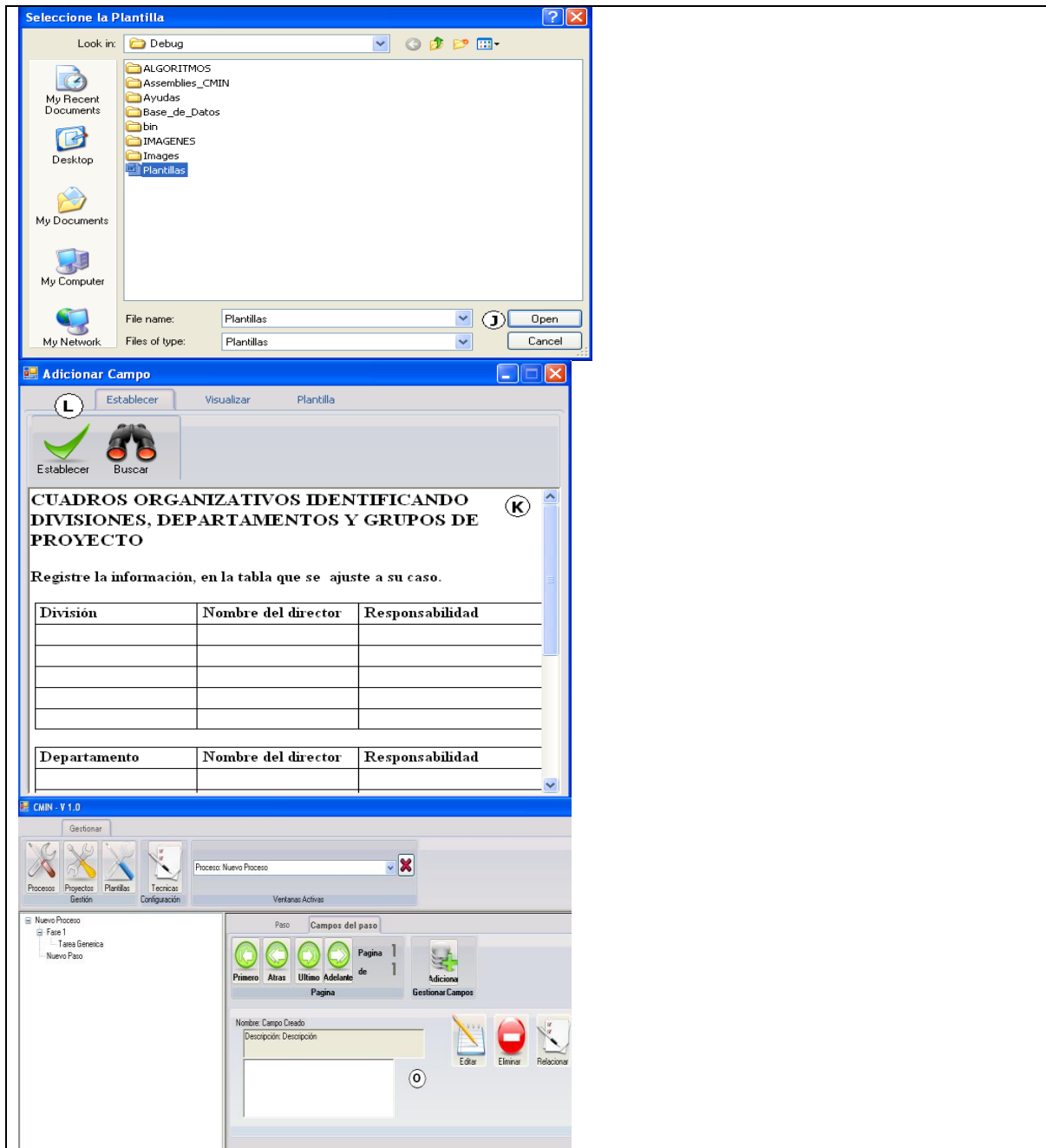
Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Pasos [A].	2. El sistema muestra la ventana de gestión de pasos.
3. El usuario da clic derecho en paso del árbol donde se ira creando la estructura del proceso [B].	4. El sistema despliega el menú para la modificación de la estructura del proceso.
5. El usuario da clic en la opción Adicionar [C].	6. El sistema limpia los controles donde ira la información general del paso [F] y [G], además en el campo del nombre [D] por defecto le pone "Nuevo Campo".
7. El usuario digita el nombre del paso nuevo [D].	
7. El usuario escoge el tipo de paso a crear [E].	
8. El usuario especifica el orden del paso [F].	
9. El usuario digita la descripción del paso [G].	
10. El usuario da clic en el botón establecer [H].	11. El sistema valida que la información este digitada y muestra la ventana de éxito.
12. El usuario da clic en el botón OK [I].	12. El sistema termina la rutina.

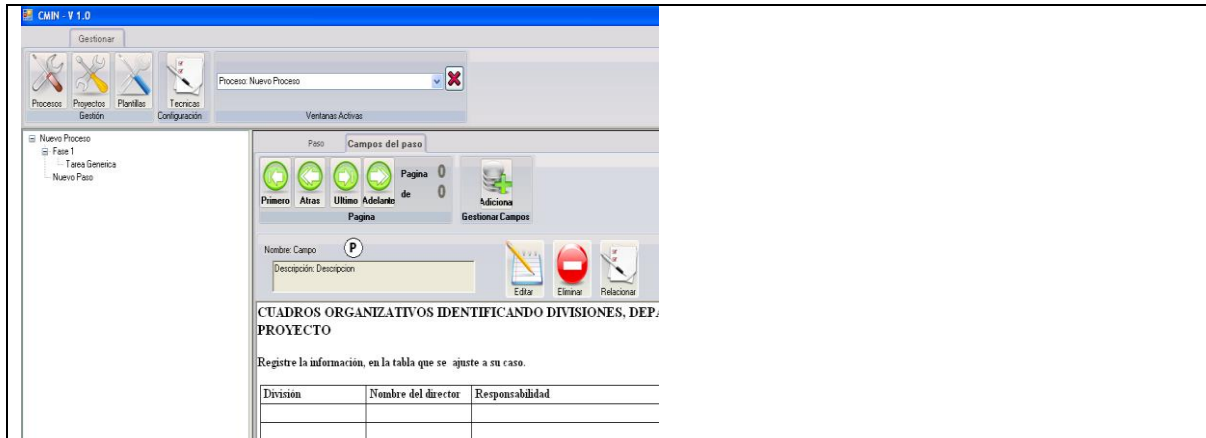
TRAZABILIDAD

Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 3. Caso de Uso Real Adicionar Pasos.



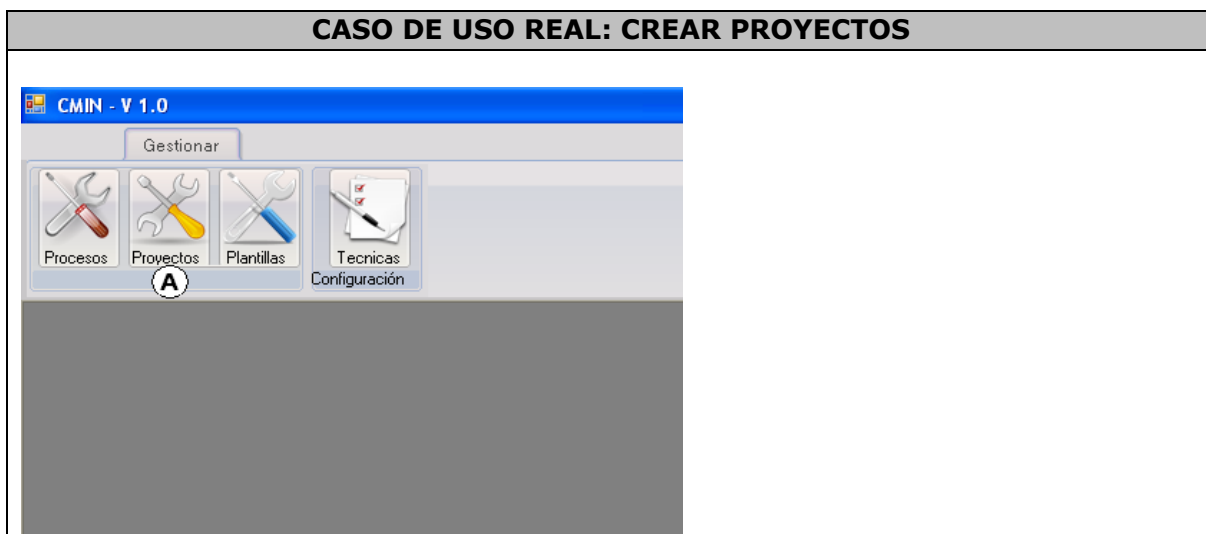


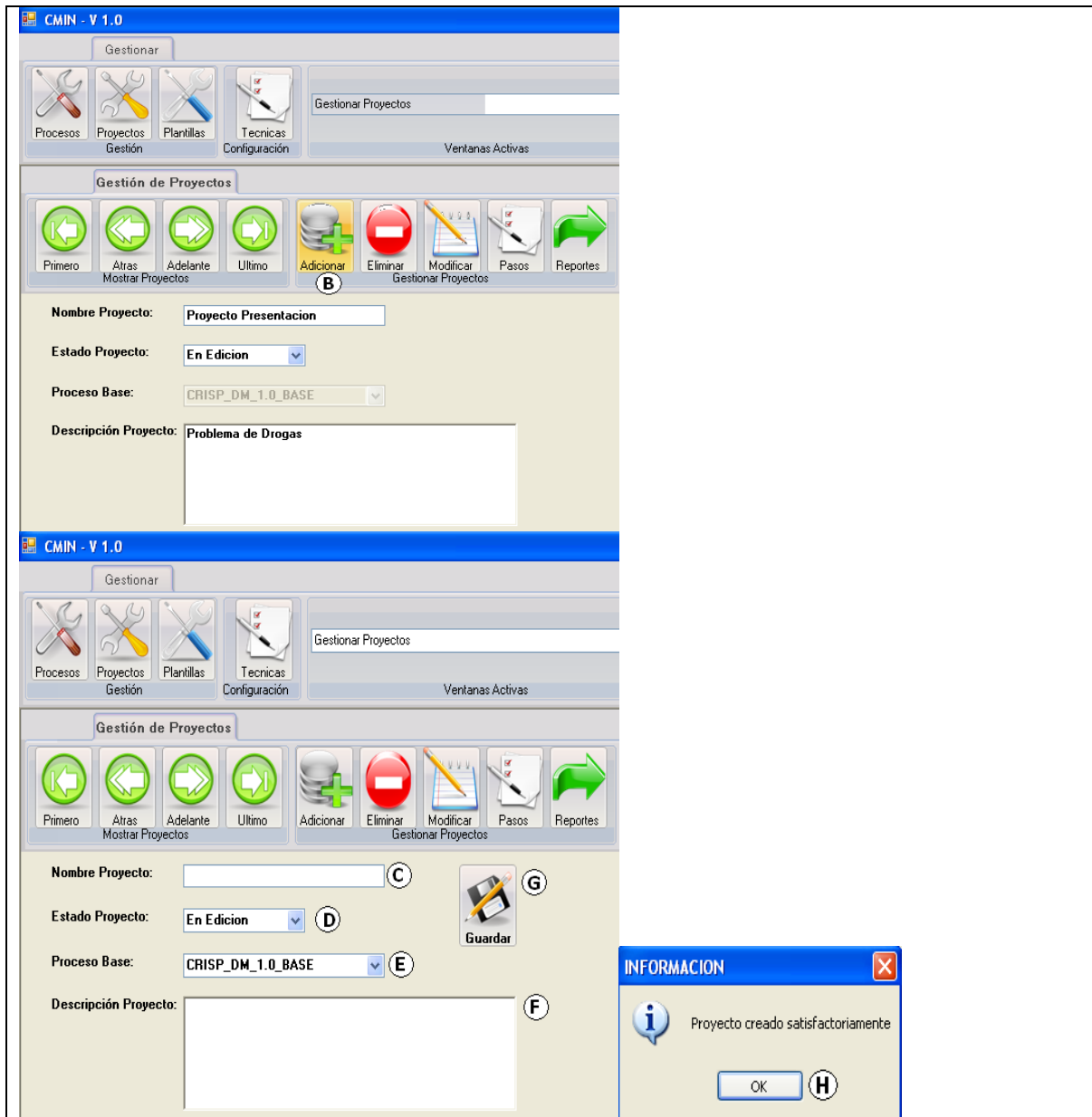


CURSO NORMAL DE LOS EVENTOS	
Acción del actor	Respuesta del sistema
1. El usuario selecciona un paso [A].	
2. El usuario da clic en el tab de "Campos del Paso" [B].	3. El sistema presenta los controles de navegación para ver todos los campos creados en paso y el botón de adicionar [C].
4. El usuario clic en el botón Adicionar [C].	5. El sistema presenta la ventana para crear un nuevo campo.
6. El usuario digita el nombre del campo [D].	
7. El usuario digita la descripción del campo [E].	
8. El usuario digita el numero de la pagina en donde se mostrara el campo [F].	
9. El usuario digita el orden en el que se presentara el campo en la pagina [G].	
10. El usuario escoge si el campo es obligatorio o no [H].	
11. El usuario escoge el tipo de campo a crear [I].	
12. El usuario especifica si el campo podrá utilizar el work flow al seleccionar la caja de chequeo [M].	
13. El usuario da clic en el botón "Establecer" [N].	14. El sistema valida que los datos estén digitados.
	15. El sistema muestra el campo creado [O] y se termina la rutina.
CURSO ALTERNO	
Acción del actor	Respuesta del sistema
11. El usuario escoge el tipo de campo plantilla	12. El sistema presenta la ventana

[I].	para escoger la plantilla.
13. El usuario escoge la plantilla y da clic en el botón "Open" [J].	14. El sistema presenta la previsualización de la plantilla para verificar que es la correcta [K].
14. El usuario da clic en el botón establecer para relacionar la plantilla al campo [L].	15. El sistema muestra de nuevo el tab donde se ingresa la información general del campo.
16. El usuario da clic en el botón "Establecer" [N].	17. El sistema valida que los datos estén digitados.
	18. El sistema muestra el campo creado [P] y se termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 4. Caso de Uso Real Crear Campos.





CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Proyectos [A].	2. El sistema muestra la ventana de gestión de proyectos.
3. El usuario da clic en el botón Adicionar [B].	4. El sistema limpia los controles para crear un nuevo proyecto [C], [D], [E] y [F], además muestra el botón Guardar [G].

5. El usuario digita el nombre del proyecto nuevo [C].	
6. El usuario escoge si el proyecto esta en edición o terminado [D].	
7. El usuario escoge el proceso del que se basara su desarrollo [E].	
8. El usuario digita la descripción del proyecto [F].	
9. El usuario da clic en el botón "Guardar" [G].	10. El sistema valida que los datos estén digitados y muestra la ventana de éxito.
11. El usuario da clic en el botón OK [H].	12. El sistema termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 5. Caso de Uso Real Crear Proyectos.

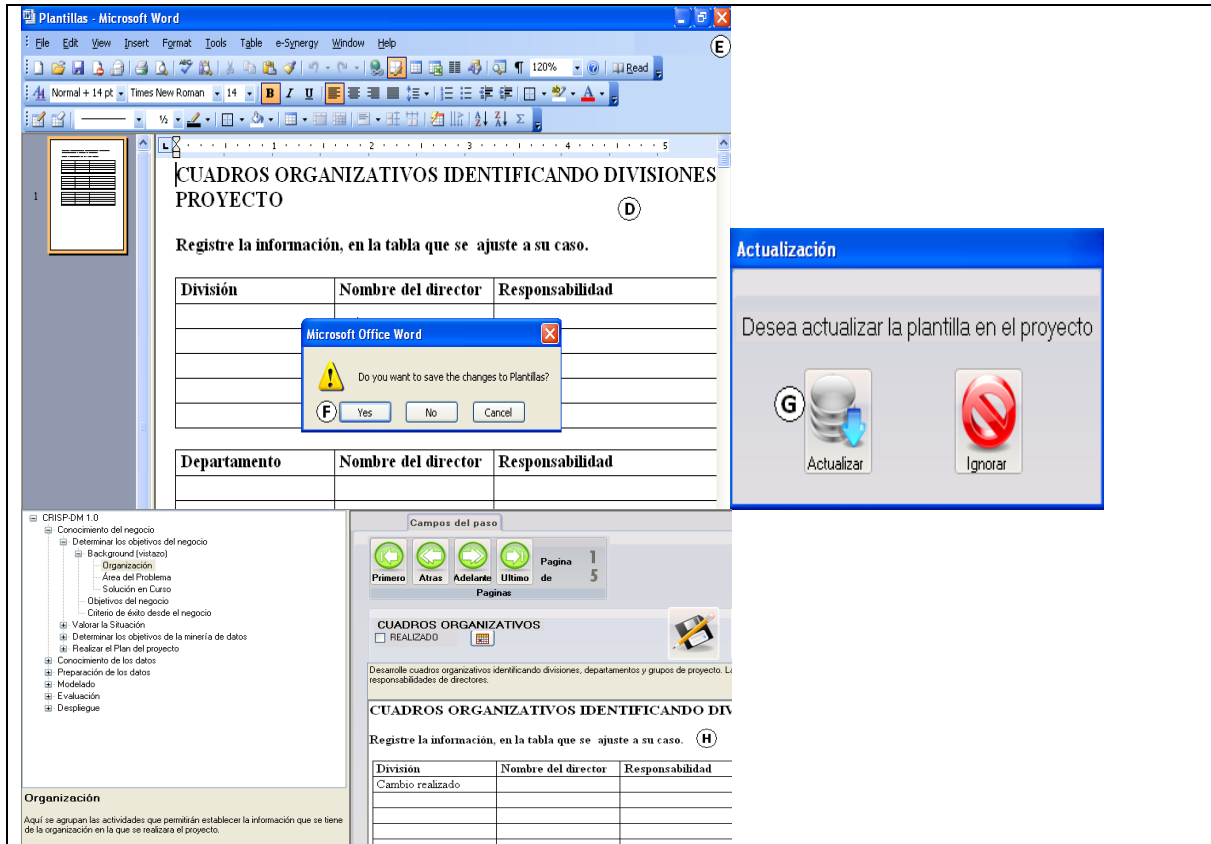
CASO DE USO REAL: DESARROLLAR PROYECTO

The screenshot displays the CMIN 1.0 software interface. The main window is titled 'CMIN - V 1.0' and features a 'Gestionar' menu. The toolbar includes icons for 'Procesos', 'Proyectos', 'Plantillas', and 'Técnicas'. The main workspace is divided into several sections:

- Left Sidebar:** A tree view showing the project structure under 'CRISPDM 1.0', including 'Conocimiento del negocio', 'Determinar los objetivos del negocio', 'Background (inicial)', 'Organización', 'Área del Problema', 'Solución en Curso', 'Objetivos del negocio', 'Criterio de éxito desde el negocio', 'Validar la Situación', 'Determinar los objetivos de la minería de datos', 'Realizar el Plan del proyecto', 'Conocimiento de los datos', 'Preparación de los datos', 'Modelado', 'Evaluación', and 'Despliegue'.
- Central Area:** Contains 'Campos del paso' (Step Fields) with navigation buttons (Primero, Atras, Adelante, Ultimo) and a 'Paginas' section (Página 1 de 5). Below this is the 'CUADROS ORGANIZATIVOS' section, which is currently active and shows a 'REALIZADO' status. The text below reads: 'Desarrolle cuadros organizativos identificando divisiones, departamentos y grupos de proyecto, responsabilidades de directores.'
- Right Sidebar:** Contains a 'Paginas' section with a 'C' icon.

At the bottom of the screenshot, there is a section labeled 'B' titled 'CUADROS ORGANIZATIVOS IDENTIFICANDO DI'. It includes the instruction: 'Registre la información, en la tabla que se ajuste a su caso.' Below this is a table with the following structure:

Division	Nombre del director	Responsabilidad



CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario el paso que desea modificar [A].	2. El sistema muestra la ventana que contiene los campos del paso.
3. El usuario modifica el campo si es de texto, si es plantilla la visualiza [B].	
5. El usuario da clic en el botón "Guardar" [C].	6. Si el campo es una plantilla el sistema ejecuta WORD con la plantilla que se desea modificar.
7. El usuario modifica la plantilla según lo necesitado [D].	
8. El usuario cierra WORD después de modificar [E].	9. El sistema de WORD muestra la ventana que pregunta que si se guardaran los cambios si no se ha guardado antes.
10. El usuario da clic en el botón "Si o Yes" [F].	11. El sistema muestra la ventana que pregunta si se guardaran los cambios en el sistema.
12. El usuario da clic en el botón "Actualizar"	13. El sistema muestra la

[G].	información de cómo quedo guardado en el sistema la plantilla [H] y termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 6. Caso de Uso Real Desarrollar Proyecto.

ANEXO B. DIAGRAMAS muestra los casos de uso u operaciones que los actores expertos del sistema pueden hacer. En el ambiente computacional, los actores expertos pueden agregar dinámicamente DLL'S y gestionar las plantillas. Al agregar dinámicamente las DLL's los expertos gestionan la tareas de minería creándolas, modificándolas y eliminándolas. Al crear o modificar tareas de minería se agregan las DLL'S que implementan un objeto del work flow. Al gestionar plantillas se crean, modifican y eliminan. Los expertos al crear o modificar una plantilla basados en la experiencia especializan un proceso en un tema de proyectos específico, eliminando los pasos que no cumplen con dicho tema.

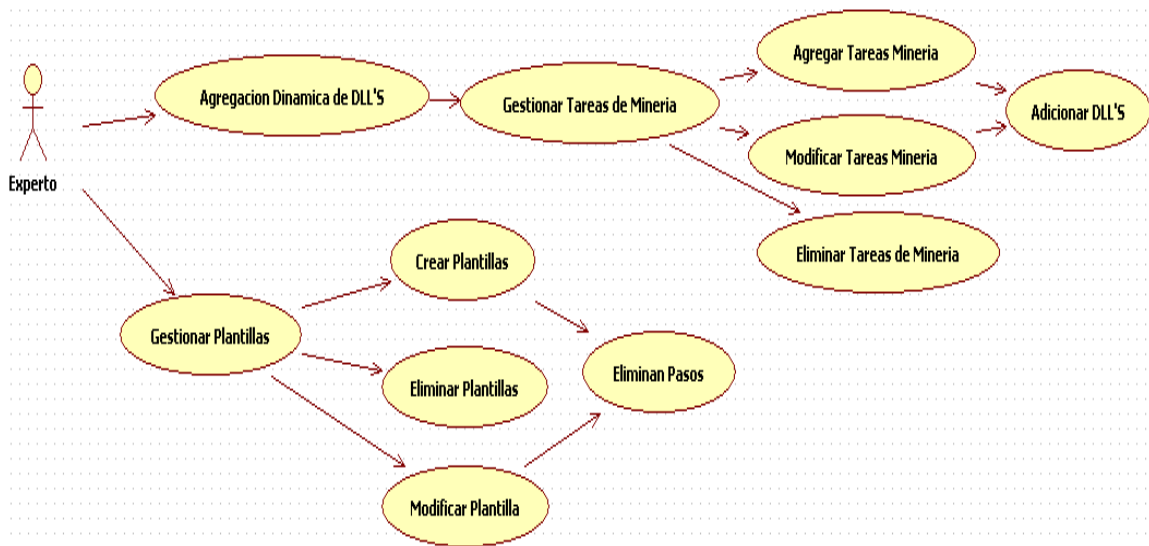
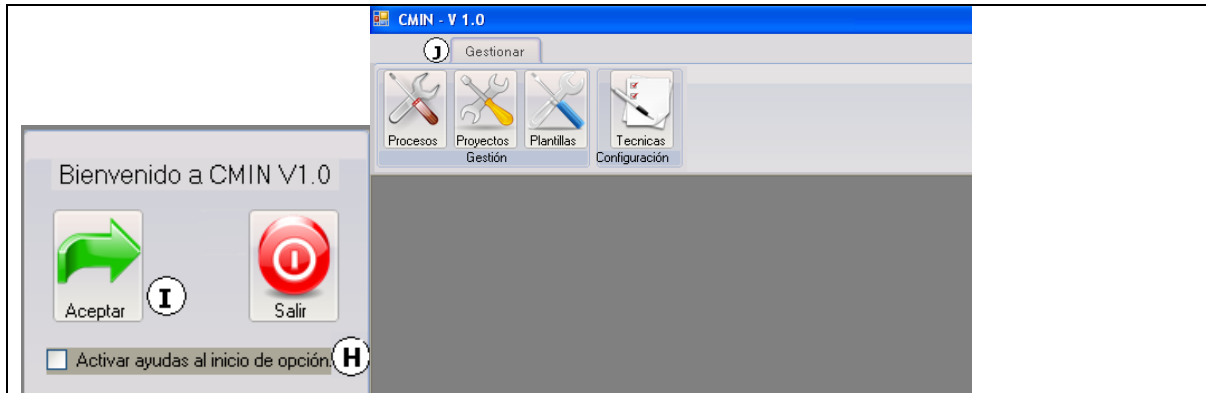


Figura 2. Casos de Uso de Usuarios Expertos.

3 CASOS DE USO REALES DEL USUARIO

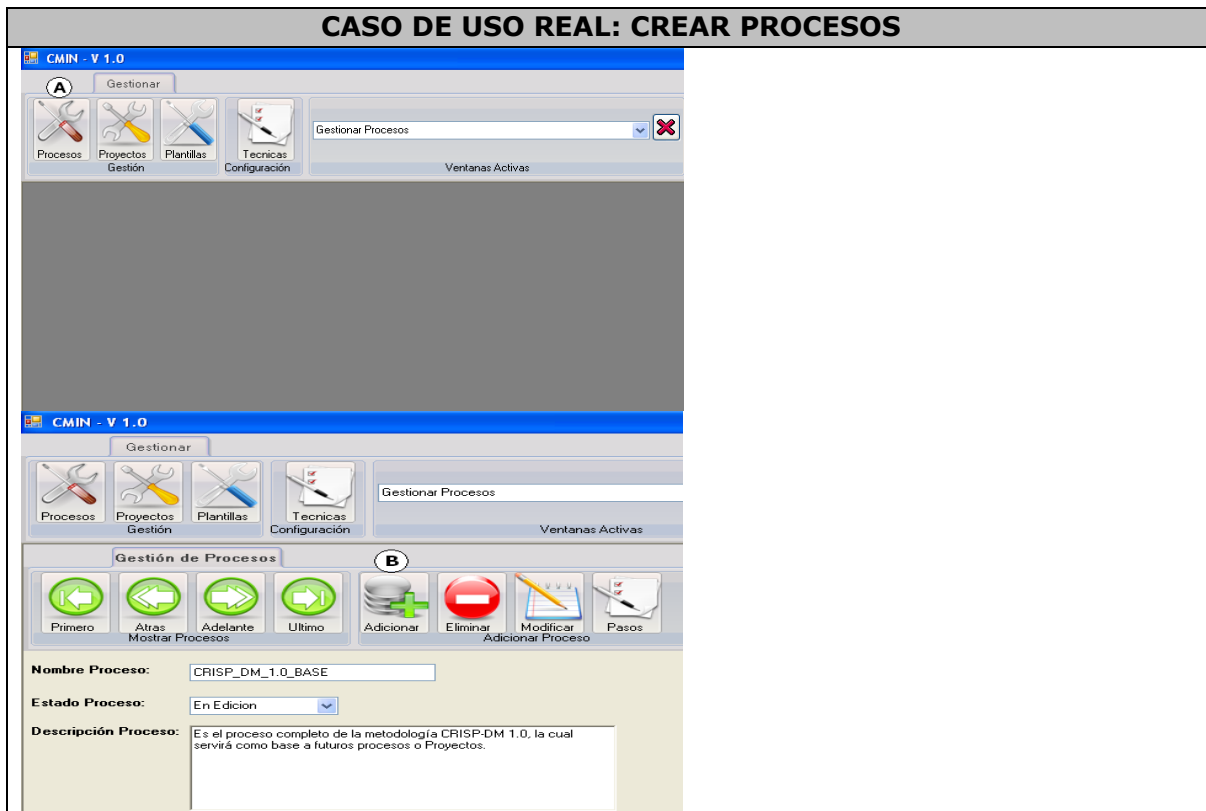


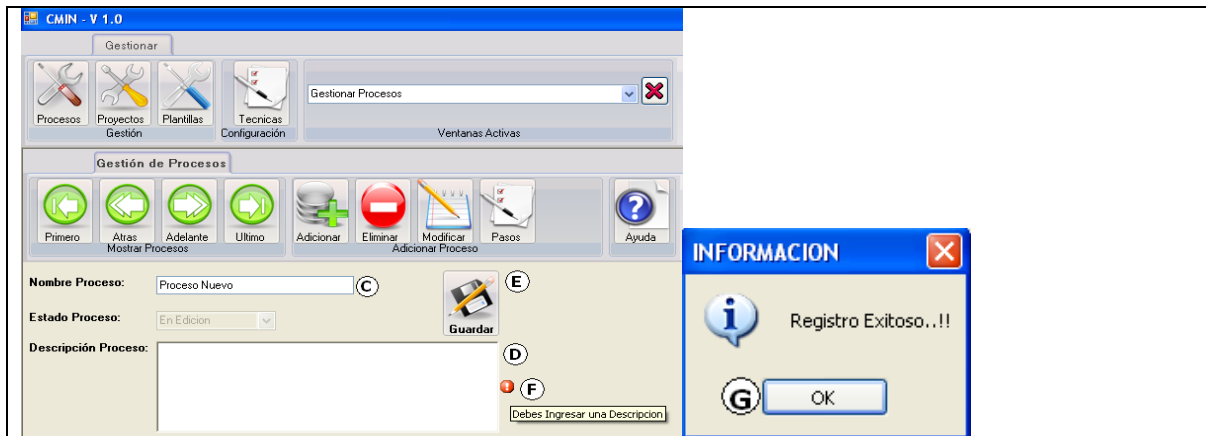


CURSO NORMAL DE LOS EVENTOS	
Acción del actor	Respuesta del sistema
15. El usuario digita el nombre del servidor de SQL Server [A].	
16. El usuario selecciona la opción de seguridad Integrada de Windows [B].	17. El sistema deshabilita los campos para usuario [D] y contraseña [E] de seguridad SQL Server.
18. El usuario selecciona la opción de seguridad SQL Server [C].	19. El sistema habilita los campos para usuario [D] y contraseña [E] de seguridad SQL Server.
20. El usuario digita el usuario [D].	
21. El usuario digita la contraseña [E].	
22. El usuario da clic en el botón actualizar [F].	23. El sistema valida que con la información suministrada se pueda ingresar al sistema.
	24. El sistema despliega la ventana de bienvenida de CMIN 1.0.
25. El usuario debe seleccionar o no la opción "Activar ayudas al inicio de opción." [H].	26. Si es seleccionado el sistema mostrara automáticamente las ayudas del sistema cuando se ejecute una opción de lo contrario no.
27. El usuario da clic en el botón Aceptar [I].	28. El sistema muestra la ventana principal del sistema con una pestaña con las acciones disponibles para todos los usuarios [J].
CURSO ALTERNO	
Acción del actor	Respuesta del sistema

	9. El usuario no digito la información obligatoria en los campos [D]y [E]. Se le informa al usuario que los datos están incompletos [G].
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 1. Caso de uso real Entrar al sistema.





CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Procesos [A].	2. El sistema muestra la ventana de gestión de procesos.
3. El usuario da clic en el botón Adicionar [B].	4. El sistema limpia los controles para crear un nuevo proceso [C] y [D], además muestra el botón Guardar [E].
5. El usuario digita el nombre del proceso nuevo [C].	
6. El usuario digita la descripción del proceso nuevo [D].	
7. El usuario da clic en el botón Guardar [E].	8. El sistema valida que los campos estén digitados, después guarda la información general del nuevo proceso en la base de datos y muestra la ventana de éxito.
9. El usuario da clic en el botón OK [G].	10. El sistema termina la rutina.

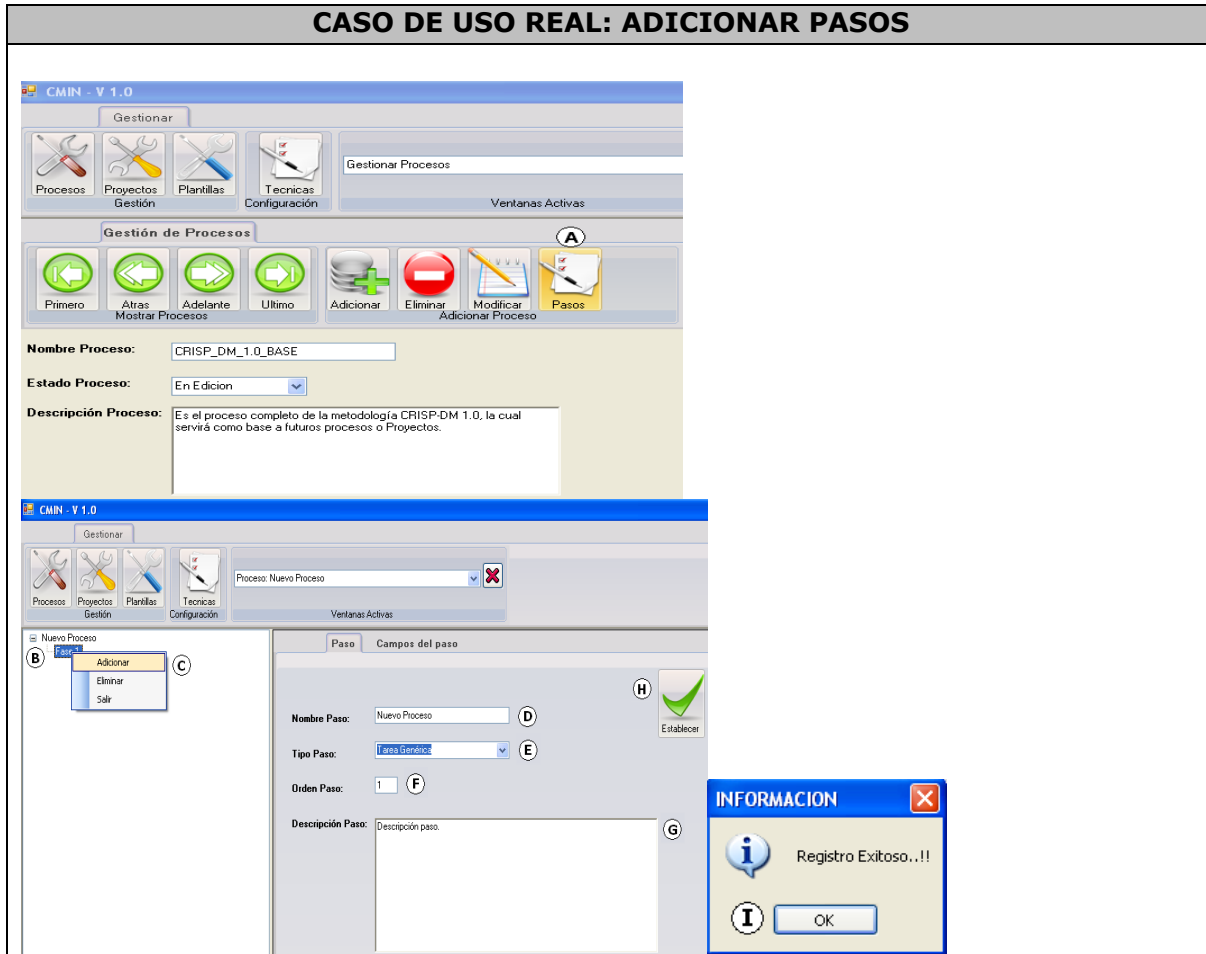
CURSO ALTERNO

Acción del actor	Respuesta del sistema
	8. El usuario no digito la información obligatoria en los campos [C]y [D]. Se le informa al usuario que los datos están incompletos [F].

TRAZABILIDAD

Artefactos anteriores	Artefactos posteriores
Artefactos del Análisis: Diagrama general de casos de uso usuario.	

Tabla 2. Caso de so Real Crear Procesos

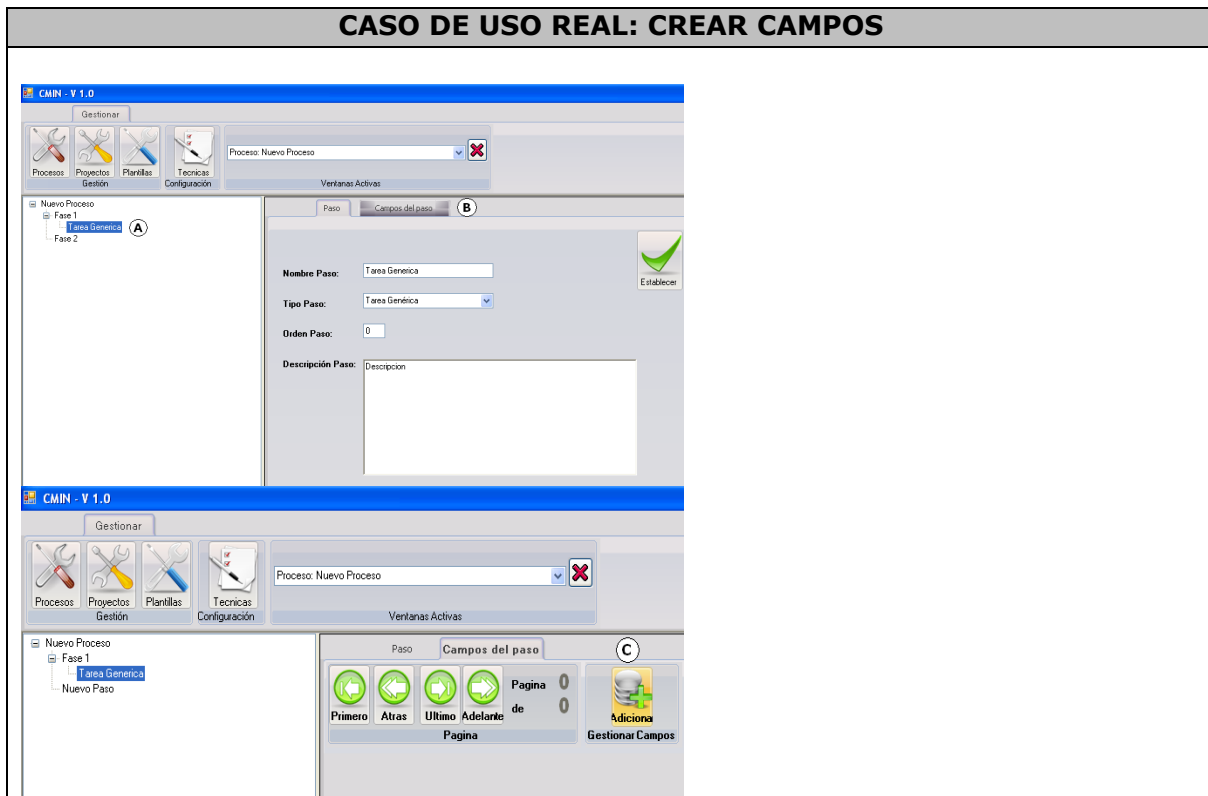


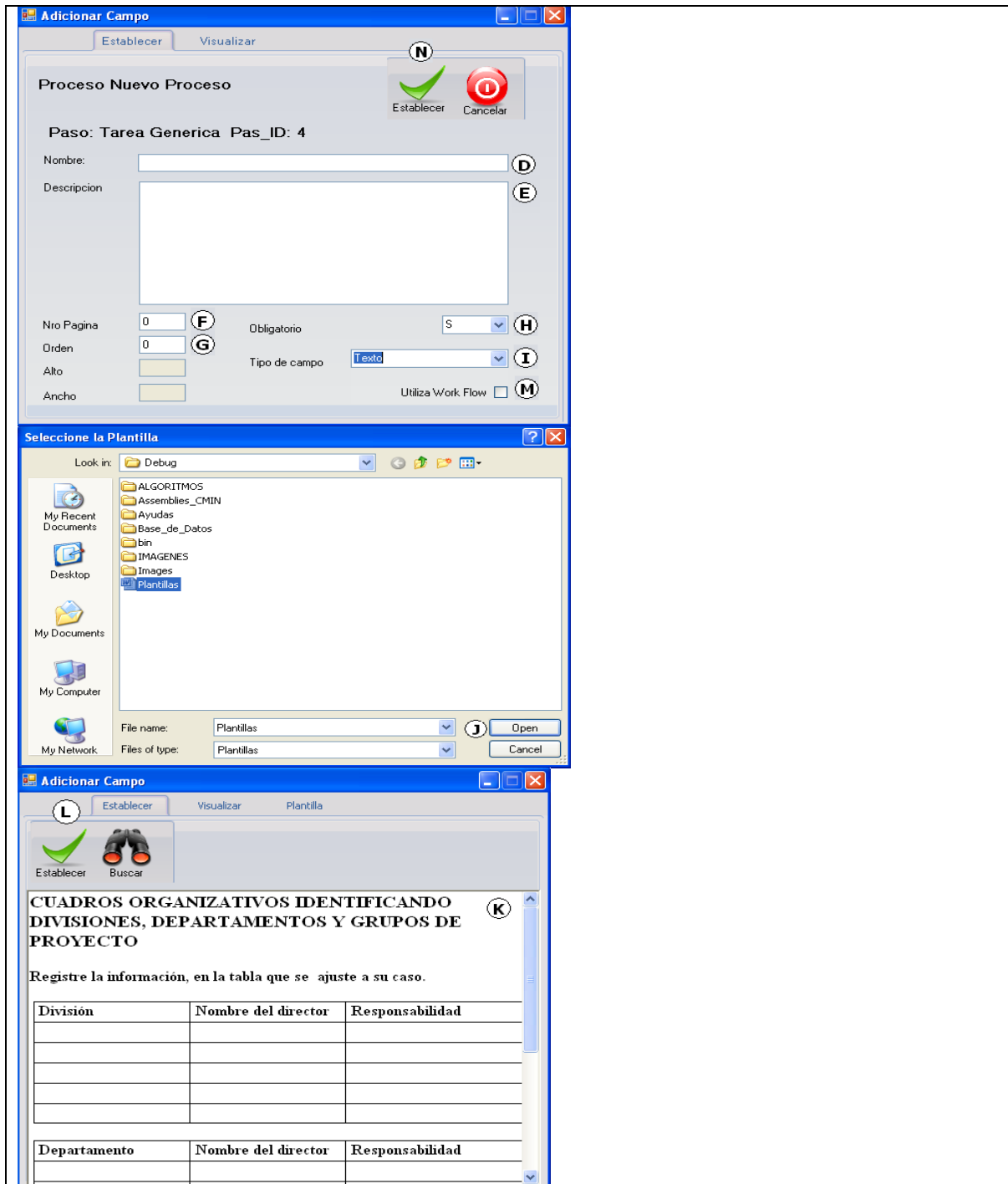
CURSO NORMAL DE LOS EVENTOS

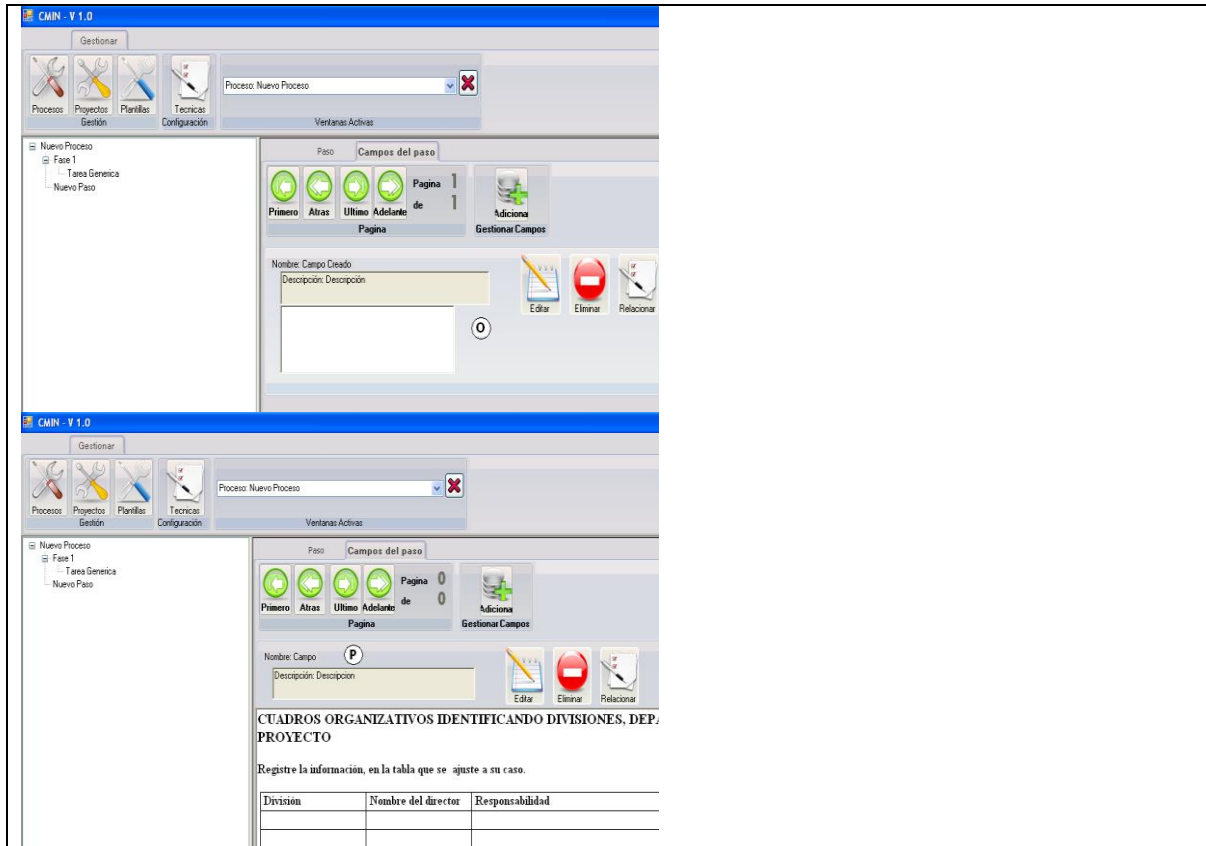
Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Pasos [A].	2. El sistema muestra la ventana de gestión de pasos.
3. El usuario da clic derecho en paso del árbol donde se ira creando la estructura del proceso [B].	4. El sistema despliega el menú para la modificación de la estructura del proceso.
5. El usuario da clic en la opción Adicionar [C].	6. El sistema limpia los controles donde ira la información general del paso [F] y [G], además en el campo del nombre [D] por defecto le pone "Nuevo Campo".
7. El usuario digita el nombre del paso nuevo [D].	

7. El usuario escoge el tipo de paso a crear [E].	
8. El usuario especifica el orden del paso [F].	
9. El usuario digita la descripción del paso [G].	
10. El usuario da clic en el botón establecer [H].	11. El sistema valida que la información este digitada y muestra la ventana de éxito.
12. El usuario da clic en el botón OK [I].	12. El sistema termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 3. Caso de Uso Real Adicionar Pasos.







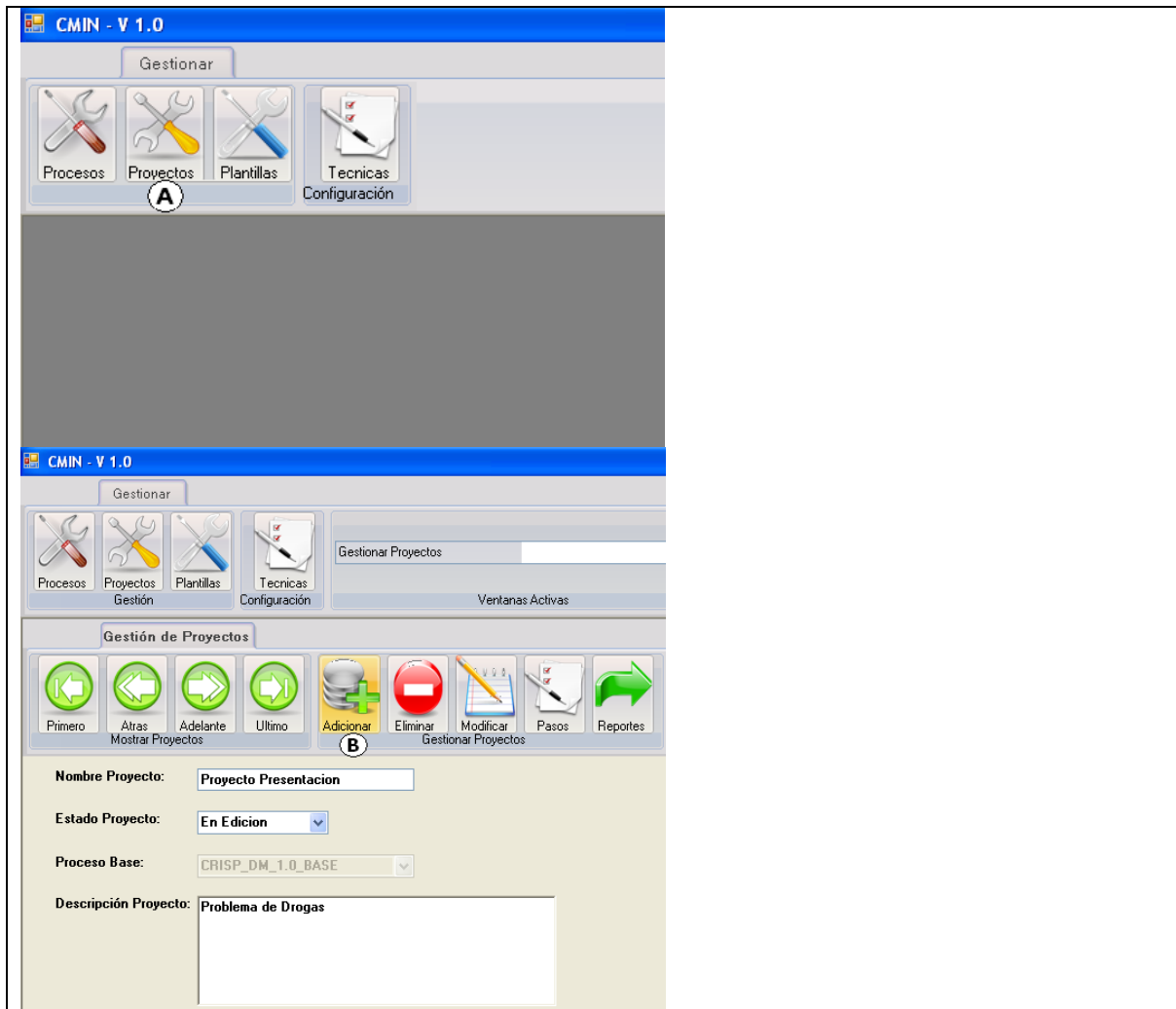
CURSO NORMAL DE LOS EVENTOS

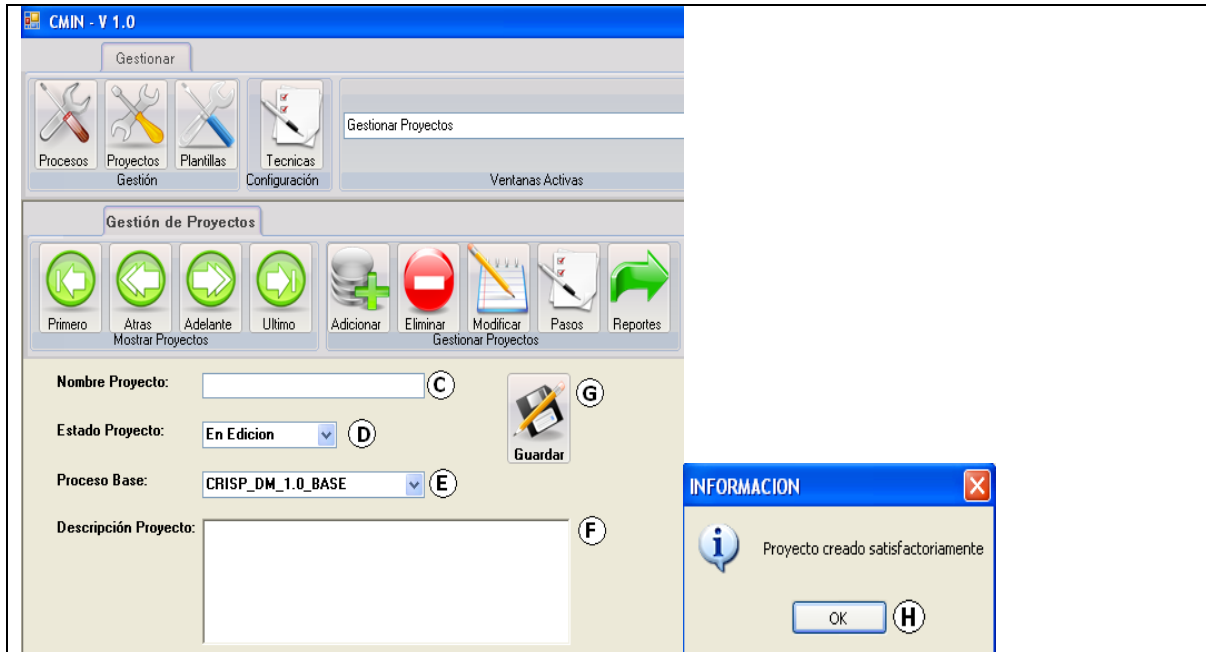
Acción del actor	Respuesta del sistema
1. El usuario selecciona un paso [A].	
2. El usuario da clic en el tab de "Campos del Paso" [B].	3. El sistema presenta los controles de navegación para ver todos los campos creados en paso y el botón de adicionar [C].
4. El usuario clic en el botón Adicionar [C].	5. El sistema presenta la ventana para crear un nuevo campo.
6. El usuario digita el nombre del campo [D].	
7. El usuario digita la descripción del campo [E].	
8. El usuario digita el numero de la pagina en donde se mostrara el campo [F].	
9. El usuario digita el orden en el que se presentara el campo en la pagina [G].	
10. El usuario escoge si el campo es obligatorio o no [H].	
11. El usuario escoge el tipo de campo a crear	

[I].	
12. El usuario especifica si el campo podrá utilizar el work flow al seleccionar la caja de chequeo [M].	
13. El usuario da clic en el botón "Establecer" [N].	14. El sistema valida que los datos estén digitados.
	15. El sistema muestra el campo creado [O] y se termina la rutina.
CURSO ALTERNO	
Acción del actor	Respuesta del sistema
11. El usuario escoge el tipo de campo plantilla [I].	12. El sistema presenta la ventana para escoger la plantilla.
13. El usuario escoge la plantilla y da clic en el botón "Open" [J].	14. El sistema presenta la previsualización de la plantilla para verificar que es la correcta [K].
14. El usuario da clic en el botón establecer para relacionar la plantilla al campo [L].	15. El sistema muestra de nuevo el tab donde se ingresa la información general del campo.
16. El usuario da clic en el botón "Establecer" [N].	17. El sistema valida que los datos estén digitados.
	18. El sistema muestra el campo creado [P] y se termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 4. Caso de Uso Real Crear Campos.

CASO DE USO REAL: CREAR PROYECTOS





CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Proyectos [A].	2. El sistema muestra la ventana de gestión de proyectos.
3. El usuario da clic en el botón Adicionar [B].	4. El sistema limpia los controles para crear un nuevo proyecto [C], [D], [E] y [F], además muestra el botón Guardar [G].
5. El usuario digita el nombre del proyecto nuevo [C].	
6. El usuario escoge si el proyecto esta en edición o terminado [D].	
7. El usuario escoge el proceso del que se basara su desarrollo [E].	
8. El usuario digita la descripción del proyecto [F].	
9. El usuario da clic en el botón "Guardar" [G].	10. El sistema valida que los datos estén digitados y muestra la ventana de éxito.
11. El usuario da clic en el botón OK [H].	12. El sistema termina la rutina.

TRAZABILIDAD

Artefactos anteriores	Artefactos posteriores
Artefactos del Análisis:	

Diagrama general de casos de uso usuario.

Tabla 5. Caso de Uso Real Crear Proyectos.

CASO DE USO REAL: DESARROLLAR PROYECTO

The screenshot shows the CMIN - V 1.0 interface with the following components:

- Top Panel:** "Gestionar" menu, "Gestión del proyecto: Nuevo Proyecto" dropdown, and "Ventanas Activas" section.
- Left Panel (CRISPDM 1.0):** A tree view showing project phases: "Conocimiento del negocio", "Determinar los objetivos del negocio", "Background (visiata)", "Organización" (highlighted with 'A'), "Área del Problema", "Solución en Curso", "Objetivos del negocio", and "Criterio de éxito desde el negocio".
- Center Panel (Campos del paso):** Navigation buttons (Primero, Atras, Adelante, Ultimo) and page indicators (Pagina 1 de 5). Below is a section for "CUADROS ORGANIZATIVOS" with a "REALIZADO" checkbox and a description: "Desarrolle cuadros organizativos identificando divisiones, departamentos y grupos de proyecto, responsabilidades de directores." A sub-section "CUADROS ORGANIZATIVOS IDENTIFICANDO DI" (with 'B') includes the instruction "Registre la información, en la tabla que se ajuste a su caso." and a table with columns: "División", "Nombre del director", "Responsabilidad".
- Bottom Panel:** A Microsoft Word window titled "Plantillas - Microsoft Word" (with 'E') showing a document with the title "CUADROS ORGANIZATIVOS IDENTIFICANDO DIVISIONES PROYECTO" (with 'D') and the same instruction and table as above. A "Microsoft Office Word" dialog box (with 'F') asks "Do you want to save the changes to Plantillas?" with "Yes", "No", and "Cancel" buttons.
- Right Panel (Actualización):** A dialog box titled "Actualización" (with 'G') asking "Desea actualizar la plantilla en el proyecto" with "Actualizar" and "Ignorar" buttons.
- Bottom Section:** A section titled "CURSO NORMAL DE LOS EVENTOS" containing a "Organización" section with the text: "Aquí se agrupan las actividades que permitirán establecer la información que se tiene de la organización en la que se realiza el proyecto." Below this is another "Campos del paso" section with navigation buttons and a "CUADROS ORGANIZATIVOS" section (with 'H') containing a table with columns: "División", "Nombre del director", "Responsabilidad", and a row for "Cambio realizado".

Acción del actor	Respuesta del sistema
1. El usuario el paso que desea modificar [A].	2. El sistema muestra la ventana que contiene los campos del paso.
3. El usuario modifica el campo si es de texto, si es plantilla la visualiza [B].	
5. El usuario da clic en el botón "Guardar" [C].	6. Si el campo es una plantilla el sistema ejecuta WORD con la plantilla que se desea modificar.
7. El usuario modifica la plantilla según lo necesitado [D].	
8. El usuario cierra WORD después de modificar [E].	9. El sistema de WORD muestra la ventana que pregunta que si se guardaran los cambios si no se ha guardado antes.
10. El usuario da clic en el botón "Si o Yes" [F].	11. El sistema muestra la ventana que pregunta si se guardaran los cambios en el sistema.
12. El usuario da clic en el botón "Actualizar" [G].	13. El sistema muestra la información de cómo quedo guardado en el sistema la plantilla [H] y termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 6. Caso de Uso Real Desarrollar Proyecto.

ANEXO B. DIAGRAMAS muestra los casos de uso u operaciones que todos los actores del sistema (expertos y novatos) pueden hacer. Los actores en este diagrama están representados por el actor Usuario. En el ambiente computacional, todos los actores pueden entrar al sistema, gestionar proyectos y gestionar procesos. Los usuarios al ingresar al sistema deben configurar el servidor de bases de datos de SQL Server, para poder cargar toda la información necesaria para el funcionamiento del sistema.

Al gestionar procesos pueden crearlos, modificarlos y eliminarlos. Cuando crean un proceso se puede adicionar, modificar y eliminar los pasos que hacen parte del proceso y los campos que pertenecen al paso, cuando modifican un proceso se puede modificar la información general del sistema así como adicionar, modificar y eliminar los pasos y los campos pertenecientes a los pasos y finalmente pueden eliminar un proceso.

Al gestionar los proyectos los usuarios pueden desarrollar los pasos propuestos por su proceso base, de tal manera que ejecutan los campos que se definen para cada paso, en algunos campos se puede utilizar el work flow si se necesita realizar practicas propias de minería de datos.

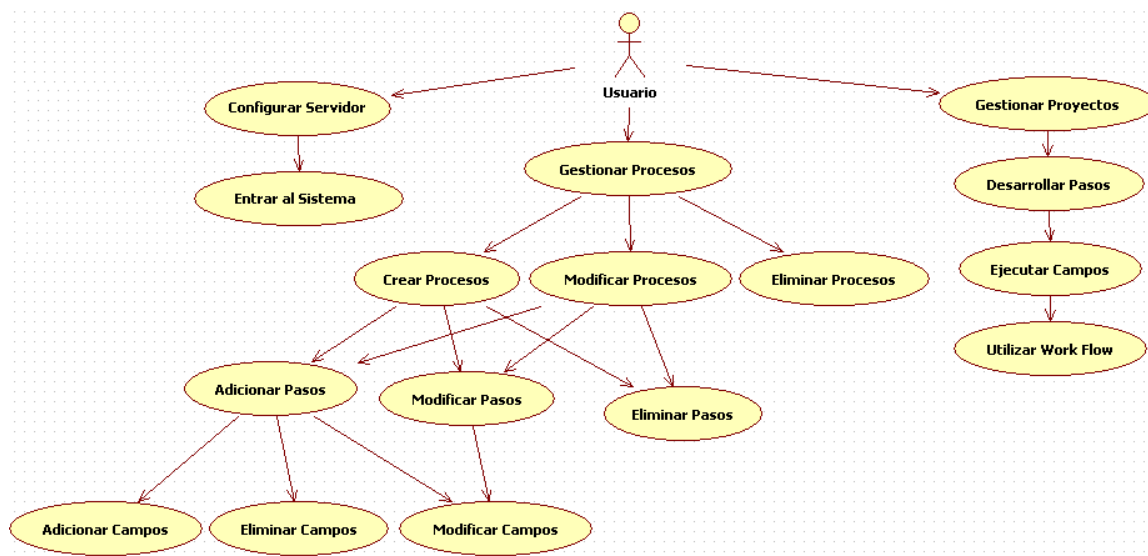


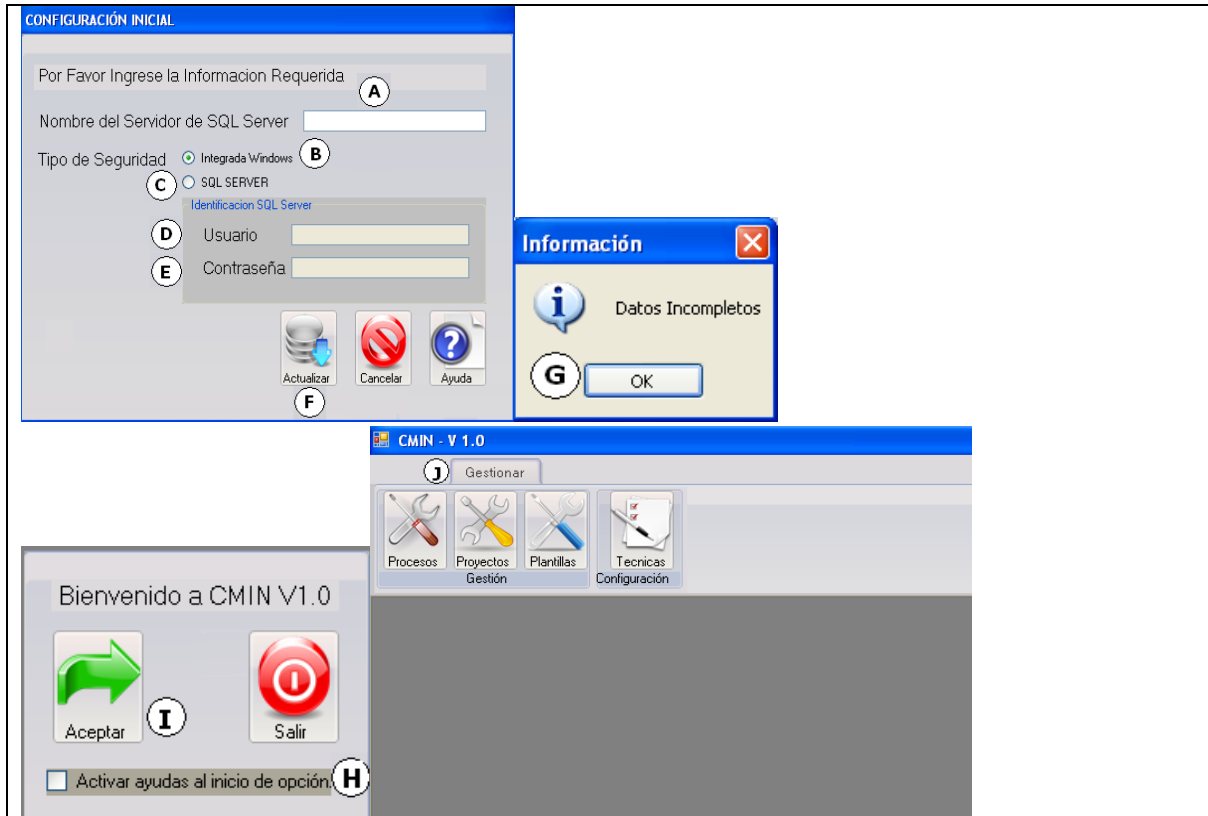
Figura 1. Caso de Uso General del Sistema.

3.1 Casos de Uso de Usuarios Expertos

La Figura 2. Casos de Uso de Usuarios Expertos.

4 CASOS DE USO REALES DEL USUARIO

CASO DE USO REAL: ENTRAR AL SISTEMA



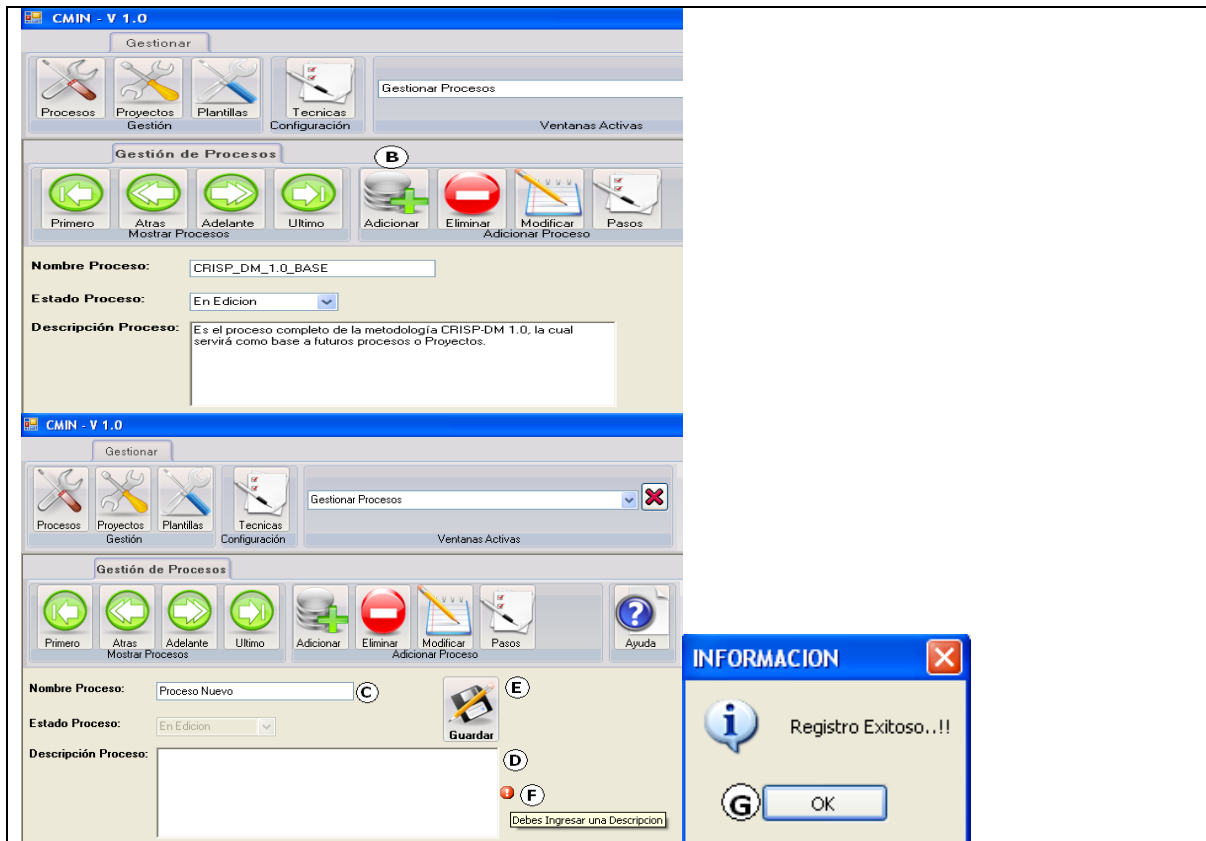
CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
29. El usuario digita el nombre del servidor de SQL Server [A].	
30. El usuario selecciona la opción de seguridad Integrada de Windows [B].	31. El sistema deshabilita los campos para usuario [D] y contraseña [E] de seguridad SQL Server.
32. El usuario selecciona la opción de seguridad SQL Server [C].	33. El sistema habilita los campos para usuario [D] y contraseña [E] de seguridad SQL Server.
34. El usuario digita el usuario [D].	
35. El usuario digita la contraseña [E].	
36. El usuario da clic en el botón actualizar [F].	37. El sistema valida que con la información suministrada se pueda ingresar al sistema.
	38. El sistema despliega la ventana de bienvenida de CMIN 1.0.
39. El usuario debe seleccionar o no la opción	40. Si es seleccionado el sistema

"Activar ayudas al inicio de opción." [H].	mostrara automáticamente las ayudas del sistema cuando se ejecute una opción de lo contrario no.
41. El usuario da clic en el botón Aceptar [I].	42. El sistema muestra la ventana principal del sistema con una pestaña con las acciones disponibles para todos los usuarios [J].
CURSO ALTERNO	
Acción del actor	Respuesta del sistema
	9. El usuario no digito la información obligatoria en los campos [D]y [E]. Se le informa al usuario que los datos están incompletos [G].
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 1. Caso de uso real Entrar al sistema.





CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Procesos [A].	2. El sistema muestra la ventana de gestión de procesos.
3. El usuario da clic en el botón Adicionar [B].	4. El sistema limpia los controles para crear un nuevo proceso [C] y [D], además muestra el botón Guardar [E].
5. El usuario digita el nombre del proceso nuevo [C].	
6. El usuario digita la descripción del proceso nuevo [D].	
7. El usuario da clic en el botón Guardar [E].	8. El sistema valida que los campos estén digitados, después guarda la información general del nuevo proceso en la base de datos y muestra la ventana de éxito.
9. El usuario da clic en el botón OK [G].	10. El sistema termina la rutina.

CURSO ALTERNO	
Acción del actor	Respuesta del sistema
	8. El usuario no digito la información obligatoria en los campos [C]y [D]. Se le informa al usuario que los datos están incompletos [F].
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 2. Caso de so Real Crear Procesos

CASO DE USO REAL: ADICIONAR PASOS

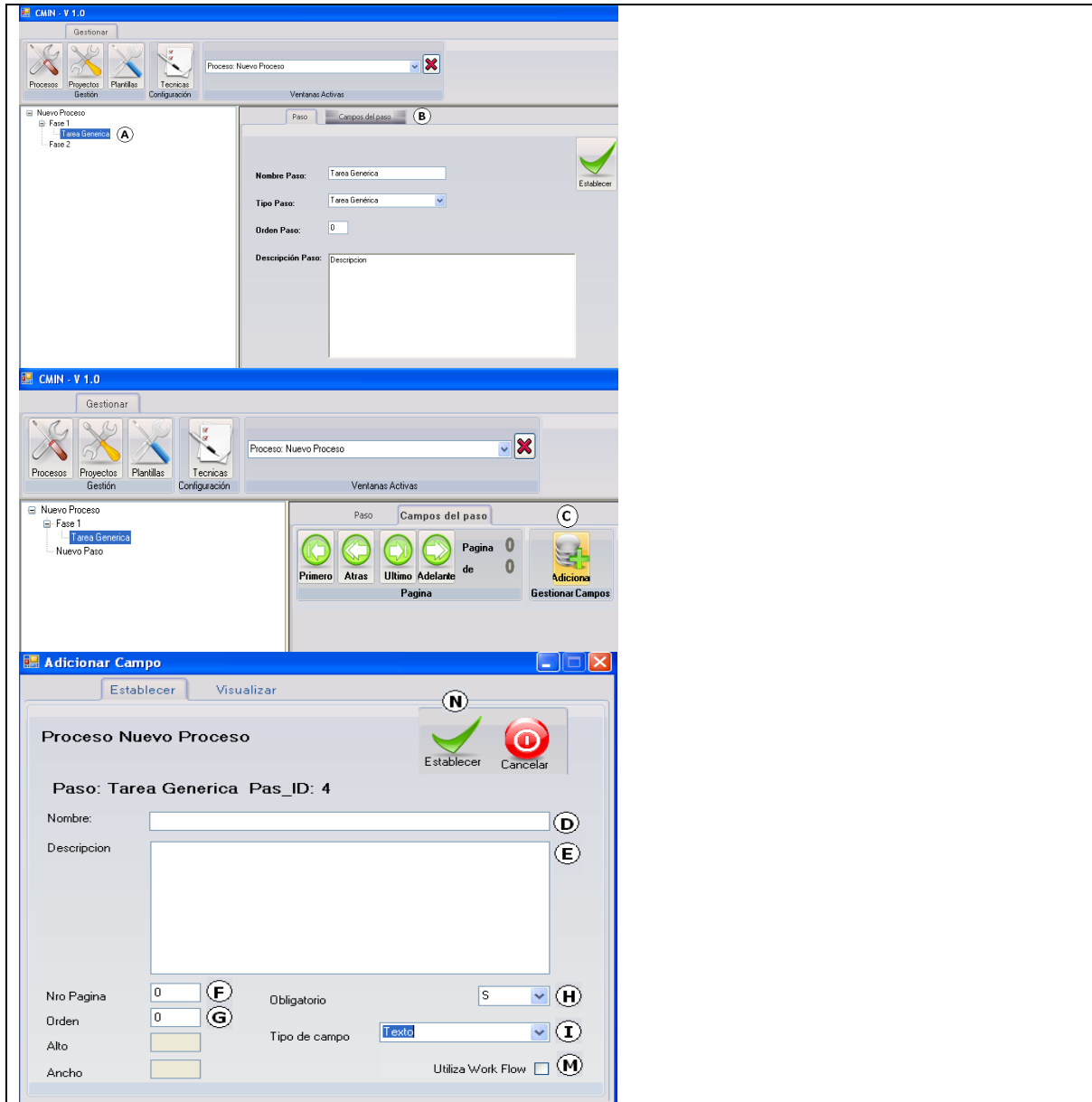
The image displays two screenshots of the CMIN 1.0 software interface. The top screenshot shows the 'Gestión de Procesos' window. The 'Nombre Proceso' field contains 'CRISP_DM_1.0_BASE', and the 'Estado Proceso' is set to 'En Edición'. The 'Descripción Proceso' field contains the text: 'Es el proceso completo de la metodología CRISP-DM 1.0, la cual servirá como base a futuros procesos o Proyectos.' The bottom screenshot shows the 'Nuevo Proceso' window. The 'Nombre Paso' field contains 'Nuevo Proceso', the 'Tipo Paso' is 'Tarea Genérica', and the 'Orden Paso' is '1'. An 'INFORMACION' dialog box is open, displaying 'Registro Exitoso...!!' and an 'OK' button.

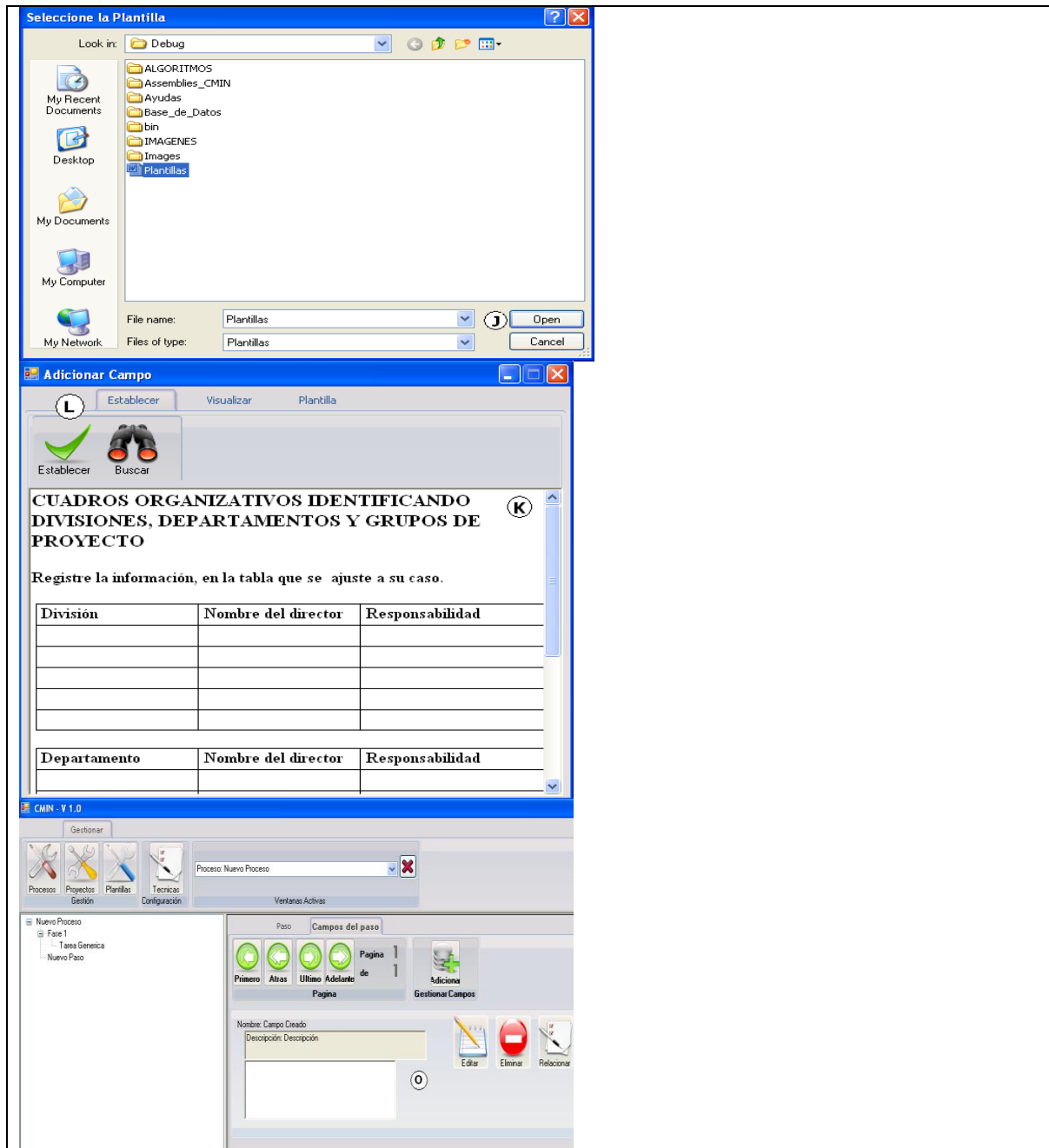
CURSO NORMAL DE LOS EVENTOS

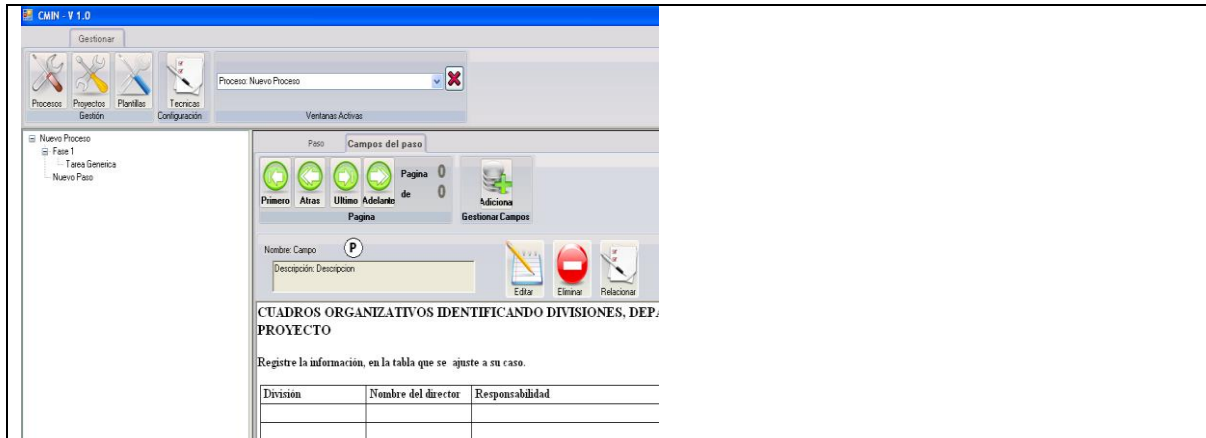
Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Pasos [A].	2. El sistema muestra la ventana de gestión de pasos.
3. El usuario da clic derecho en paso del árbol donde se ira creando la estructura del proceso [B].	4. El sistema despliega el menú para la modificación de la estructura del proceso.
5. El usuario da clic en la opción Adicionar [C].	6. El sistema limpia los controles donde ira la información general del paso [F] y [G], además en el campo del nombre [D] por defecto le pone "Nuevo Campo".
7. El usuario digita el nombre del paso nuevo [D].	
7. El usuario escoge el tipo de paso a crear [E].	
8. El usuario especifica el orden del paso [F].	
9. El usuario digita la descripción del paso [G].	
10. El usuario da clic en el botón establecer [H].	11. El sistema valida que la información este digitada y muestra la ventana de éxito.
12. El usuario da clic en el botón OK [I].	12. El sistema termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 3. Caso de Uso Real Adicionar Pasos.

CASO DE USO REAL: CREAR CAMPOS







CURSO NORMAL DE LOS EVENTOS

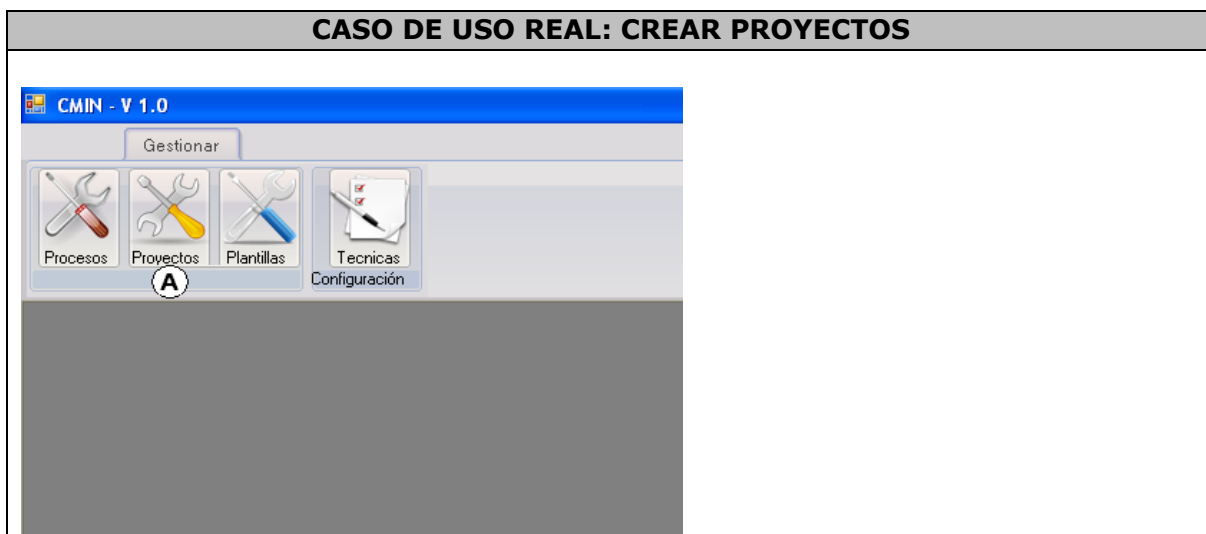
Acción del actor	Respuesta del sistema
1. El usuario selecciona un paso [A].	
2. El usuario da clic en el tab de "Campos del Paso" [B].	3. El sistema presenta los controles de navegación para ver todos los campos creados en paso y el botón de adicionar [C].
4. El usuario clic en el botón Adicionar [C].	5. El sistema presenta la ventana para crear un nuevo campo.
6. El usuario digita el nombre del campo [D].	
7. El usuario digita la descripción del campo [E].	
8. El usuario digita el numero de la pagina en donde se mostrara el campo [F].	
9. El usuario digita el orden en el que se presentara el campo en la pagina [G].	
10. El usuario escoge si el campo es obligatorio o no [H].	
11. El usuario escoge el tipo de campo a crear [I].	
12. El usuario especifica si el campo podrá utilizar el work flow al seleccionar la caja de chequeo [M].	
13. El usuario da clic en el botón "Establecer" [N].	14. El sistema valida que los datos estén digitados.
	15. El sistema muestra el campo creado [O] y se termina la rutina.

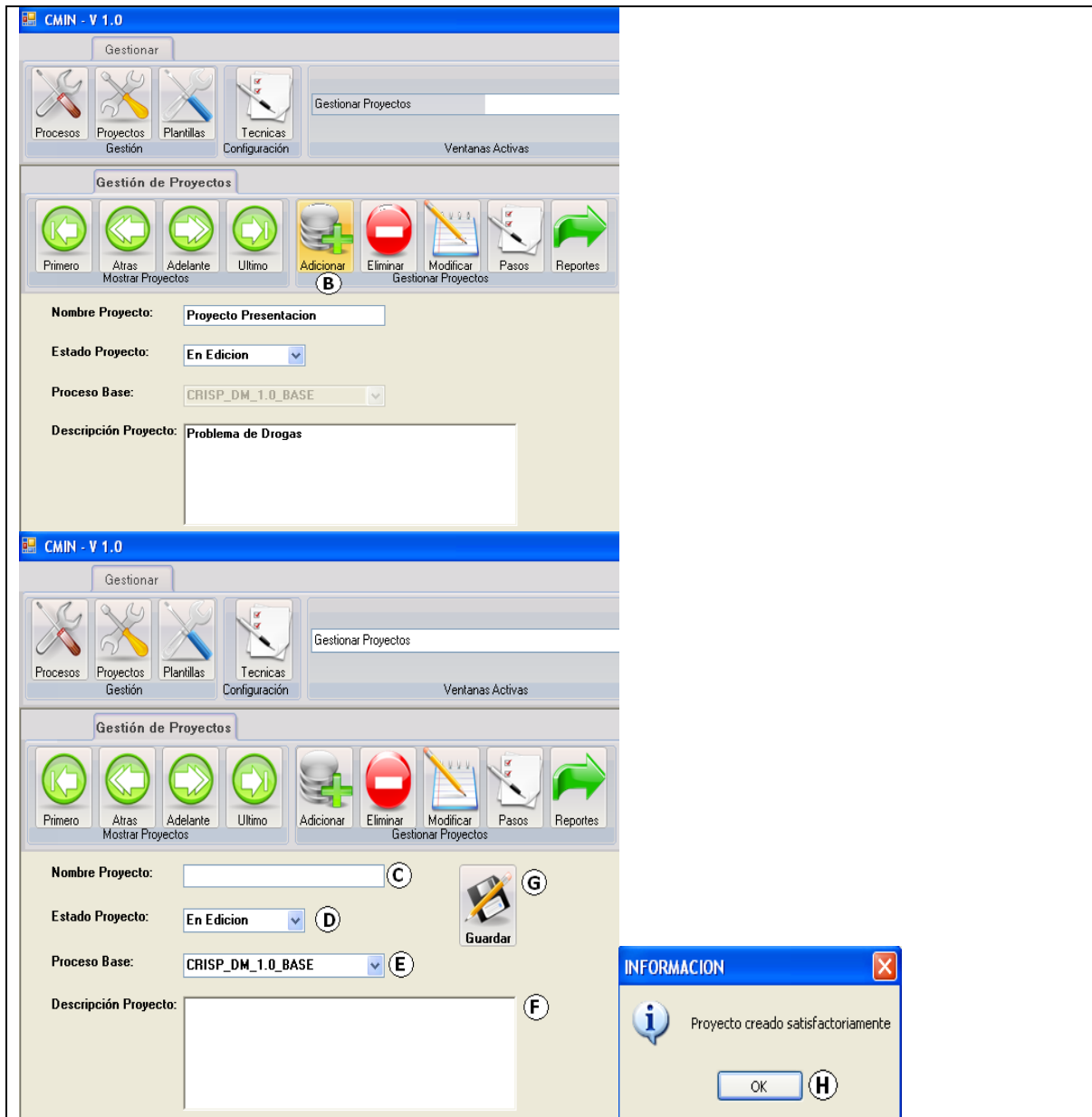
CURSO ALTERNO

Acción del actor	Respuesta del sistema
11. El usuario escoge el tipo de campo plantilla	12. El sistema presenta la ventana

[I].	para escoger la plantilla.
13. El usuario escoge la plantilla y da clic en el botón "Open" [J].	14. El sistema presenta la previsualización de la plantilla para verificar que es la correcta [K].
14. El usuario da clic en el botón establecer para relacionar la plantilla al campo [L].	15. El sistema muestra de nuevo el tab donde se ingresa la información general del campo.
16. El usuario da clic en el botón "Establecer" [N].	17. El sistema valida que los datos estén digitados.
	18. El sistema muestra el campo creado [P] y se termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 4. Caso de Uso Real Crear Campos.





CURSO NORMAL DE LOS EVENTOS

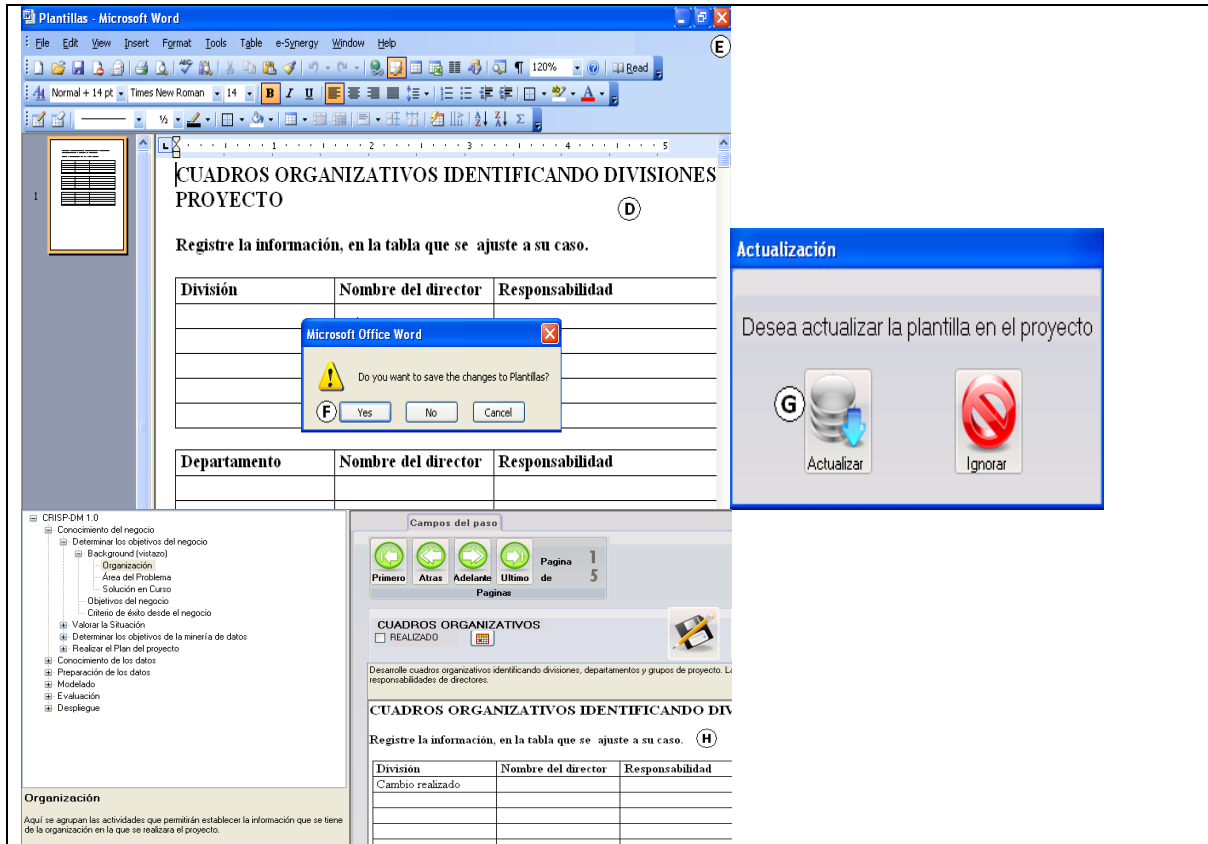
Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Proyectos [A].	2. El sistema muestra la ventana de gestión de proyectos.
3. El usuario da clic en el botón Adicionar [B].	4. El sistema limpia los controles para crear un nuevo proyecto [C], [D], [E] y [F], además muestra el botón Guardar [G].

5. El usuario digita el nombre del proyecto nuevo [C].	
6. El usuario escoge si el proyecto esta en edición o terminado [D].	
7. El usuario escoge el proceso del que se basara su desarrollo [E].	
8. El usuario digita la descripción del proyecto [F].	
9. El usuario da clic en el botón "Guardar" [G].	10. El sistema valida que los datos estén digitados y muestra la ventana de éxito.
11. El usuario da clic en el botón OK [H].	12. El sistema termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 5. Caso de Uso Real Crear Proyectos.

CASO DE USO REAL: DESARROLLAR PROYECTO

The screenshot displays the 'Organización' step of a project management process. The left sidebar shows a hierarchical tree of project steps, with 'Organización' selected and marked with a circled 'A'. The main workspace features a 'Campos del paso' section with navigation buttons and a 'Paginas' indicator showing page 1 of 5. Below this, the 'CUADROS ORGANIZATIVOS' section is active, with a 'REALIZADO' checkbox and a table for recording organizational information. The table has columns for 'Division', 'Nombre del director', and 'Responsabilidad'. A circled 'B' is placed near the table, and a circled 'C' is near the 'Paginas' indicator.



CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario el paso que desea modificar [A].	2. El sistema muestra la ventana que contiene los campos del paso.
3. El usuario modifica el campo si es de texto, si es plantilla la visualiza [B].	
5. El usuario da clic en el botón "Guardar" [C].	6. Si el campo es una plantilla el sistema ejecuta WORD con la plantilla que se desea modificar.
7. El usuario modifica la plantilla según lo necesitado [D].	
8. El usuario cierra WORD después de modificar [E].	9. El sistema de WORD muestra la ventana que pregunta que si se guardaran los cambios si no se ha guardado antes.
10. El usuario da clic en el botón "Si o Yes" [F].	11. El sistema muestra la ventana que pregunta si se guardaran los cambios en el sistema.
12. El usuario da clic en el botón "Actualizar"	13. El sistema muestra la

[G].	información de cómo quedo guardado en el sistema la plantilla [H] y termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 6. Caso de Uso Real Desarrollar Proyecto.

ANEXO B. DIAGRAMAS muestra los casos de uso u operaciones que los actores expertos del sistema pueden hacer. En el ambiente computacional, los actores expertos pueden agregar dinámicamente DLL'S y gestionar las plantillas. Al agregar dinámicamente las DLL's los expertos gestionan la tareas de minería creándolas, modificándolas y eliminándolas. Al crear o modificar tareas de minería se agregan las DLL'S que implementan un objeto del work flow. Al gestionar plantillas se crean, modifican y eliminan. Los expertos al crear o modificar una plantilla basados en la experiencia especializan un proceso en un tema de proyectos específico, eliminando los pasos que no cumplen con dicho tema.

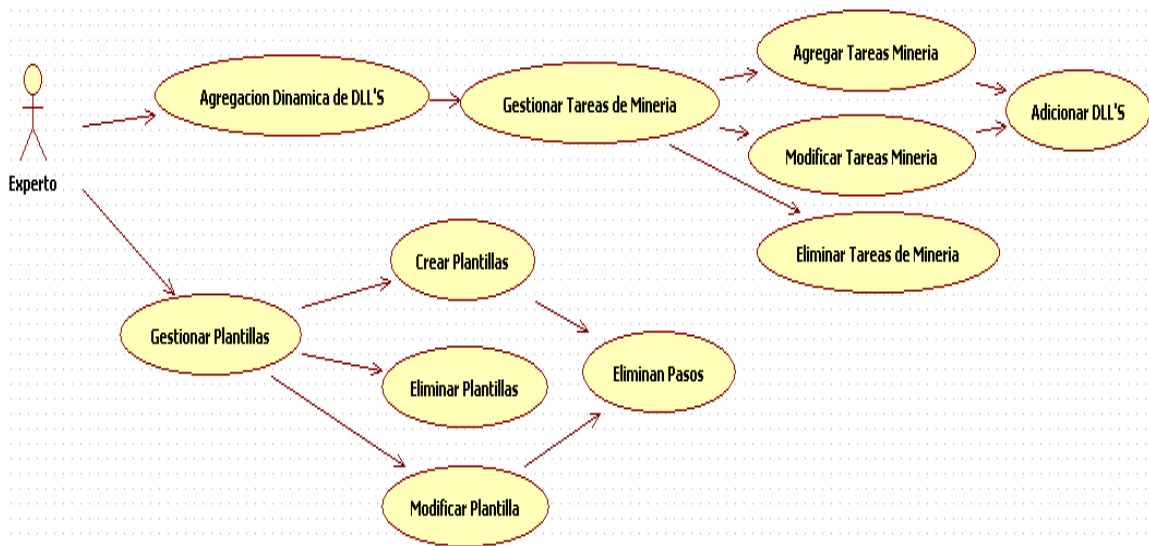
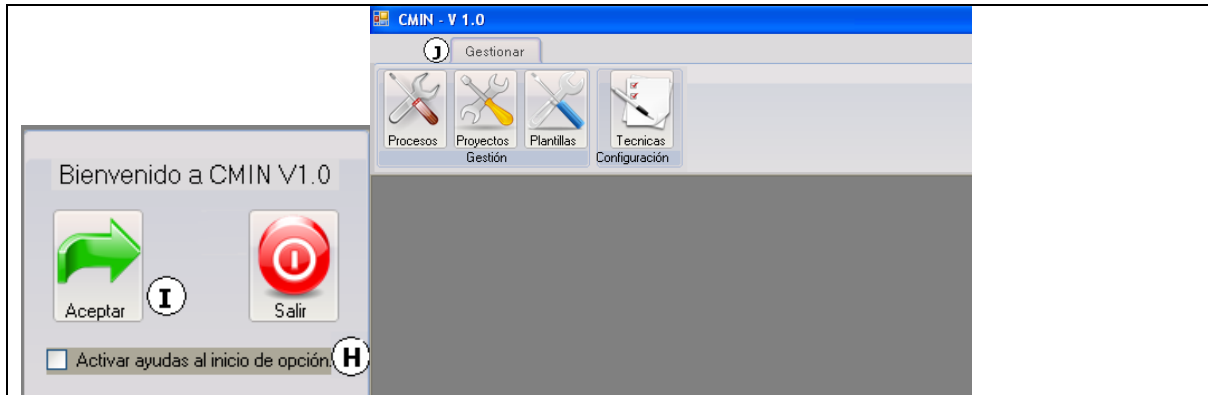


Figura 2. Casos de Uso de Usuarios Expertos.

5 CASOS DE USO REALES DEL USUARIO

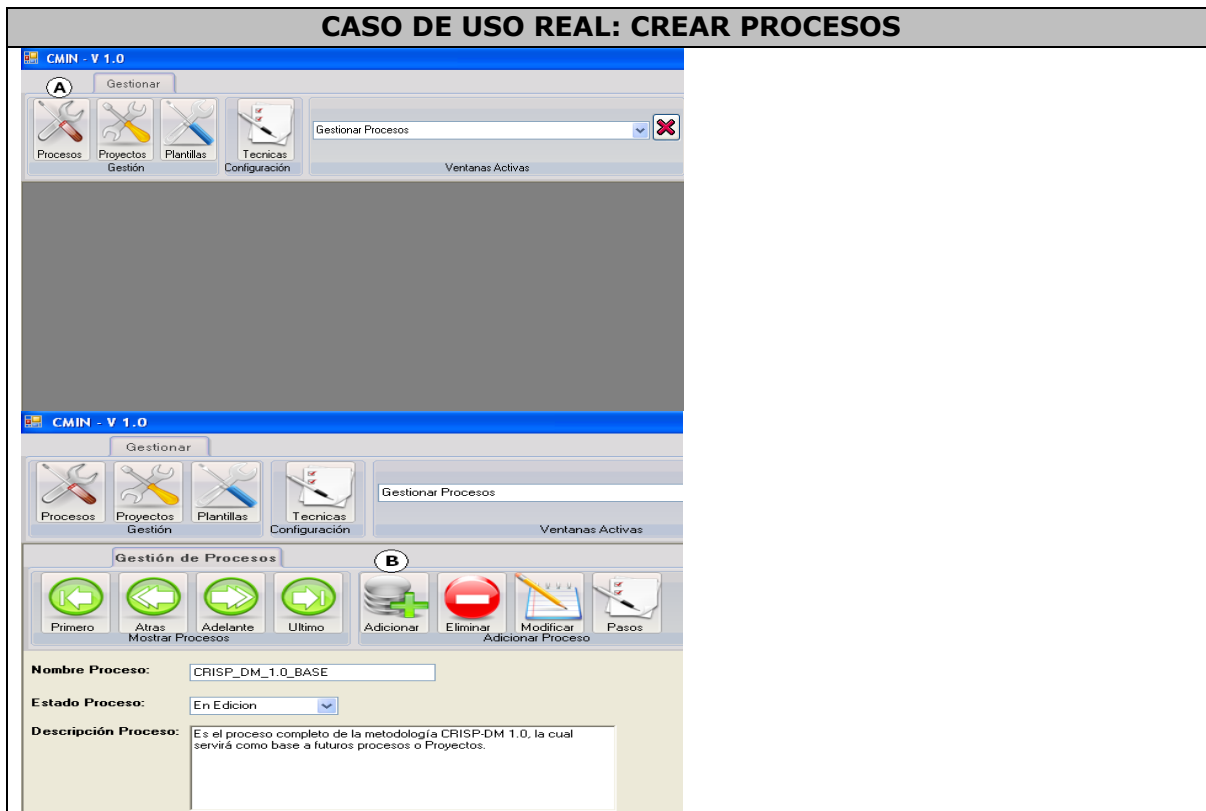


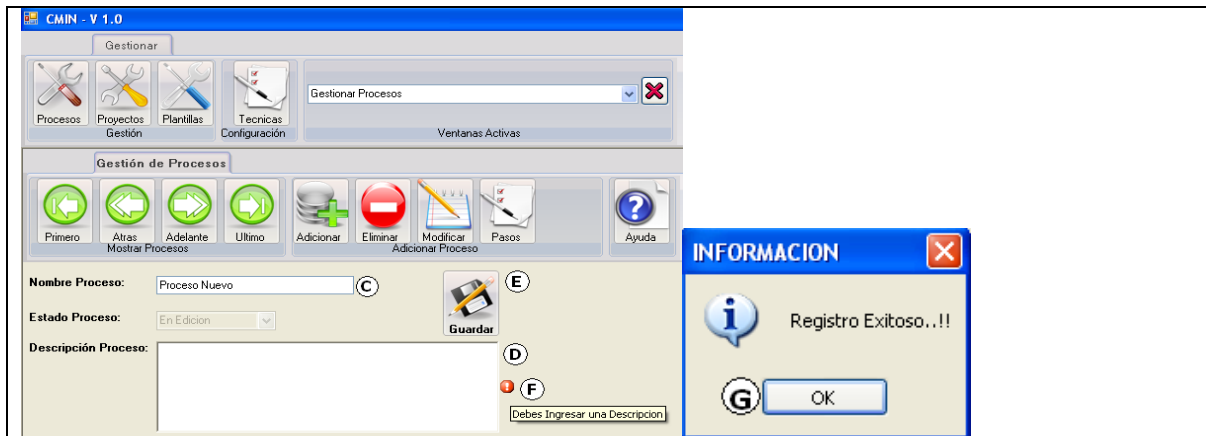


CURSO NORMAL DE LOS EVENTOS	
Acción del actor	Respuesta del sistema
43. El usuario digita el nombre del servidor de SQL Server [A].	
44. El usuario selecciona la opción de seguridad Integrada de Windows [B].	45. El sistema deshabilita los campos para usuario [D] y contraseña [E] de seguridad SQL Server.
46. El usuario selecciona la opción de seguridad SQL Server [C].	47. El sistema habilita los campos para usuario [D] y contraseña [E] de seguridad SQL Server.
48. El usuario digita el usuario [D].	
49. El usuario digita la contraseña [E].	
50. El usuario da clic en el botón actualizar [F].	51. El sistema valida que con la información suministrada se pueda ingresar al sistema.
	52. El sistema despliega la ventana de bienvenida de CMIN 1.0.
53. El usuario debe seleccionar o no la opción "Activar ayudas al inicio de opción." [H].	54. Si es seleccionado el sistema mostrara automáticamente las ayudas del sistema cuando se ejecute una opción de lo contrario no.
55. El usuario da clic en el botón Aceptar [I].	56. El sistema muestra la ventana principal del sistema con una pestaña con las acciones disponibles para todos los usuarios [J].
CURSO ALTERNO	
Acción del actor	Respuesta del sistema

	9. El usuario no digito la información obligatoria en los campos [D]y [E]. Se le informa al usuario que los datos están incompletos [G].
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 1. Caso de uso real Entrar al sistema.





CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Procesos [A].	2. El sistema muestra la ventana de gestión de procesos.
3. El usuario da clic en el botón Adicionar [B].	4. El sistema limpia los controles para crear un nuevo proceso [C] y [D], además muestra el botón Guardar [E].
5. El usuario digita el nombre del proceso nuevo [C].	
6. El usuario digita la descripción del proceso nuevo [D].	
7. El usuario da clic en el botón Guardar [E].	8. El sistema valida que los campos estén digitados, después guarda la información general del nuevo proceso en la base de datos y muestra la ventana de éxito.
9. El usuario da clic en el botón OK [G].	10. El sistema termina la rutina.

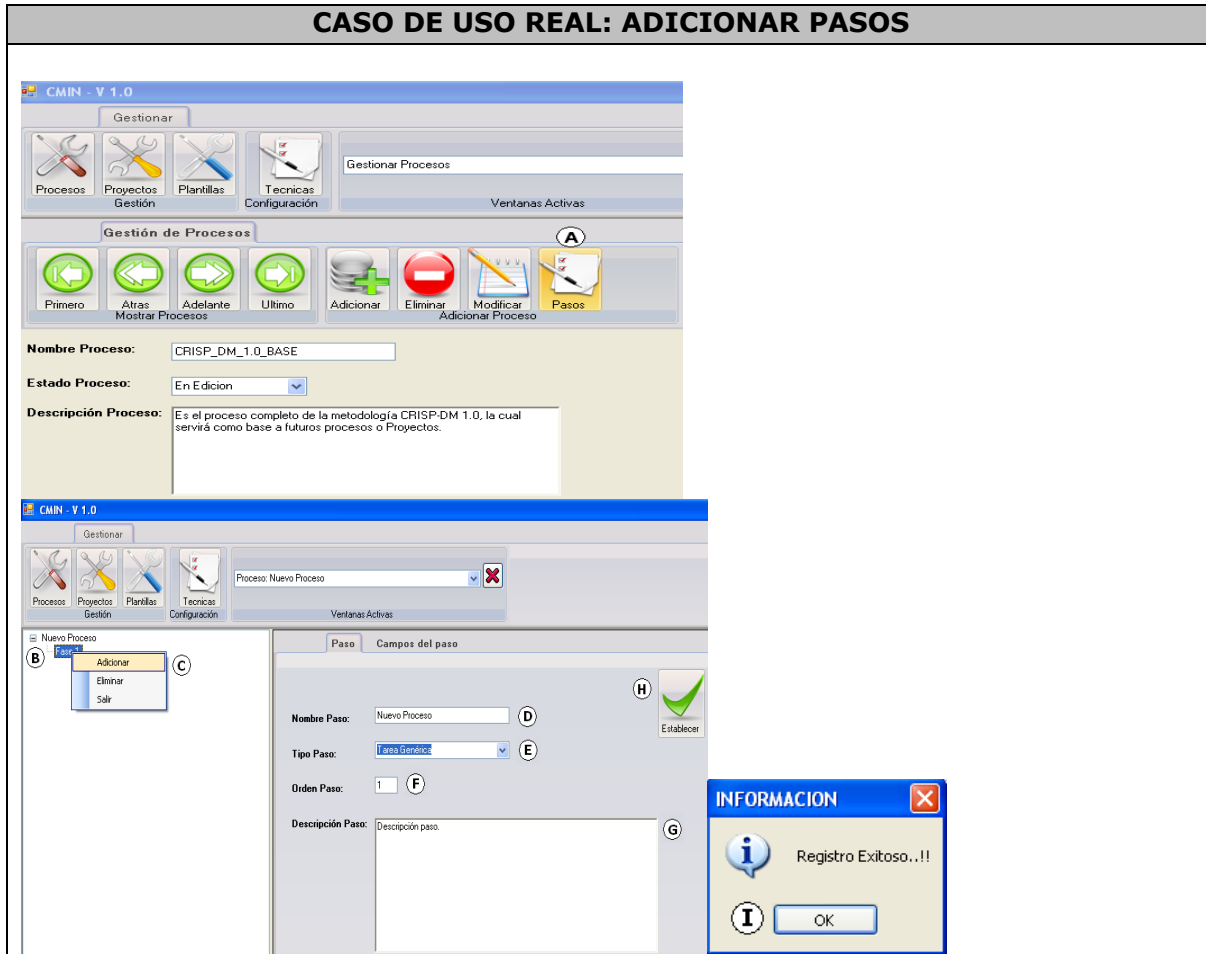
CURSO ALTERNO

Acción del actor	Respuesta del sistema
	8. El usuario no digito la información obligatoria en los campos [C]y [D]. Se le informa al usuario que los datos están incompletos [F].

TRAZABILIDAD

Artefactos anteriores	Artefactos posteriores
Artefactos del Análisis: Diagrama general de casos de uso usuario.	

Tabla 2. Caso de uso Real Crear Procesos



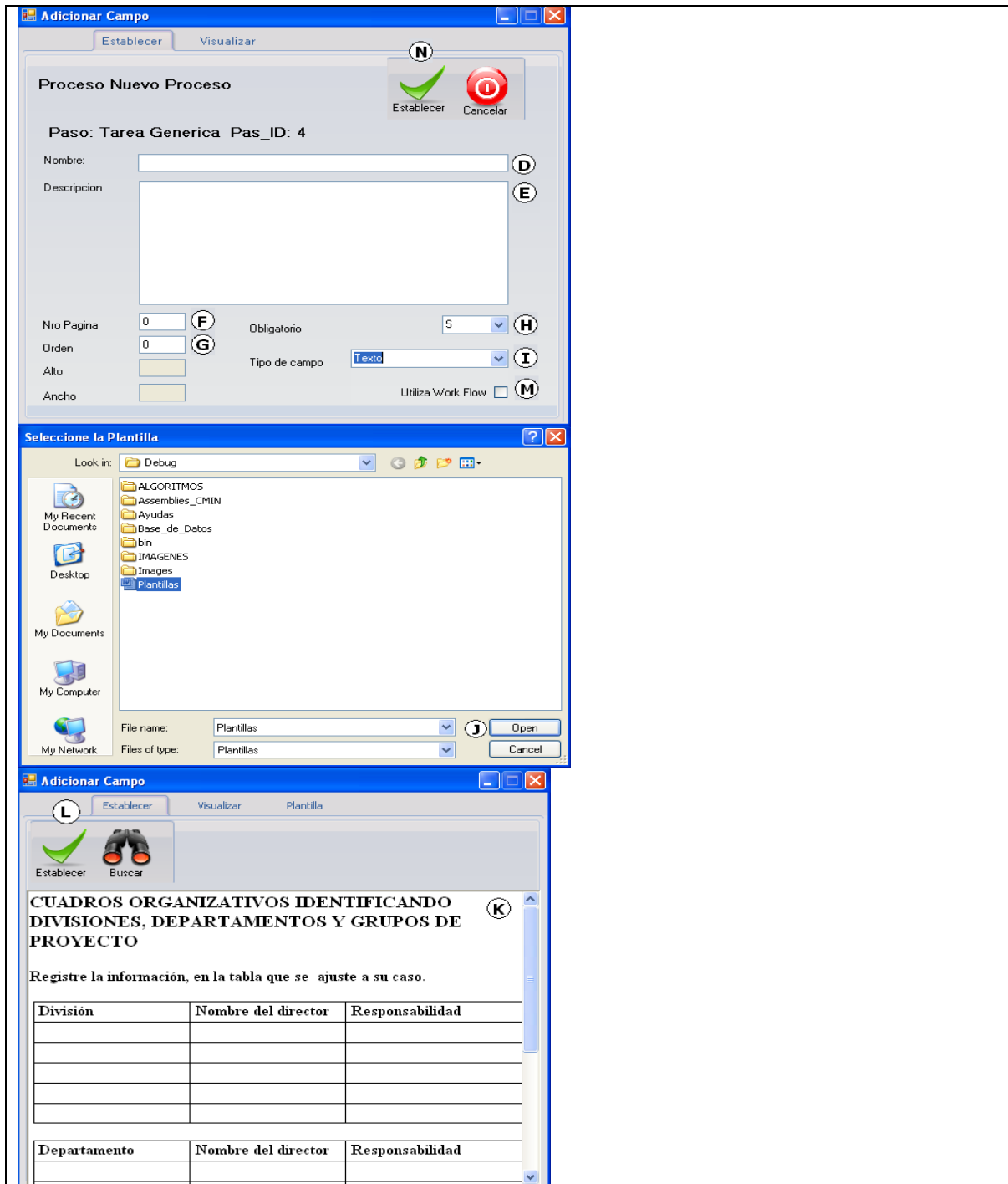
CURSO NORMAL DE LOS EVENTOS

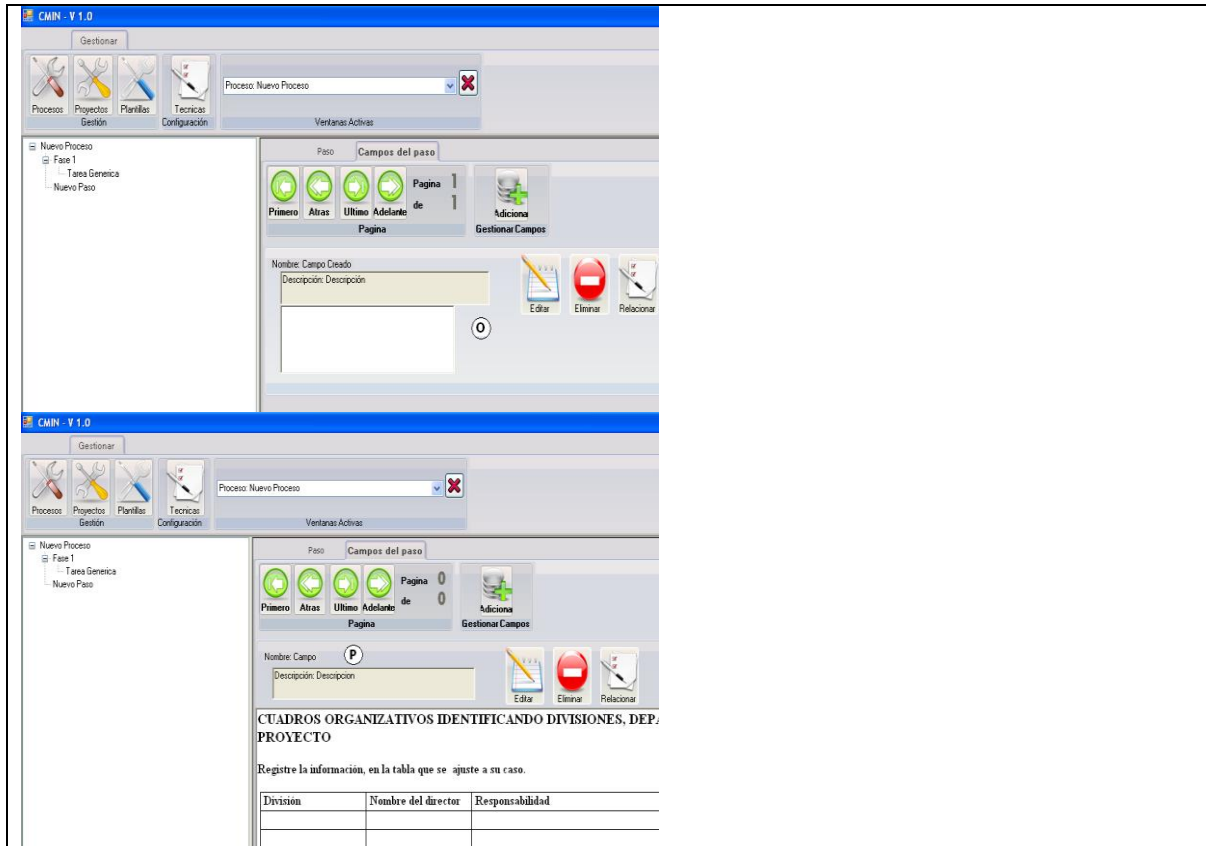
Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Pasos [A].	2. El sistema muestra la ventana de gestión de pasos.
3. El usuario da clic derecho en paso del árbol donde se ira creando la estructura del proceso [B].	4. El sistema despliega el menú para la modificación de la estructura del proceso.
5. El usuario da clic en la opción Adicionar [C].	6. El sistema limpia los controles donde ira la información general del paso [F] y [G], además en el campo del nombre [D] por defecto le pone "Nuevo Campo".
7. El usuario digita el nombre del paso nuevo [D].	

7. El usuario escoge el tipo de paso a crear [E].	
8. El usuario especifica el orden del paso [F].	
9. El usuario digita la descripción del paso [G].	
10. El usuario da clic en el botón establecer [H].	11. El sistema valida que la información este digitada y muestra la ventana de éxito.
12. El usuario da clic en el botón OK [I].	12. El sistema termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 3. Caso de Uso Real Adicionar Pasos.







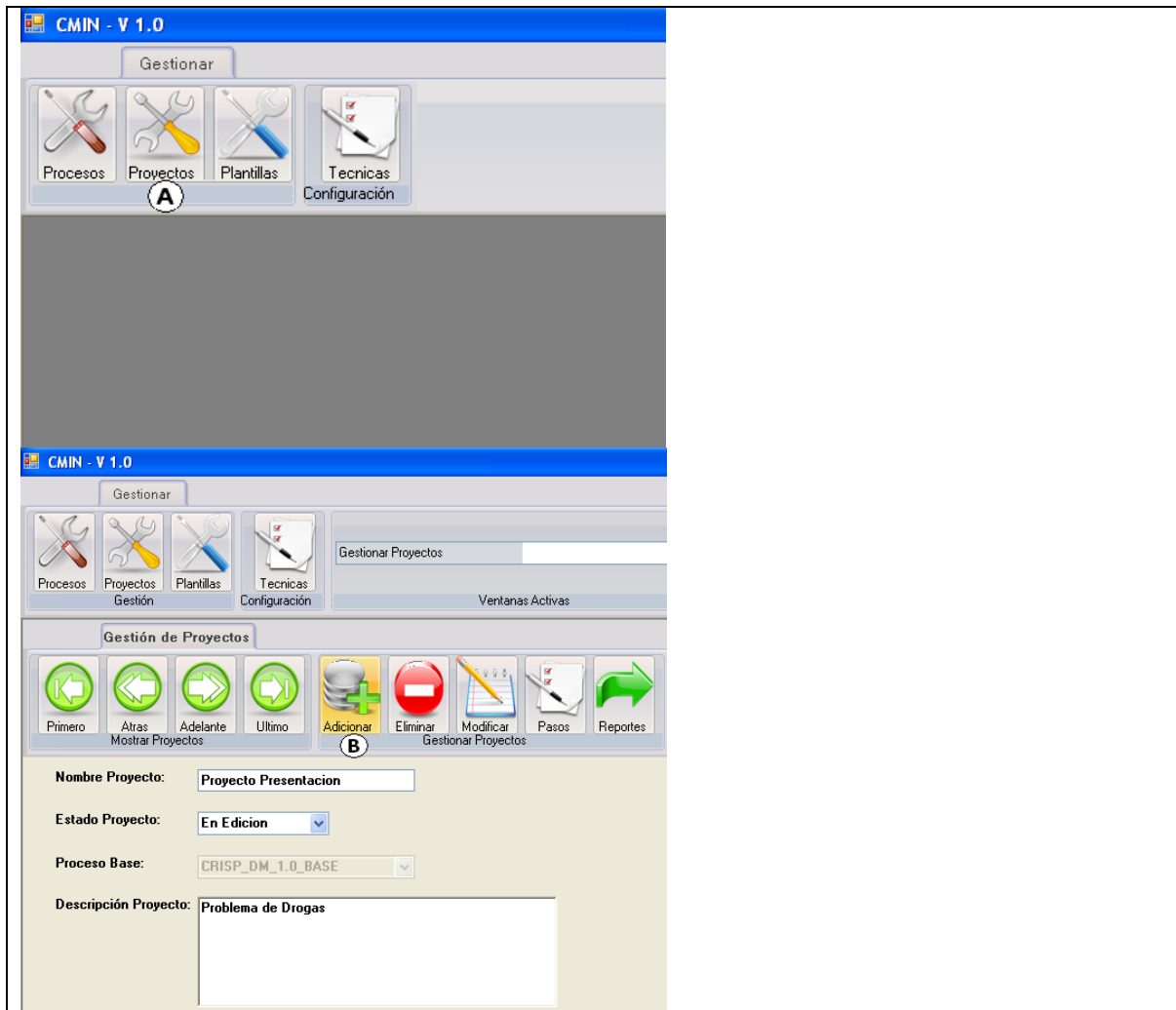
CURSO NORMAL DE LOS EVENTOS

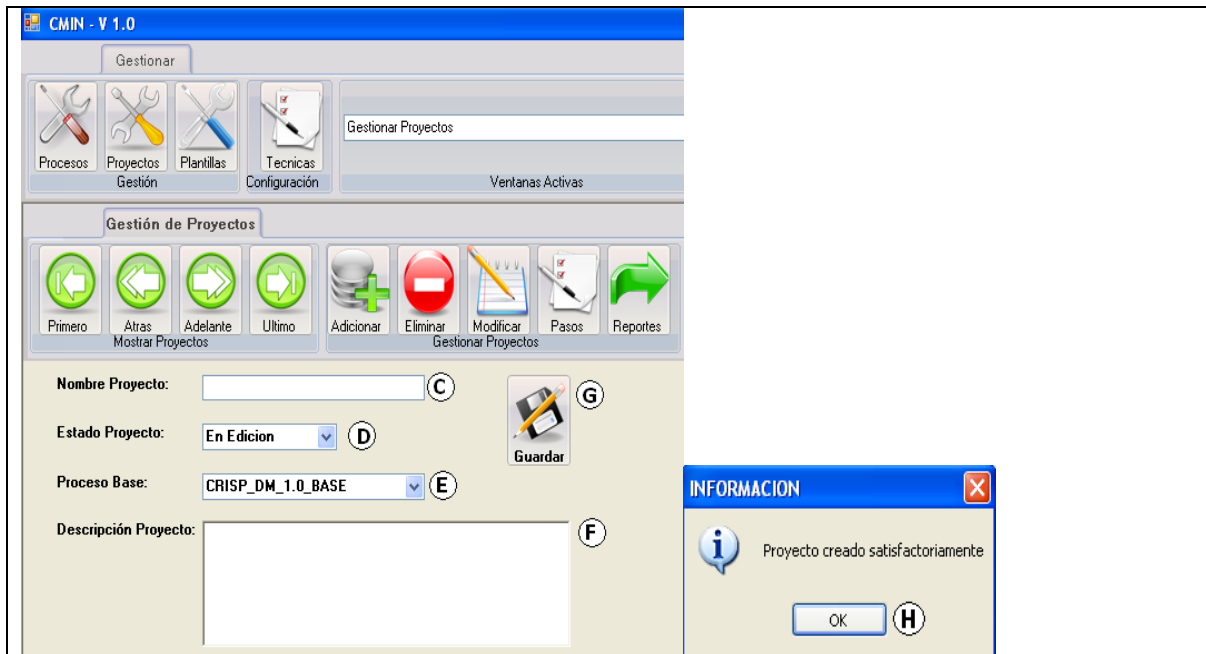
Acción del actor	Respuesta del sistema
1. El usuario selecciona un paso [A].	
2. El usuario da clic en el tab de "Campos del Paso" [B].	3. El sistema presenta los controles de navegación para ver todos los campos creados en paso y el botón de adicionar [C].
4. El usuario clic en el botón Adicionar [C].	5. El sistema presenta la ventana para crear un nuevo campo.
6. El usuario digita el nombre del campo [D].	
7. El usuario digita la descripción del campo [E].	
8. El usuario digita el numero de la pagina en donde se mostrara el campo [F].	
9. El usuario digita el orden en el que se presentara el campo en la pagina [G].	
10. El usuario escoge si el campo es obligatorio o no [H].	
11. El usuario escoge el tipo de campo a crear	

[I].	
12. El usuario especifica si el campo podrá utilizar el work flow al seleccionar la caja de chequeo [M].	
13. El usuario da clic en el botón "Establecer" [N].	14. El sistema valida que los datos estén digitados.
	15. El sistema muestra el campo creado [O] y se termina la rutina.
CURSO ALTERNO	
Acción del actor	Respuesta del sistema
11. El usuario escoge el tipo de campo plantilla [I].	12. El sistema presenta la ventana para escoger la plantilla.
13. El usuario escoge la plantilla y da clic en el botón "Open" [J].	14. El sistema presenta la previsualización de la plantilla para verificar que es la correcta [K].
14. El usuario da clic en el botón establecer para relacionar la plantilla al campo [L].	15. El sistema muestra de nuevo el tab donde se ingresa la información general del campo.
16. El usuario da clic en el botón "Establecer" [N].	17. El sistema valida que los datos estén digitados.
	18. El sistema muestra el campo creado [P] y se termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 4. Caso de Uso Real Crear Campos.

CASO DE USO REAL: CREAR PROYECTOS





CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario da clic en el botón Proyectos [A].	2. El sistema muestra la ventana de gestión de proyectos.
3. El usuario da clic en el botón Adicionar [B].	4. El sistema limpia los controles para crear un nuevo proyecto [C], [D], [E] y [F], además muestra el botón Guardar [G].
5. El usuario digita el nombre del proyecto nuevo [C].	
6. El usuario escoge si el proyecto esta en edición o terminado [D].	
7. El usuario escoge el proceso del que se basara su desarrollo [E].	
8. El usuario digita la descripción del proyecto [F].	
9. El usuario da clic en el botón "Guardar" [G].	10. El sistema valida que los datos estén digitados y muestra la ventana de éxito.
11. El usuario da clic en el botón OK [H].	12. El sistema termina la rutina.

TRAZABILIDAD

Artefactos anteriores	Artefactos posteriores
Artefactos del Análisis:	

Diagrama general de casos de uso usuario.

Tabla 5. Caso de Uso Real Crear Proyectos.

CASO DE USO REAL: DESARROLLAR PROYECTO

The screenshot displays the CMIN - V 1.0 application window. The main workspace shows the 'CUADROS ORGANIZATIVOS IDENTIFICANDO DIVISIONES PROYECTO' step. It includes a table for recording information:

División	Nombre del director	Responsabilidad

Below this table, there is another table with columns 'Departamento', 'Nombre del director', and 'Responsabilidad'. A dialog box from Microsoft Office Word asks: 'Do you want to save the changes to Plantillas?'. To the right, a confirmation dialog asks: 'Desea actualizar la plantilla en el proyecto' with 'Actualizar' and 'Ignorar' buttons.

CURSO NORMAL DE LOS EVENTOS

Acción del actor	Respuesta del sistema
1. El usuario el paso que desea modificar [A].	2. El sistema muestra la ventana que contiene los campos del paso.
3. El usuario modifica el campo si es de texto, si es plantilla la visualiza [B].	
5. El usuario da clic en el botón "Guardar" [C].	6. Si el campo es una plantilla el sistema ejecuta WORD con la plantilla que se desea modificar.
7. El usuario modifica la plantilla según lo necesitado [D].	
8. El usuario cierra WORD después de modificar [E].	9. El sistema de WORD muestra la ventana que pregunta que si se guardaran los cambios si no se ha guardado antes.
10. El usuario da clic en el botón "Si o Yes" [F].	11. El sistema muestra la ventana que pregunta si se guardaran los cambios en el sistema.
12. El usuario da clic en el botón "Actualizar" [G].	13. El sistema muestra la información de cómo quedo guardado en el sistema la plantilla [H] y termina la rutina.
TRAZABILIDAD	
Artefactos anteriores	Artefactos posteriores
<i>Artefactos del Análisis:</i> Diagrama general de casos de uso usuario.	

Tabla 6. Caso de Uso Real Desarrollar Proyecto.

ANEXO B. DIAGRAMAS

6 DIAGRAMA FISICO DEL SISTEMA

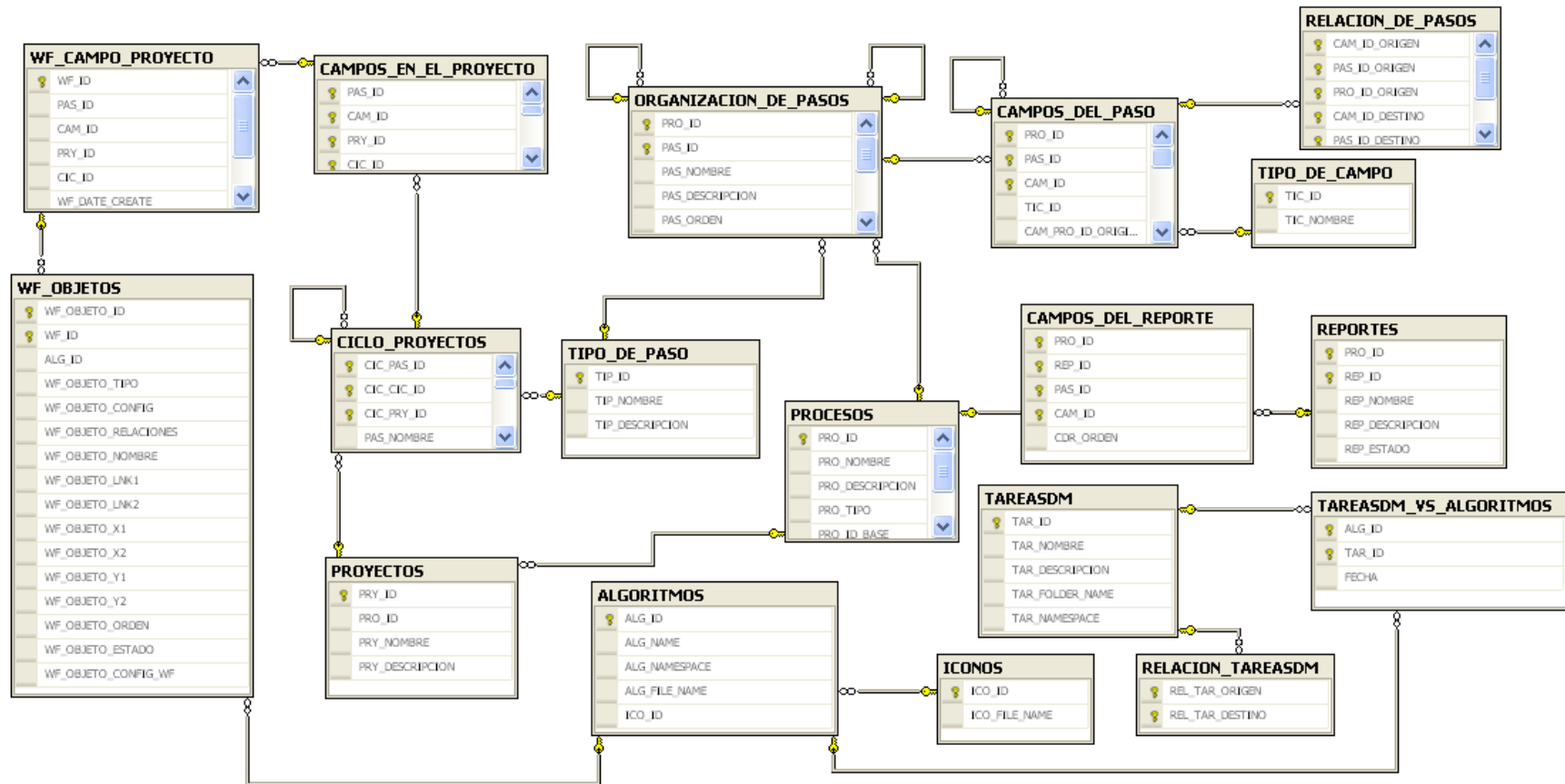


Figura 3. Diagrama Físico de la Base de Datos.

6.1 LISTA DE TABLAS

TABLA	FUNCIÓN
TbIORGANIZACION_DE_PASOS	Contiene la información de los pasos que conforman un proceso o metodología.
TbICAMPOS_DEL_PASO	Contiene la información referente a los campos que pertenecen a un paso.
TbIRELACION_DE_PASOS	Contiene la información referente a las relaciones que existe entre los pasos.
TbITIPO_DE_CAMPO	Contiene la información referente a los tipos de campos existentes en la metodología o proceso.
TbIREPORTES	Contiene la información referente a los reportes que se definen.
TbIPROCESOS	Contiene la información referente a los procesos o metodologías a ingresar.
TbICAMPOS_DEL_REPORTE	Contiene la información que define los campos que pertenecen a un reporte.
TbICICLO_PROYECTOS	Contiene la copia de los pasos de un proceso o metodología para poder utilizarlos en los desarrollos de proyectos.
TbITIPO_DE_PASO	Contiene la información referente a los tipos de pasos pertenecientes aun proceso o metodología.
TbICAMPOS_EN_EL_PROYECTO	Contiene la información generada en cada campo después de ejecutarlo en el desarrollo de un proyecto.
TbIPROYECTOS	Contiene la información referente a los proyectos creados.
TbIWF_CAMPO_PROYECTO	Contiene la información referente a los campos que tienen relacionado un Work Flow.
TbIWF_OBJETOS	Contiene la información referente a los objetos creados en un Work Flow.
TbIALGORITMOS	Contiene la información de los algoritmos de los que se basan los objetos del Work Flow.
TbIICONOS	Contiene la información referente al icono relacionado con un algoritmo.
TbITAREASDM	Contiene la información de las clases de tareas que contienen los algoritmos.
TbIRELACION_TAREASDM	Contiene la información referente a las relaciones que pueden existir entre las tareas.
TbITAREASDM_VS_ALGORITMOS	Contiene la información que especifica a que tarea pertenece un algoritmo.

Tabla 1. Lista de Tablas de la Base de Datos.

7 DIAGRAMAS DE CLASES DEL SISTEMA

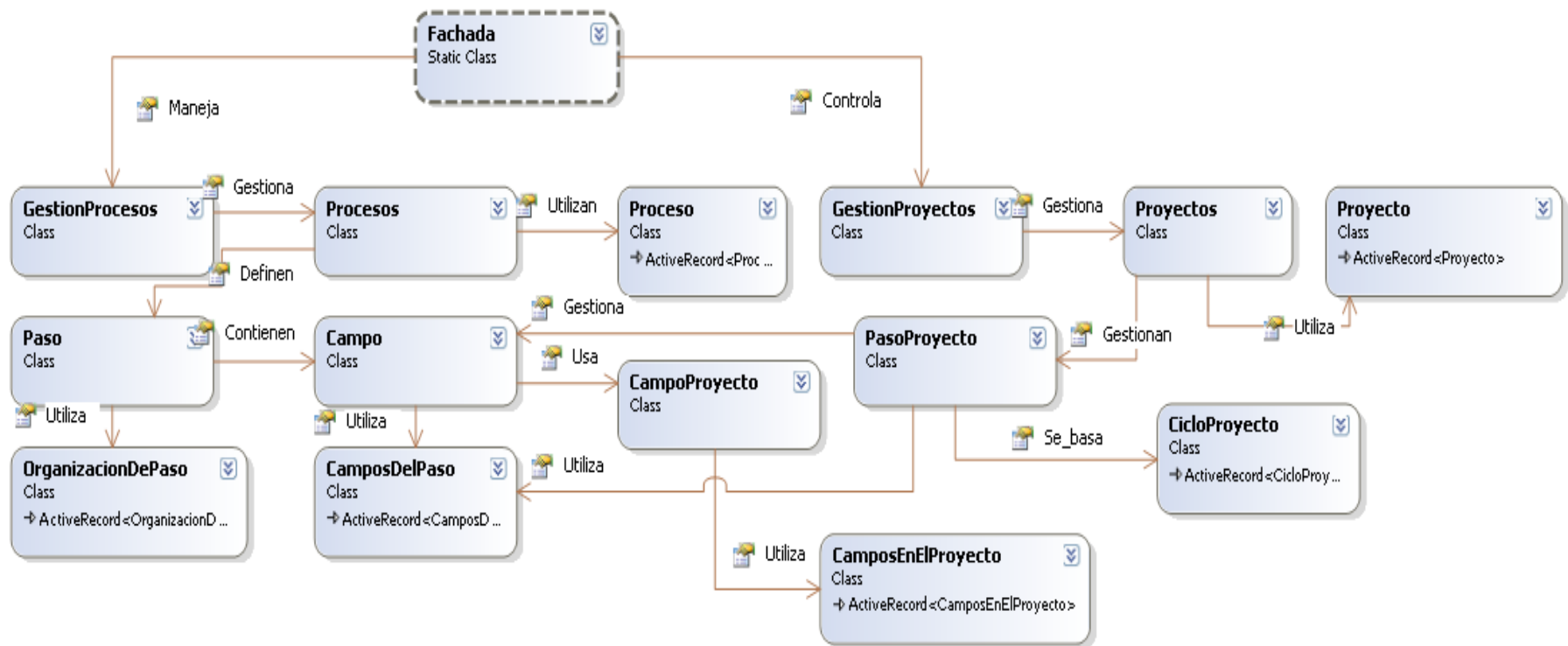


Figura 4. Diagrama de Clases De los Módulos de Gestion de Procesos, Proyectos y Plantillas.

7.1 LISTA DE CLASES

CLASE	FUNCIÓN
Fachada	Provee los servicios para enlazar la capa de interfaz con la capa del negocio.
GestionProcesos	Provee los servicios para la gestión de los procesos.
Procesos	Provee los servicios para la gestión de los pasos de un proceso.
Proceso	Es una entidad que provee la funcionalidad para el manejo de la tabla ROCESOS de la base de datos.
Paso	Provee los servicios para la gestión de los campos de un paso.
OrganizacionDePaso	Es una entidad que provee la funcionalidad para el manejo de la tabla ORGANIZACION_DE_PASOS de la base de datos.
Campo	Provee los servicios para manejar las especificaciones de los campos en un proceso.
CampoDelPaso	Es una entidad que provee la funcionalidad para el manejo de la tabla CAMPOS_DEL_PASO de la base de datos.
CampoProyecto	Provee los servicios para la gestión de los campos que ya han sido ejecutados y proveen de unos resultados.
GestionProyectos	Provee los servicios para la gestión de los proyectos.
CamposEnElProyecto	Es una entidad que provee la funcionalidad para el manejo de la tabla CAMPOS_EN_EL_PROYECTO de la base de datos.
Proyectos	Provee de la lógica para la gestión de los pasos de un proyecto.
Proyecto	Es una entidad que provee la funcionalidad para el manejo de la tabla PROYECTOS de la base de datos.
PasoProyecto	Provee los servicios para la gestión de los pasos de un proyecto en ejecución.
CicloProyecto	Es una entidad que provee la funcionalidad para el manejo de la tabla CICLO_PROYECTOS de la base de datos.

Tabla 2. Lista de Clases del sistema.

8 DIAGRAMA DE CLASES DEL MODULO DE WORK FLOW

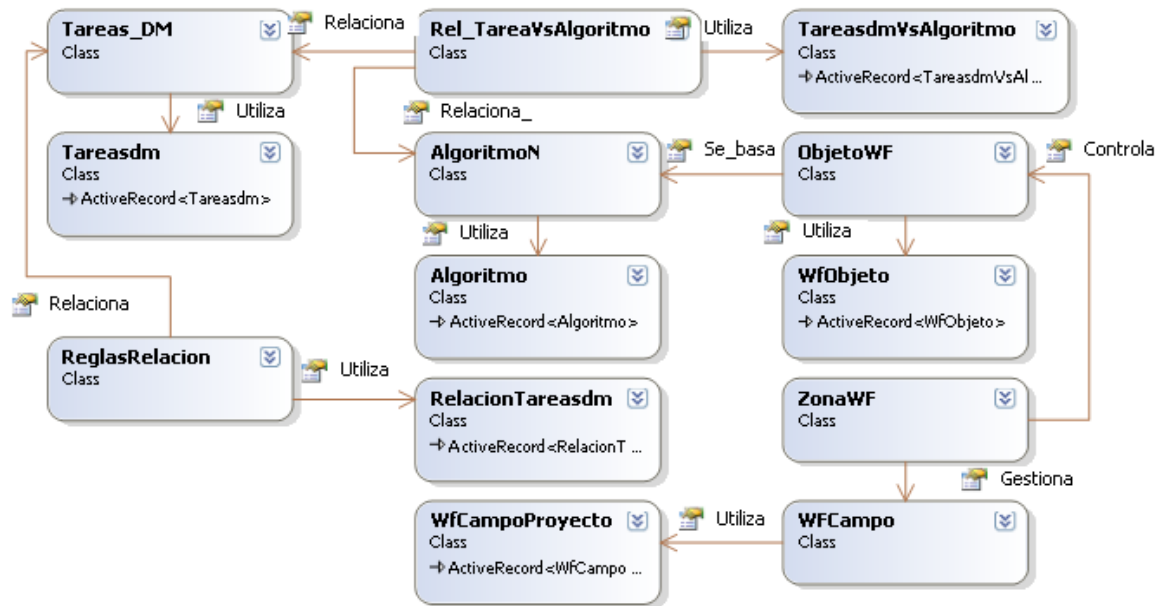


Figura 5 Diagrama de Clases del Modulo de Work Flow.

8.1 LISTA DE CLASES DEL MODULO DE WORK FLOW

CLASE	FUNCIÓN
Treas_DM	Provee los servicios para la gestión de las tareas de minería en el work flow.
Tareasdm	Es una entidad que provee la funcionalidad para el manejo de la tabla TAREASDM de la base de datos.
ReglasRelacion	Provee los servicios para la gestión de las relaciones entre las tareas.
RelacionTareasdm	Es una entidad que provee la funcionalidad para el manejo de la tabla RELACION_TAREASDM de la base de datos.
Rel_TareaVsAlgoritmo	Provee los servicios para la gestión de las relaciones entre las tareas y los algoritmos que pertenecen a estas.
TareasdmVsAlgoritmo	Es una entidad que provee la funcionalidad para el manejo de la tabla TAREASDM_VS_ALGORITMOS de la base de datos.
AlgoritmoN	Provee los servicios para la gestión de los algoritmos.
Algoritmo	Es una entidad que provee la funcionalidad para el manejo de la tabla ALGORITMOS de la base de datos.
ObjetoWF	Provee los servicios para la gestión de los objetos del work flow.
WfObjeto	Es una entidad que provee la funcionalidad para el manejo de la tabla WF_OBJETOS de la base de datos.
ZonaWF	Provee los servicios para la gestión del manejo grafico del work flow.
WFCampo	Provee los servicios para la gestión de los campos que tienen un work flow relacionado.

CLASE	FUNCIÓN
WFCampoProyecto	Es una entidad que provee la funcionalidad para el manejo de la tabla WF_CAMPO_PROYECTO de la base de datos.

Tabla 3. Lista de Clases del Modulo de Work Flow.

ANEXO C. POSTER CIENTÍFICO PRESENTADO EN: Demofest Microsoft Research Academic Summit

9 POSTER CIENTÍFICO



CMIN V1.0 - Herramienta CASE para el soporte de proyectos de Minería de Datos basado en CRISP-DM

C. A. Cobos, J. C. Guarín, J. E. Zuñiga
Departamento de Sistemas, Universidad del Cauca



Abstract

CMIN V1.0 es una herramienta CASE integrada que permite realizar el seguimiento y desarrollo de proyectos de minería de datos a diferentes clases de organizaciones, en especial, porque se basa en una metodología ampliamente confiable, flexible y aprobada a nivel mundial, como lo es CRISP-DM, además cuenta con la capacidad de entender su funcionalidad dinámicamente por parte de cualquier usuario, permitiendo que se pueda adaptar a diferentes situaciones empresariales que no se contemplan inicialmente en la distribución.

Introducción

La minería de datos es una tecnología que se utiliza principalmente para obtener información previamente desconocida en grandes volúmenes de datos, y en los últimos años se han reportado excelentes resultados en su uso en diversos ámbitos de la ciencia y la tecnología, como por ejemplo en medicina, marketing y agricultura. Para la explotación de esta tecnología existen diferentes herramientas, las cuales son tema central de este proyecto debido a que se han determinado varios inconvenientes en ellas [1] [2] [3] entre los más relevantes:

Desarrollo Interfaz

No cuentan con un método interactivo para guiar al usuario en el proceso de minería de datos, lo que es una dificultad para usuarios que tienen poco conocimiento del proceso.

Metodología

No se basan en una metodología estándar para seguir el proceso de minería de datos, sino por el contrario usan una propia, lo que es una desventaja, debido a que actualmente existen metodologías estándar para el proceso de minería de datos que pretenden facilitar la realización de nuevos proyectos con características similares, optimizar la planificación y dirección de los mismos, reducir su complejidad y permitir realizar un mejor seguimiento a estos [4]. Entre las que se destacan CRISP-DM (Cross – Industry Standard Process for Data Mining) y SEMMA (Simple, Explore, Modify, Model, Assess) [5]. SEMMA se centra más en las características técnicas del desarrollo del proceso, mientras que CRISP-DM mantiene como foco los objetivos empresariales del proyecto.

Extensibilidad

No cuentan con la capacidad de adicionar dinámicamente algunos al conjunto de tareas que se entregan en cada versión. Las herramientas evolucionan con versiones, al ritmo de las empresas que las producen y no permiten un desarrollo en comunidad similar al de software libre.

Herramienta	Acceso	Desarrollo Interfaz	Metodología	Extensibilidad
Clustream	Catálogo	Sencillo	Parte de CRISP-DM	No, sólo con el propio software
Intelligent Miner	Catálogo	Complejo	PRODA	No, sólo con el propio software
WEKA	Libre	Complejo	PRODA	No, por su licencia open source
CRIT	Catálogo	Complejo	PRODA	No, sólo con el propio software
Procesador	Catálogo	Complejo	PRODA	No, sólo con el propio software
MSI	Catálogo	Sencillo	SEMMA	No, sólo con el propio software

Tabla 1 Herramientas evaluadas

Metodologías Para Proyectos de Minería de Datos

Entre las que se destacan están CRISP-DM (Cross – Industry Standard Process for Data Mining) [5] y SEMMA (Simple, Explore, Modify, Model, Assess) [6]. SEMMA se centra más en las características técnicas del desarrollo del proceso, mientras que CRISP-DM mantiene como foco los objetivos empresariales del proyecto. La diferencia se establece ya desde la primera fase donde SEMMA comienza realizando un muestreo de datos, mientras que CRISP-DM comienza realizando un análisis del problema empresarial para su transformación en un problema técnico. Desde ese punto de vista se puede considerar que CRISP-DM está más cercano al concepto real de un proyecto de minería de datos, pudiendo ser integrada con una Metodología de Gestión de proyectos específica que completaría las tareas administrativas y técnicas, además CRISP-DM es de libre distribución sin ningún costo a diferencia de SEMMA[4].

CMIN V1.0 Estado Actual

CMIN V1.0 tiene como base la metodología CRISP-DM, para convertir en un Proceso guía para los proyectos en la Herramienta, se realizó lo siguiente:

PASO1

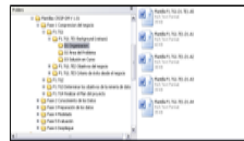


Fig 2 Plantillas por cada actividad de CRISP-DM

Consiste en crear Plantillas por cada Actividad propuesta en la metodología, que requieren ser Documentada.

PASO2

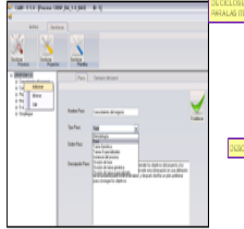


Fig 3 Adición de Pasos de la Metodología.

Adicionar los Pasos definidos en la Metodología CRISP-DM al Proceso en la CASE desde se almacena información como tipo de Paso (Metodología, Fase, Tarea General, etc.) descripción, Orden etc. Tal y como se muestra en la figura 3.

PASO3

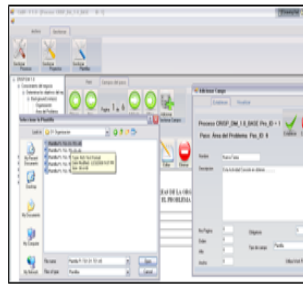


Fig 4 Adicionar campos a los Pasos.

Consiste en adicionar a cada Paso creado, los campos o Tareas establecidas en la Metodología para este. Cada tarea a las cuales seguimos la Plantilla específica (creada en el Paso 1) si es requerido y establecemos la descripción, se utiliza Workflow o entre otros. Las Tareas son almacenadas en el sistema junto con sus plantillas. Mostrado en la figura 4.

Proyecto

Ahora ya estamos listos para la creación de Proyectos que se basan en CRISP-DM. De esta manera CMIN V1.0 permite almacenar la información del Proyecto en cada actividad de la metodología guía. Además como en todo Proceso se tienen las mismas: CMIN permite la creación de ciclos de cualquier Paso e igualmente almacenar sus resultados. Es de notar que la descripción del Paso y la Tarea o Actividad son presentadas al usuario para su comprensión. Verifiquemos 5.

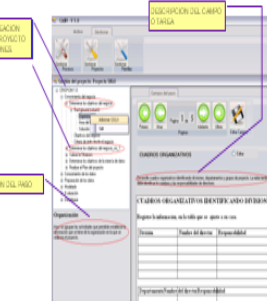
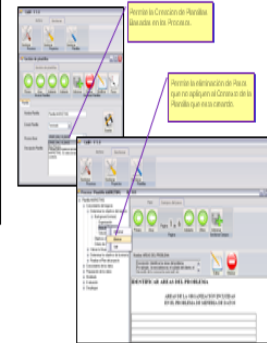


Fig 5 Proyectos CMIN.

Plantillas los Proyectos no solo pueden basarse en un Proceso también pueden hacerlo de una instancia de algún Proceso o Plantilla. Debido a que los procesos son metodologías generales que tratan de abarcar todas las posibles situaciones de un proyecto de minería de datos, las instancias o plantillas tratan de seleccionar lo que necesitan de la Metodología para un proyecto o contexto específico.



Conclusiones

- >CMIN V1.0 permite la adición de cualquier Metodología para proyectos de minería de datos. Es decir puede tener lugar adicionar como otro proceso base CRISP-DM Version 2 u otra.
- >El aprendizaje del Proceso de minería de datos es casi asegurado por quien usa esta herramienta.

Trabajo Futuro

En la fecha el desarrollo del Proyecto está centrado en la creación del Workflow el cual es un hit para las tareas de minería de datos, se pretende que sea fácil de usar y que permita la adición de algunos dinámicamente. Además la Arquitectura de la herramienta está pensada para que por cada campo o tarea que necesite del Workflow, se almacene el trabajo realizado, es decir permitirá ver como avanza el Workflow del Proyecto a través de los pasos.

REFERENCIAS

[1] Microsoft (M) [Aplicación: Minería de Datos]. (Citado: 15 Julio de 2007).
[2] Kluugroup (2004) [Minería de Datos]. (Citado: 2007, 15 Agosto)
[3] Briva, P., Brivante, E., Ochoa, M., Molino, H., Ochoa, E., y García, R. METODOLOGÍA DE SELECCIÓN DE HERRAMIENTAS DE EXPLORACIÓN DE DATOS. Dpto. Ingeniería del Software. Facultad de Ingeniería. Universidad de Cauca. (Citado: 2005, 14 Julio).
[4] Guadalupe, J. del Valle. (Citado: 2007, 10 Mayo). Metodología para la explotación de Proyectos de Datos Mining. [Aplicación: Minería de Datos].
[5] Proceso CRISP-DM (CRISP-DM). (Citado: 2006, 01 Noviembre). [Página: Introducción]. URL: <http://www.crisp-dm.org/>
[6] SAS (SASMA). (Citado: 2006, 01 Noviembre). [Página: Introducción]. URL: <http://www.sas.com/>

Figure 6. Póster Científico

ANEXO D. REPORTE: Evaluación de Prototipo CMIN por estudiantes de la Electiva Minería de Datos.

10 RESULTADOS DE LA PRUEBA

Los resultados después de la prueba a un prototipo de CMIN 1.0 por los estudiantes se presentan a continuación, es de anotar que los reportes se incorporan como originalmente fueron enviados por los estudiantes.

Cada grupo que se organizó para realizar la evolución encargándose de evaluar una fase específica.

10.1 FASE 1

Entendimiento del negocio Revisión de la herramienta CMIN V 1.0 Minería de Datos

Taller 2

**Alba Viviana Camayo Otero
Adrian Fernando Martinez Molina**

11 de Febrero de 2008

Revisión De La Herramienta CMIN 1.0 En La Etapa De Entendimiento Del Negocio

Fase 1: Determinar los objetivos del Negocio					
Tarea	Salida	Actividad	Se encuentra:		Observaciones
			SI	NO	
Determinar los objetivos del negocio	Background	Organización:			
		Elaboración de organigramas identificando los nombres y responsabilidades de los administradores.	X		Se considera que las tablas propuestas no reflejan de una manera clara los datos requeridos por la actividad
		Identificar las personas claves de la empresa y sus funciones.	X		
		Identificar un sponsor interno (patrocinador financiero y un usuario principal / dominio de expertos)	X		
		Identificar la existencia de un comité de dirección y quienes son los miembros	X		
		Identificación de las unidades de negocio que son impactadas	X		
		Área del Problema:			
		Identificación del problema	X		En esta actividad se considera que se evalúen los títulos asignados a la tabla y que se adicione una columna de descripción

		Descripción del problema en términos generales.		X	
		Comprobar el estado actual del proyecto		X	
		Aclarar los prerrequisitos del proyecto		X	Estas actividades pueden ser sugerencias para el usuario
		Preparación de presentaciones y presentación de la minería de datos para el negocio. (opcional)		X	
		Identificar los grupos objetivo para el resultado del proyecto	X		
		Identificar las necesidades y expectativas de los usuarios	X		
		Solución Actual:			
		Descripción de cualquier solución utilizada actualmente para el problema.		X	Esta solo como un titulo, debería haber un campo o columna para llenar.
		Descripción de las ventajas y desventajas de la solución actual y el nivel de aceptación de los usuarios.	X		
	Objetivos del negocio	Descripción informal del problema que se supone que hay que resolver con la minería de datos	X		Se deberían separar porque no cuando se llene la información no se va logra diferenciar entre cada ítem.
		Especificación de todas las cuestiones de negocios con la mayor precisión posible.	X		
		Especificación de otros requerimientos del negocio (por ejemplo, la empresa no quiere perder ninguno de sus clientes)	X		
		Especificación de los beneficios previstos en términos del negocio	X		
	Criterios de éxito del negocio				
		Especificación los criterios de éxito de la empresa	X		
		Identificar quien evalúa los criterios de éxito.		X	Se debería agregar una columna para esta descripción.
Fase 2: Evaluar la situación					

Tarea	Salida	Actividad	Se encuentra:		Observaciones	
			SI	NO		
Inventario de los recursos		Recursos Hardware:				
		Identificación del hardware base.	X			
		Establecer la disponibilidad del hardware base para el proyecto de minería de datos.	X			
		Revise si el calendario del mantenimiento del hardware presenta conflictos con la disponibilidad del hardware para el proyecto de minería de datos.		X	Estas actividades pueden ser sugerencias para el usuario	
		Identificar el hardware disponible para la herramienta de minería de datos que se utilizaran. Fuentes de datos y conocimiento.		X		
		Identificar las fuentes de datos	X		La ubicación y el tipo de las fuentes de datos puede estar en una sola tabla	
		Identificar el tipo de fuentes de datos.	X			
		Identificar fuentes de conocimiento.	X		La ubicación y el tipo de las fuentes de conocimiento puede estar en una sola tabla	
		Identificar el tipo de fuentes de conocimiento.	X			
		Revise las herramientas y las técnicas disponibles.		X	La tabla para este paso puede ser una lista de las herramientas y las técnicas disponibles.	
		Describir los antecedentes relevantes al conocimiento (informal o formal) fuentes personales.		X		
		Identificar los sponsor del proyecto	X			
		Identificar el administrador del sistema, administrador de bases de datos y el equipo de soporte de técnico para preguntas futuras.	X			
		Identificar los analistas de mercado, los expertos en minería de datos y estadística y comprobar	X			

		su disponibilidad.			
		Ver la disponibilidad de los expertos del dominio para las fases posteriores.	X		
Evaluar la situación	Requisitos, supuestos y limitaciones	Requisitos:			
		Especificar perfil del público objetivo.	X		(La tabla esta muy a la izquierda, no se ve la "N" de nombre).
		Capturar todos los requisitos dentro de la programación.	X		Una tabla para poder organizar mejor la información, cuando sean llenados todos los datos.
		Captura de los requisitos teniendo en cuenta la comprensibilidad, la precisión, capacidad de despliegue, mantenimiento y redundancia de los datos del proyecto de minería de datos y el modelo resultante.	X		Una tabla para poder organizar mejor la información, cuando sean llenados todos los datos.
		Captura de los requisitos teniendo en cuenta la seguridad, las restricciones legales, de privacidad, la presentación de informes y el calendario del proyecto	X		Una tabla para poder organizar mejor la información, cuando sean llenados todos los datos.
		Supuestos:			
		Aclarar todas las hipótesis (incluida la implícita) y hacerla explícita.		X	Estas actividades pueden ser sugerencias para el usuario
		Lista de las hipótesis sobre la calidad de los datos.	X		El nombre no corresponde a la actividad
		Lista de las hipótesis sobre los factores externos.	X		
		Clarificación de cualquiera de los supuestos que dan lugar a cualquiera de las estimaciones.		X	Estas actividades pueden ser sugerencias para el usuario
		Lista de todas las hipótesis sobre si es necesario entender y describir o explicar el modelo.	X		
		Limitaciones:			

		Revise las limitaciones generales.	X		Se sugiere que esta información este contenida en una tabla	
		Revise los permisos de acceso a las fuentes de datos.	X			
		Revise la accesibilidad técnica de los datos (sistemas operativos, sistema de gestión de datos, formato de la base de datos o de los archivos).	X		Todo lo relevante a la disponibilidad y acceso a las fuentes de datos se puede encontrar en una sola tabla	
		Revise si se tiene acceso a los conocimientos requeridos.	X		Todo lo relevante a la disponibilidad y acceso a las fuentes de conocimiento se puede encontrar en una sola tabla	
		Revise las limitaciones del presupuesto (costos fijos, gastos de ejecución, etc.).	X		Se sugiere que esta información este contenida en una tabla	
	riesgos y contingencias	Identificar los riesgos:				
		Identificar los riesgos del negocio.	X		Se sugiere que esta información este contenida en una tabla, igual se cree que se debe proporcionar una opción que le permita clasificar al usuario el impacto del riesgo	
		Identificar los riesgos de la organización.	X			
		Identificar los riesgos financieros.	X			
		Identificar riesgos técnicos.	X			
		Identificar los riesgos que dependen de los datos y las fuentes de datos.	X			
		Elaborar planes de contingencia:				
		Determinar las condiciones en la que cada riesgo puede ocurrir.	X		Se sugiere que esta información este contenida en una tabla	
	Elaborar los planes de contingencia.	X		Se sugiere que sean vínculos a documentos		
	Terminología	Revise la disponibilidad de glosarios previos, de otro proyecto antes de comenzar a construir los glosarios.	X	X	Estas actividades pueden ser sugerencias para el usuario. Se encuentra una tabla en la cual se encuentra el término y una descripción.	
		Hable con el experto en el tema para comprender su terminología.		X		
		Familiarícese con la terminología del negocio.		X		

	costos y beneficios	Estimación del costo de la recolección de los datos.	X		Se sugiere que la tabla contenga la clasificación de los costos estimados, propuestos por el modelo
		Estimación de los costos de la elaboración y aplicación de la solución.	X		
		Estimación para gastos de funcionamiento.	X		
		Identificar los beneficios del despliegue de la solución.	X		Se sugiere que esta información este contenida en una tabla
Fase 3: Determinar los objetivos de la minería de datos					
Tarea	Salida	Actividad	Se encuentra:		Observaciones
			SI	NO	
Determinar los objetivos de la minería de datos	objetivos de la minería de datos	Traducir las preguntas del negocio a objetivos de la minería de datos	X		Se sugiere que esta información este contenida en una tabla
		Especificar el tipo de problema de minería de datos		X	
	criterios de éxito para la minería de datos	Especificar los criterios de evaluación del modelo		X	
		Definir puntos de referencia para los criterios de evaluación	X		
		Especificar los criterios que se refieren a la evaluación de criterios subjetivos		X	
	Fase 4: Producir Plan del Proyecto				
Tarea	Salida	Actividad	Se encuentra:		Observaciones
			SI	NO	
Producir plan del proyecto	Plan del proyecto	Definir el proceso inicial y discutir el plan de viabilidad con todo el personal involucrado.		X	El plan del proyecto es un documento muy importante que define los lineamientos del proceso a seguir en el proyecto, se sugiere que lo que se encuentre en esta actividad, sea la estructura básica del documento, pero el documento en si debe ser un link a un documento más formal.
		Coloque todos los objetivos identificados y las técnicas seleccionadas para formar un procedimiento coherente que resuelva las preguntas del negocio y que reúna los criterios de éxito del negocio.		X	
		Estimación del esfuerzo y los recursos necesarios para alcanzar e implementar la solución		X	
		Identificación los pasos críticos		X	

		Marcar los puntos de decisión		X		
		Marcar los puntos de revisión		X		
		Identificación de las principales iteraciones		X		
	evaluación inicial de las técnicas y herramientas		Crear una lista de criterios de selección de los instrumentos y de las técnicas		X	Se sugiere que esta información este contenida en una tabla y hace falta el criterio de selección
			Elección de las técnicas y herramientas potenciales.	X		Se sugiere que se adicione una columna que permita saber porque estas técnicas y herramientas fueron seleccionadas.
			Evaluar que tan adecuadas son las técnicas		X	Estas actividades pueden ser sugerencias para el usuario.
		Examinar y priorizar las técnicas aplicables de acuerdo a la evaluación de las alternativas de solución	X			

Recomendaciones Adicionales:

1. **agrupar las actividades de una salida, en pestañas o incluir los botones de navegación (adelante y atrás)** en el mismo marco de las actividades porque cuando uno ingresa por primera vez no se da cuenta que hay más actividades y quitar los textos de explicación de la funcionalidad de algunos botones.
2. las tablas se encuentran muy mal justificadas y en muchas partes pasa que se pierden algunas letras, además las tablas y los textos se pueden borrar por completo y cuando se escribe sobre las tablas no se tiene límite de escritura y se sigue escribiendo en la columna continua y fuera de la tabla.
3. la sintaxis de los textos y la semántica no esta muy bien estructurada y los mensajes a veces no son claros, además el tipo de letra no es muy adecuada y debería ser de la **misma fuente** del resto de la aplicación.

10.2 FASE 2

DOCUMENTO DE EVALUACION DE LA HERRAMIENTA C-MIN PARA LA ETAPA DE CONOCIMIENTO DE LOS DATOS

Por
Oscar Eduardo Rendón
Alexander Ortiz Rosada

En este documento se realiza una evaluación del nivel de seguimiento que presenta la herramienta CASE C-MIN v1.0 respecto a la metodología CRISP-DM para la fase de *Entendimiento de los Datos*, la cual corresponde a la segunda fase dentro del ciclo de vida para un proyecto de minería de datos propuesto por la metodología mencionada.

La fase de *Entendimiento de los Datos* consta de 4 etapas, a saber:

- Recolección inicial de los datos.
- Descripción de los datos.
- Exploración de los datos.
- Verificación de los datos.

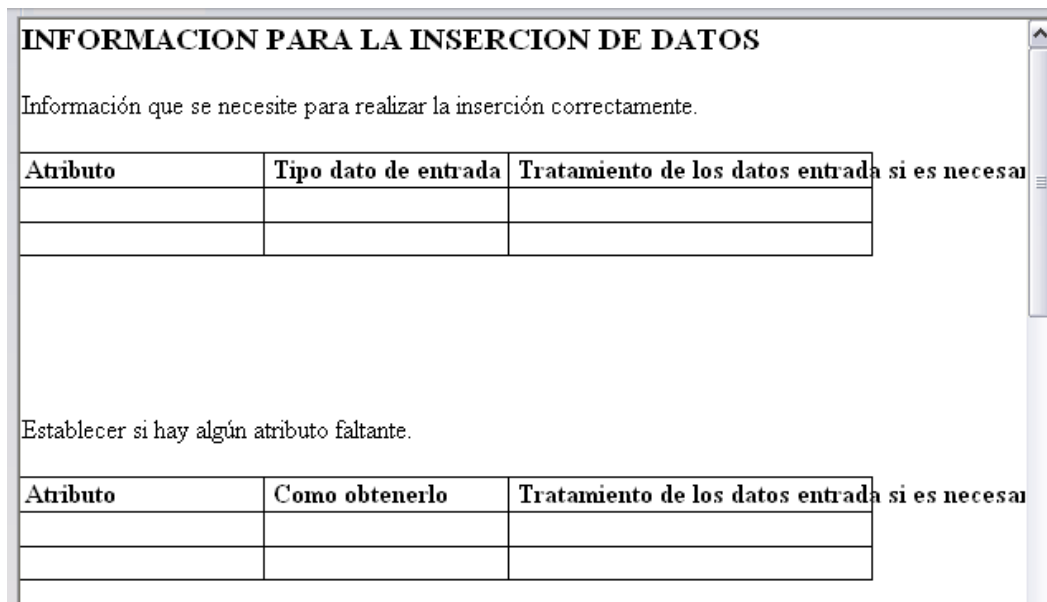
La herramienta presenta claramente la identificación de estas etapas, y ofrece la posibilidad de realizar los documentos correspondientes a los reportes de cada tarea de estas etapas.

VALORACION DEL SEGUIMIENTO DE LAS ETAPAS.

Etapas 1.

RECOLECCION INICIAL DE LOS DATOS.

- La herramienta proporciona tablas para la definición de definición de los datos, pero no permite especificar la disponibilidad y las fuentes de los datos. La descripción de las tablas y su usabilidad aun son pobres.
- Para las tareas correspondientes a criterios de selección e inserción de los datos, la herramienta debería incluir valores predeterminados para la elección de tipos y descripción de tablas y atributos.



INFORMACION PARA LA INSERCIÓN DE DATOS

Información que se necesite para realizar la inserción correctamente.

Atributo	Tipo dato de entrada	Tratamiento de los datos entrada si es necesar

Establecer si hay algún atributo faltante.

Atributo	Como obtenerlo	Tratamiento de los datos entrada si es necesar

Etapa 2.

DESCRIPCION DE LOS DATOS.

- La herramienta proporciona tabla para registrar las actividades captura de datos y evaluación de supuestos y objetivos, sin embargo la metodología sugiere descripciones a nivel estadístico de los datos, que no son tomados en cuenta por la herramienta y que son de gran importancia en la definición de los datos.
- Para la formulación de supuestos se debería tener en cuenta un reporte mucho más detallado y no limitarlos a la utilización de una tabla, la herramienta podría sugerir formatos predeterminados para estas tareas.
- La herramienta podría utilizar listas de chequeo u otros controles para los atributos o tablas en fin de determinar cuáles resultan definitivos para la minería de datos.
- La herramienta no presenta un modulo para el análisis de atributos clave y sus relaciones.

Etapa 3.

EXPLORACION DE LOS DATOS.

- La herramienta no cuenta con una funcionalidad que guie la evaluación de los datos, su descripción en detalle y la identificación de sub-poblaciones, para este modulo la herramienta debería incluir controles gráficos.
- La herramienta no cuenta con funcionalidad que guie la formulación de hipótesis, ni la redefinición de los objetivos de la minería de datos basados en hipótesis.

Etapa 4.

VERIFICACION DE LOS DATOS.

- La herramienta provee el manejo de la información que reside en archivos planos y define delimitadores y uso de delimitadores.
- La herramienta debería permitir la validación y la confrontación de datos provenientes de fuentes distintas.
- La herramienta no cuenta con funcionalidad que permita plantear estrategias encaminadas al manejo del ruido.
- El manejo de los datos faltantes parece claro.

CONSIDERACIONES GENERALES

La herramienta presenta problemas evidentes de usabilidad, entendibles por la temprana etapa en la que se encuentra el desarrollo de la misma, sin embargo se debe tener en cuenta que una herramienta CASE debe facilitar al usuario la aplicación de una metodología y debe en lo posible evitar que el desarrollador cometa errores en el diseño, así, es necesario que utilizar controles que conduzcan al usuario a seleccionar parámetros en lugar de digitarlos. Siguiendo consideraciones de usabilidad en el diseño de interfaces, los comandos, opciones o botones no deberían combinar texto con imágenes, es decir, que los iconos utilizados deben ser los suficientemente claros para que no tengan necesidad de incluir texto, en caso de necesitarlo el texto debe ir, como en efecto lo cumple la herramienta, en controles tooltip.

10.3 FASE 3

MINERIA DE DATOS

TALLER 2: EVALUACION CMIN

FASE 3 CRISP-DM: PREPARACION DE LOS DATOS

German Velasco cod. 46012223

Jose Luis López cod. 46012111

PREPARACIÓN DE DATOS

- La herramienta se probó con una resolución de 1024 x 768 píxeles.

1. Datos seleccionados

Decidir los datos a ser usados para el análisis. Los criterios incluyen la importancia a los objetivos de minería de datos, la calidad, y las restricciones técnicas como los límites en el volumen de datos o en los tipos de datos.

Salida Razonamiento para inclusión/exclusión

Listar los datos a ser usados / excluidos y los motivos para estas decisiones.

- La herramienta muestra los datos, pero no tiene forma de saber cuales han sido incluidos y cuales excluidos.
- Se debe usar mucho el scrollbar

Actividades

- Recogen datos adicionales apropiados (de diferentes fuentes - internos así como externos)
 - La herramienta permite especificar los datos y las fuentes de datos de las cuales se han obtenido.
- Realizar las pruebas de importancia y correlación para decidir si los campos son incluidos
 - La herramienta permite registrar las pruebas significativas y de correlación, para futuros usos.
- Reconsiderar Criterios de Selección de Datos en la luz de experiencia de modelado (esto es, la evaluación del modelo puede mostrar que otros conjuntos de datos son necesarios)
 - No existe manera de registrar si existe algún cambio en los criterios de selección y las razones del cambio.
 - **Sugerencia: implementar un mecanismo que permita importar el listado de criterios, además se puedan editar, adicionando comentarios de las razones para el cambio del criterio.**
- Seleccionar diferentes subconjuntos de datos (por ejemplo, atributos diferentes, sólo los datos que encuentran ciertas condiciones)
 - La herramienta no tiene mecanismos para cumplir esta actividad.
- Considerar el uso de técnicas de muestreo
 - La herramienta no tiene mecanismos para cumplir esta actividad.
 - Sugerencia: La herramienta podría mostrar un listado de las técnicas de muestreo, con una pequeña descripción de cada una.
- Documentar el razonamiento para la inclusión/exclusión

- La herramienta no tiene mecanismos para cumplir esta actividad.
- **Sugerencia: implementar una plantilla con el listado de los datos, que permita incluir comentarios sobre las razones de inclusión/exclusión.**
- Comprobar técnicas disponibles para el muestreo de datos
 - La herramienta no tiene mecanismos para cumplir esta actividad.
 - Sugerencia: La herramienta podría mostrar un listado de las técnicas de muestreo, con una pequeña descripción de cada una.

2. Limpiar datos

Elevar la calidad de datos al nivel requerido por las técnicas de análisis seleccionadas. Esto puede implicar la selección de subconjuntos limpios de los datos, la inserción de faltas apropiadas, o técnicas más ambiciosas como la estimación de datos omitidos por modelado.

Salida Informe de la limpieza de datos

Describir las decisiones y las acciones que fueron tomados para dirigir los problemas de calidad de datos informados durante la Tarea de Verificación de Calidad de Datos. Si los datos están para ser usados en el ejercicio de minería de datos, el informe debería dirigir cuestiones de calidad de datos excepcionales y el efecto posible que esto podría tener sobre los resultados.

- La herramienta si permite generar el informe de limpieza de datos.

Actividades

- Reconsiderar como tratar con cualquier tipo de ruido observado
 - La herramienta no tiene mecanismos para cumplir esta actividad.
 - Sugerencia: incluir un listado de los posibles “ruidos” en los datos, dando una explicación clara de cada uno.
- Corregir, remover, o ignorar el ruido
 - La herramienta no tiene mecanismos para cumplir esta actividad.
 - Sugerencia: implementar un mecanismo que permite registrar los cambios hechos en los datos por causa del “ruido” en los mismos.
- Decidir como tratar con valores especiales y su significado.
 - La herramienta permite registrar los valores especiales y las decisiones que se tomaron sobre ellos, **no se puede registrar el significado de los valores especiales.**
 - Sugerencia: brindar al usuario herramientas para facilitar la decisión en base al significado del valor especial.
- Reconsiderar Criterios de Selección de Datos en la luz de las experiencias de los datos limpiados (esto es, usted puede desea incluir/excluir otros conjuntos de datos).
 - **No existe manera de registrar si existe algún cambio en los criterios de selección y las razones del cambio.**
 - Sugerencia: implementar un mecanismo que permita importar el listado de criterios, además se puedan editar, adicionando comentarios de las razones para el cambio del criterio.

3. Construir datos

Esta tarea incluye la construir de operaciones de preparación de datos tales como la producción de atributos derivados, completar registros nuevos, o transformar valores para atributos existentes.

Actividades

- Comprobar los mecanismos de construcción disponibles con la lista de herramientas sugeridas para el proyecto
 - La herramienta no tiene mecanismos para cumplir esta actividad.
 - Sugerencia: presentar un listado de herramientas con una descripción, dando al usuario mecanismos para su elección, para futuras verificaciones.
- Decidir si esto es lo mejor para realizar la construcción dentro de la herramienta o fuera de ella (esto es, que es más eficiente, exacto, repetible)
 - La herramienta no tiene mecanismos para cumplir esta actividad.
- Reconsiderar Criterios de Selección de Datos en la luz de las experiencias de construcción de datos (esto es, usted puede desear incluir/excluir otros conjuntos de datos)
 - No existe manera de registrar si existe algún cambio en los criterios de selección y las razones del cambio.
 - Sugerencia: implementar un mecanismo que permita importar el listado de criterios, además se puedan editar, adicionando comentarios de las razones para el cambio del criterio.

Salida Atributos derivados

Los atributos derivados son los atributos nuevos que son construidos de uno o atributos más existentes en el mismo registro

- La herramienta permite ver el listado de atributos derivados

Actividades Derivar atributos

- Decidir si cualquier atributo puede ser normalizado
 - La herramienta permite ver el listado de atributos normalizados
- Considerar agregar nueva información sobre la importancia relevante de los atributos para agregar de nuevos atributos
 - La herramienta permite adicionar características de los atributos normalizados
- Agregar atributos nuevos a los datos de acceso
 - La herramienta permite ver el listado de atributos agregados

Actividades Transformaciones de atributo individual

- Especificar los pasos de transformaciones necesarias en los términos de facilitar las transformación disponibles
 - La herramienta permite ver el listado de pasos para transformar los datos.
- Realizar pasos de transformación
 - La herramienta permite ver el listado de pasos para transformar los datos.
 - Sugerencia: implementar mecanismos que permita marcar los pasos que el usuario ha realizado, y permitir comentar los posibles errores y dificultades presentados al realizar los pasos.

Salida Registros generados

Los registros generados son registros completamente nuevos, que agregan nuevo conocimiento o representan nuevos datos que de otro modo no son representado (por ejemplo, habiendo segmentado los datos, puede ser útil generar un registro para represente al miembro prototípico de cada segmento para un tratamiento futuro).

- La herramienta provee la opción, pero no tienen implementado ninguna plantilla, ni ningún mecanismo para ver esta salida.

Actividades

- Comprobar por técnicas disponibles si es necesario (por ejemplo, mecanismos para construir prototipos para cada segmento de datos segmentados).
 - La herramienta no tiene mecanismos para cumplir esta actividad.

4. Integrar datos

Estos son métodos para combinar la información de múltiples tablas u otras fuentes de información para crear nuevos registros o valores.

Salida Datos combinados

La combinación de tablas se refiere a la unión de dos o más tablas que tienen diferente información sobre los mismos objetos. En esta etapa, también puede ser aconsejable generar registros nuevos.

También puede ser recomendado para generar valores agregados.

La agregación se refiere a operaciones donde los nuevos valores son calculados por información resumida de múltiples registros y/o tablas.

- La herramienta provee la opción, pero no tienen implementado ninguna plantilla, ni ningún mecanismo para ver esta salida.

Actividades

- Comprobar si las aplicaciones de integración son capaces de integrar las fuentes de entrada como se requiere
 - La herramienta no tiene mecanismos para cumplir esta actividad.
- Integrar fuentes y resultados almacenados
 - La herramienta no tiene mecanismos para cumplir esta actividad.
- Reconsiderar Criterios de Selección de Datos en la luz de las experiencias de integración de datos (esto es, usted puede desear incluir/excluir otros conjuntos de datos)
 - No existe manera de registrar si existe algún cambio en los criterios de selección y las razones del cambio.
 - Sugerencia: implementar un mecanismo que permita importar el listado de criterios, además se puedan editar, adicionando comentarios de las razones para el cambio del criterio.

10.4 FASE 4

EVALUACION DE HERRAMIENTA CMIN

**JENNIFER ANDRADE
WILLIAM CONSTAIN**

INGENIERO CARLOS ALBERTO COBOS L.

**UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA DE TELECOMUNICACIONES
PROGRAMA DE INGENIERIA DE SISTEMAS
ELECTIVA MINERIA DE DATOS
POPAYÁN
2008**

EVALUACION DE HERRAMIENTA CMIN

Fase de Modelado

La fase de modelado de la Metodología CRISP-DM comprende cuatro tareas, en donde cada una ellas genera sus respectivas las salidas. El objetivo del presente informe es dar desde nuestro punto de vista, sugerencias a cerca de la herramienta CMIN en cuanto la presentación del contenido y al cumplimiento de su función como herramienta que guía la documentación de un proceso de minería de datos.

Tareas

1. Seleccionar las Técnicas de Modelado

Salida Técnicas de Modelado

Creemos que en la documentación de esta salida debería también presentarse una plantilla sencilla que guíe al usuario en la documentación de la técnica escogida. Además dado que los usuarios podrían utilizar CMIN solo para documentar un proceso de minería y realizar este proceso como tal en otra herramienta, debería poder especificar la herramienta que se utilizó.

Los campos que podría presentar la plantilla serían:

Nombre de la técnica:	
Breve descripción de la técnica:	
Razones por las que se eligió la técnica:	
Herramienta de minería de datos en la que se implementa la técnica:	

Salida Suposiciones del Modelado

Comentario 1

En cuanto a las suposiciones del modelado también nos parece que debería existir una plantilla que guíe al usuario, porque es más fácil que la documentación quede de forma ordenada:

La plantilla podría ser como se muestra a continuación:

Suposiciones de la	Definir suposiciones resultantes por la
---------------------------	---

Técnica:	técnica de minería que se utiliza
Comparación de suposiciones:	Comparar con las suposiciones hechas en el informe de descripción de los datos
Conclusiones de las suposiciones	Establecer si las suposiciones son validas y si es necesario concluir volver la fase de preparación de datos

Comentario 2

Por otra parte, en la salida "*Modelado de Suposiciones*", encontramos que cuando se quiere documentar un Modelado de Suposiciones asociado a una técnica en particular hay una situación que consideramos no debería ser así, la cual explicamos con un ejemplo: cuando tengo cuatro técnicas de modelado documentadas y solo a la primera y a la última técnica les quiero asociar un modelado de suposiciones, debo primero crear un modelado de suposiciones para cada una de las técnicas anteriores a la última a la que le deseo asociar un modelado de suposiciones.

Creemos que cada técnica de modelado documentada al darle clic derecho debería permitir crear un modelado de suposiciones asociado a ella, o cada modelado de suposiciones documentado debería permitir que se edite su nombre. La figura 1 muestra esta situación presentada en la herramienta.

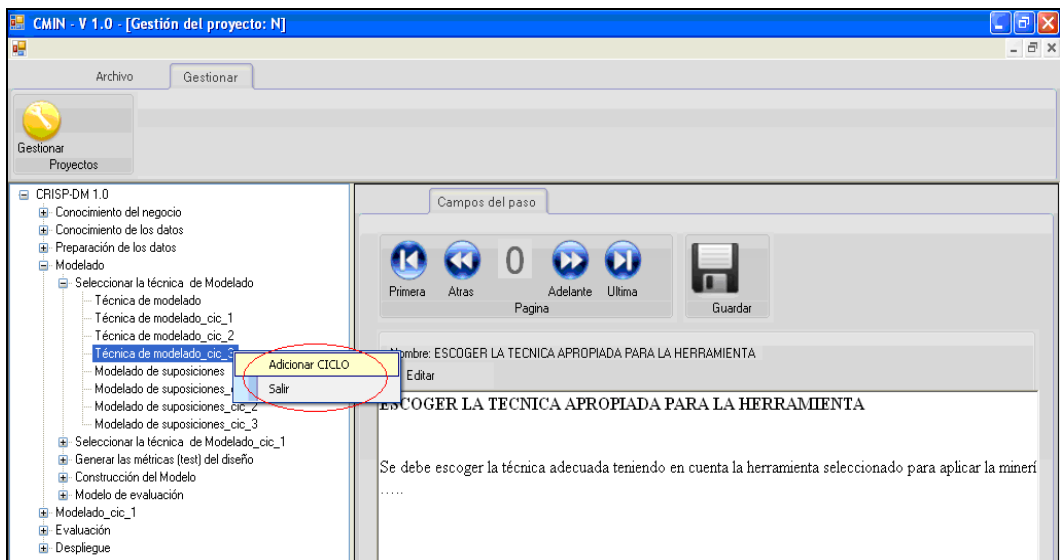


Figura 1. Salida Técnica de Modelado

2. Generar la Prueba del Diseño

Salida Prueba de Diseño

También consideramos que debería presentarse al usuario una plantilla para documentar esta salida de tal forma que guíe al usuario explícitamente. Esta plantilla podría contener lo presentado a continuación.

En cuanto a las pruebas para cada objetivo de la minería de datos:

Prueba de Diseño	Pasos para Realizar la Prueba

En cuanto a los datos requeridos para las pruebas:

Datos	Descripción
Datos de Entrenamiento	
Datos de Pruebas	
Datos de Validación	

3. Construir el Modelo

Salida Parámetros de Ajuste

En cuanto a la plantilla que se presenta para documentar los parámetros escogidos, su descripción y las razones de su escogencia, nos parece que existe una situación no deseada. Por ejemplo si voy a describir un parámetro y luego su descripción y ésta es más extensa que el tamaño de la celda de descripción la presentación no sería adecuada, como se puede observar en la figura 2.

Creemos que la tabla se debería poder llenar como se hace con una tabla creada en Microsoft Word.

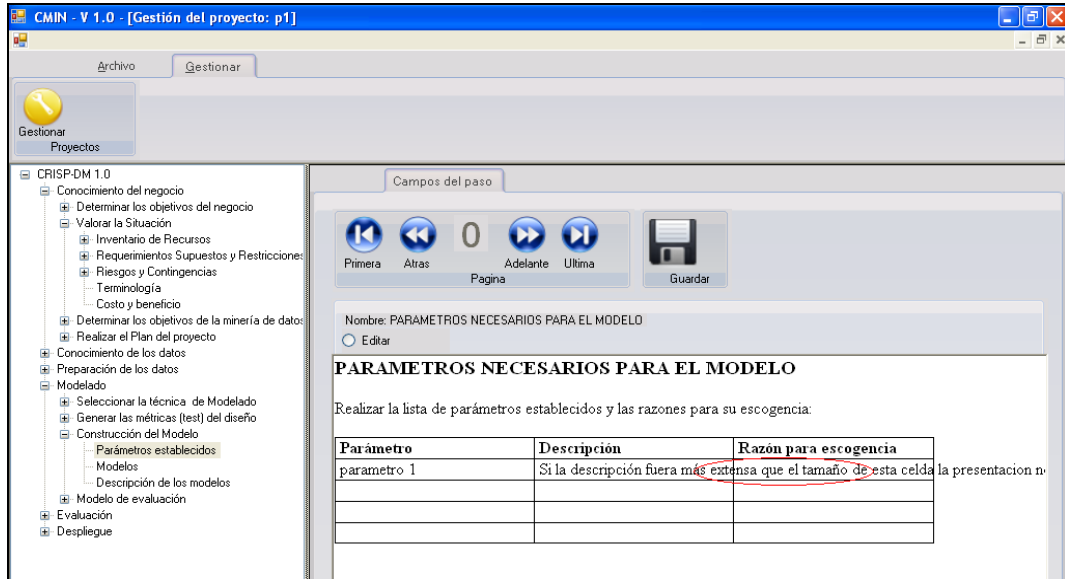


Figura 2. Parámetros establecidos

Modelos

Consideramos que la documentación guía de esta salida debería ser más clara al referirse a un proceso. Por ejemplo podría aparecer: "Registrar los procesos de minería de datos realizados para obtener los modelos:". La figura 3 muestra el texto al que nos referimos.

Además hacemos a esta salida (y a las otras salidas que presentan tablas que se pueden llenar) la misma observación que se hizo a la salida anterior en cuanto a la forma como se llenan los campos de la tabla que aparece en esta salida.

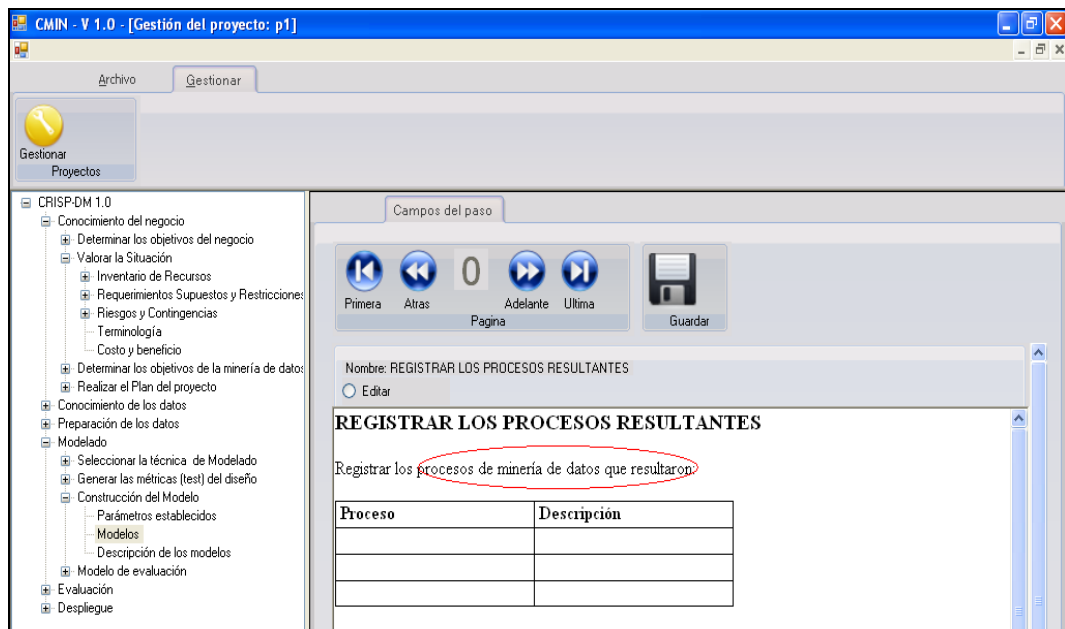


Figura 3. Salida Modelos

Descripción del Modelo

En cuanto a esta salida (“*Descripción de los Modelos*”) solo tenemos sugerencias para algunas de las páginas que se pueden observar. Por ejemplo en la página dos aparece una tabla cuya presentación no es la más adecuada por cuanto el título de una de las columnas está corrido y otro no se puede ubicar por completo dentro de su celda como se señala en la figura 4.

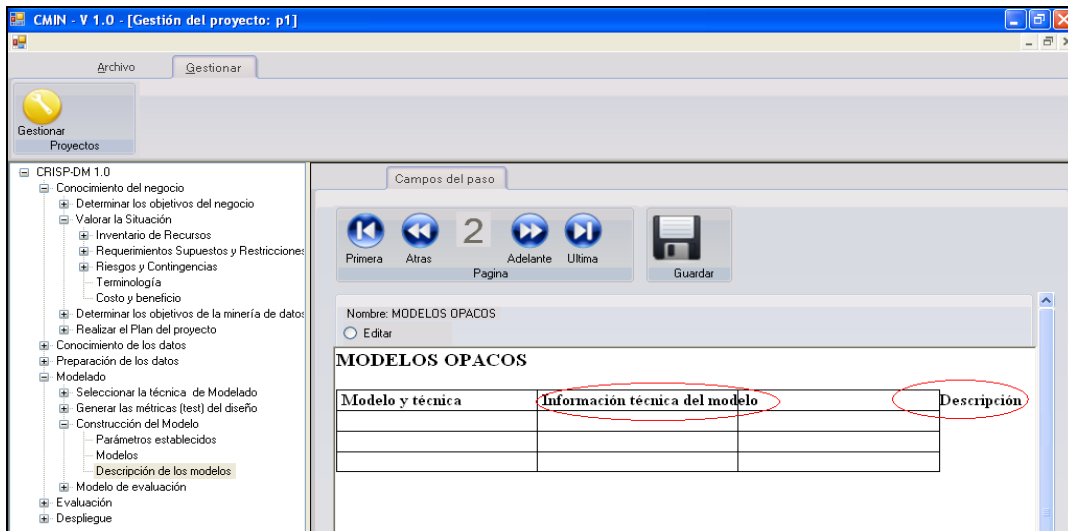


Figura 4. Salida Descripción de los Modelos-Página 2

En cuanto a la página tres sugerimos que debería aparecer una columna en la tabla que permita registrar las conclusiones asociadas a cada modelo, como se indica en la figura 5.

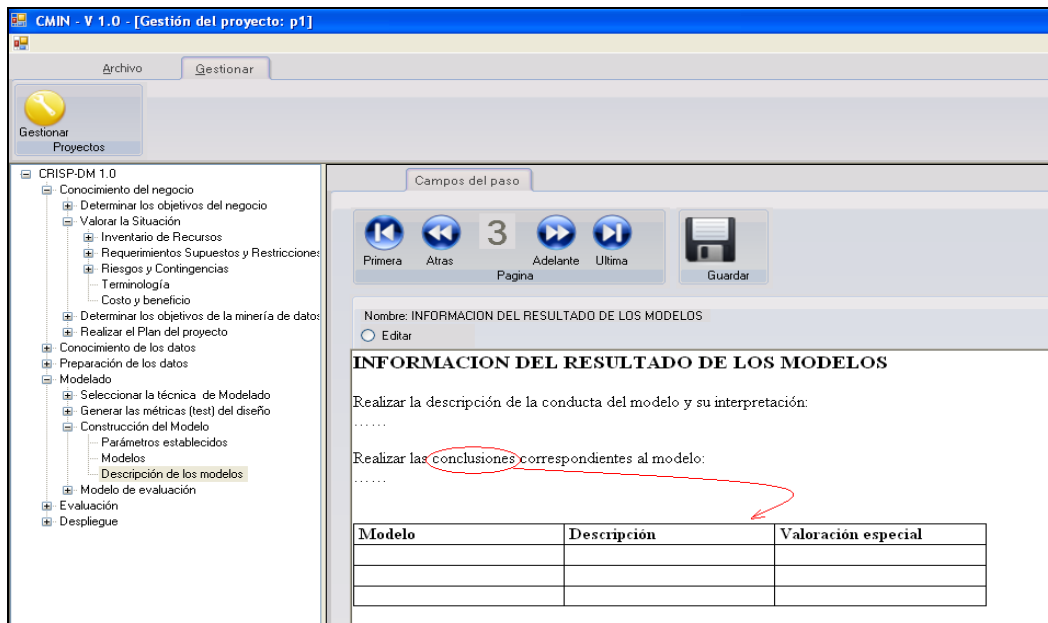


Figura 5. Salida Descripción de los Modelos-Página 3

Otra sugerencia que tenemos es en cuanto a la navegación entre página en esta salida. Consideramos que hace falta presentar no solo el número de página en que se encuentra el usuario sino también presentar el número total de las páginas, como se muestra en la figura 6. Además el número de páginas debería comenzar en uno (1) y no en cero (0) y también se debería habilitar y deshabilitar los botones de atrás, adelante, primera y última una vez ya no hayan más páginas hacia delante o hacia atrás respectivamente.

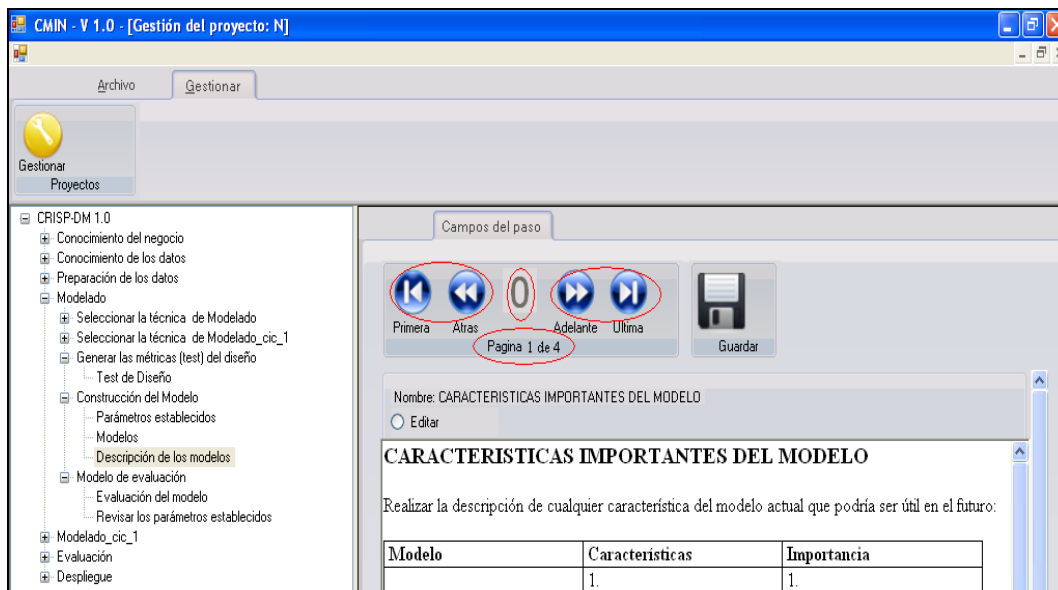


Figura 6. Salida Descripción del Modelo-Botones de Navegación

4. Valorar el Modelo

Valoración del Modelo

En cuanto a los resultados del modelo de esta salida creemos que debería presentarse una tabla para la comparación de los resultados de la evaluación y la interpretación. De igual forma se podría hacer para la clasificación jerárquica de los modelos, para la interpretación de resultados y conclusiones o recomendaciones especiales. La figura 7 muestra los elementos a los que nos referimos. La plantilla para mejor esta salida se podría presentar de la siguiente forma:

Para la comparación de resultados.

Modelo	Resultados de la Evaluación	Resultados de la	Conclusiones o Comentarios
--------	-----------------------------	------------------	----------------------------

		Interpretación	especiales

Para la jerarquía de modelos.

Modelos	Criterios de Éxito	Criterios de Evaluación	Clasificación

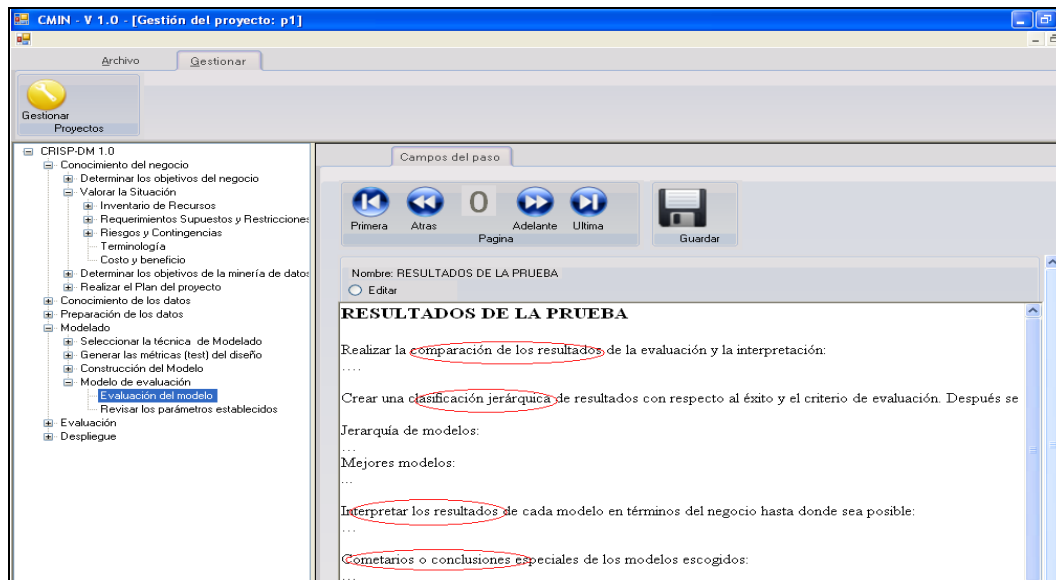


Figura 7. Salida Evaluación del Modelo

Parámetros de Ajuste revisados

En cuanto a esta salida solo tenemos que hacer una sugerencia para que se corrija un error de escritura en el texto presentado en la salida. La figura 8 indica esta situación.

En cuanto a lo demás creemos que la salida cumple con lo que se debería documentar para el ajuste de parámetros.

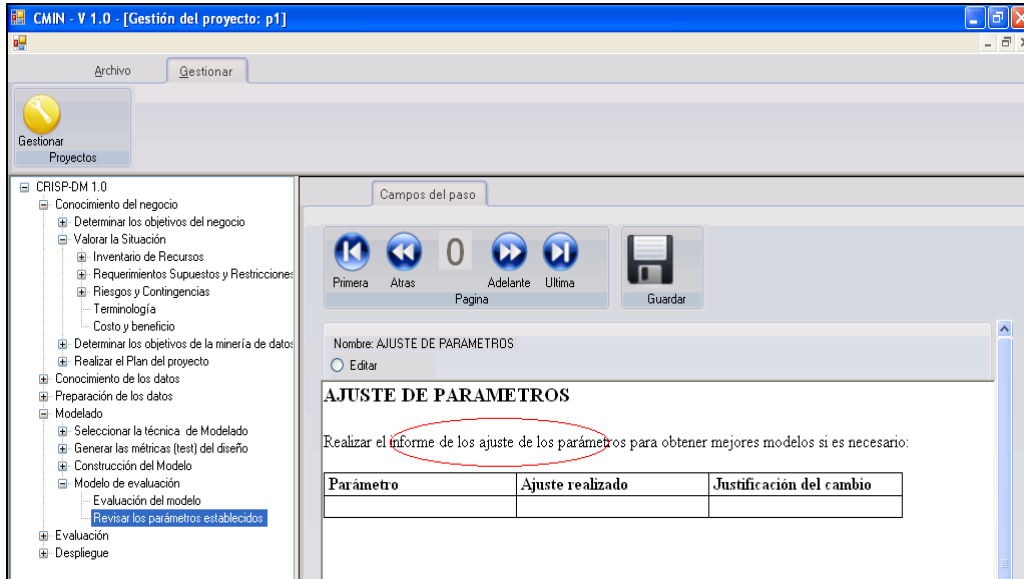


Figura 8. Salida Reajustar los parámetros establecidos

Sugerencia general

- El texto que se despliega al pasar por cada fase, tarea o salida de CRISP-DM, al ser algunas veces tan largo, no es posible leerlo completamente pues cada pocos segundos se desaparece el texto, y para observarlo de nuevo hay que volver a pasar sobre el texto que se quiere mirar perdiendo la continuidad en la lectura, situación que no es muy agradable para el usuario.

Creemos que el texto mostrado al deslizarse por cada paso es muy importante pero se debería de presentar de otra forma, por ejemplo podría haber otra pestaña al lado de "Campos del Paso" que me permitiera observar, al igual que esta pestaña, a medida que me ubico en un determinado paso, la descripción que corresponde a este en el lado derecho en el espacio que se señala en la figura 9.

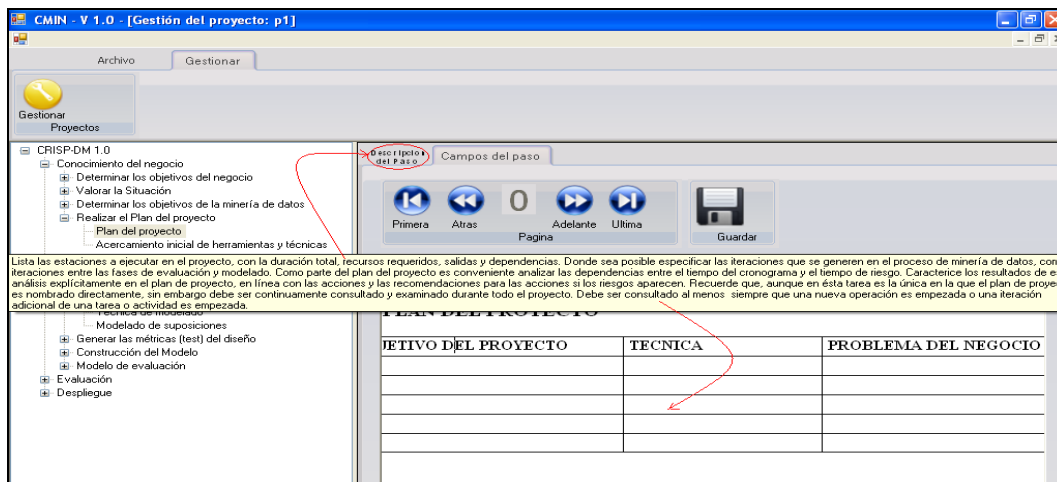


Figura 9. Pasos en CRISP-DM

10.5 FASE 5

TALLER EVALUACION CMIN
 FASE DE EVALUACION

DIEGO BENAVIDES 46012193
 ANDRES BENAVIDES 46012194

MODELO DE REFERENCIA CRISP-DM - EVALUACION

LA GUÍA DE USUARIO DE CRISP-DM

1. Evaluación de los resultados

Tarea: Evaluar los resultados

Salida: Evaluación de los resultados de minería de datos en lo que respecta a criterios de éxito de negocio

ACTIVIDAD	SI	NO
Comprender los resultados de la minería de datos		X
Interpretar los resultados en términos de la aplicación (del uso) <i>Observaciones: debería poder guardar comentarios de los resultados obtenidos, por que si el proyecto es muy largo, recordarlos en etapas posteriores seria muy difícil. Se podría guardar en un archivo de texto.</i>		X
Comprobar efectos sobre los objetivos de minería de datos <i>Observaciones: debería poder guardar comentarios de los resultados obtenidos, por que si el proyecto es muy largo, recordarlos en etapas posteriores seria muy difícil. Se podría guardar en una lista de efectos para cada objetivo.</i>		X
Comprobar los resultados de minería de datos contra la base de un conocimiento determinado para ver si la información descubierta es nueva y útil. <i>Observaciones: se debería poder documentar información que pueda llegar a ser útil, como por ejemplo si los resultados son o no coherentes, nuevos, etc.</i>		X
Evaluar y estimar los resultados en lo que respecta a criterios de éxito de negocio (esto es, el proyecto ha alcanzado los Objetivos de Negocio originales) <i>Observaciones: Debería poderse llevar un registro de los objetivos que se han alcanzado o no.</i>		X
Comparar los resultados de la evaluación y la interpretación. <i>Observaciones: Debería poder llenar una lista para comparar los resultados.</i>		X

Clasificar los resultados en lo que respecta a criterios de éxito de negocio	X	
Comprobar el efecto de los resultados sobre el objetivo (fin) de la aplicación inicial	X	
Determinar si hay nuevos objetivos de negocio para ser dirigidos más tarde en el proyecto, o en nuevos proyectos <i>Observaciones: Debería poder documentarse si hay nuevos objetivos, no solo conformarse con crear un nuevo ciclo.</i>		X
Expresar recomendaciones para proyectos futuros de minería de datos	X	

Salida: Modelos aprobados

TAREA	SI	NO
Seleccionar y aprobar los modelos generados que encontraron los criterios seleccionados. <i>Observación: En el treeview se encuentra "Modelos aprobados", pero al seleccionar esta opción, la herramienta no muestra nada</i>		X

2. Proceso de revisión

Tarea: Revisar el proceso

Salida: Revisión de procesos

ACTIVIDAD	SI	NO
Proporcionar una descripción del proceso de minería de datos usado	X	
Analizar el proceso de minería de datos. Para cada etapa del proceso pregunte: o ¿Esto fue necesario? o ¿Esto fue ejecutado óptimamente? o ¿En que modo podría ser mejorado?	X	
Identificar fracasos	X	
Identificar pasos desviados (de engaños)	X	
Identificar acciones alternativas posibles y/o caminos inesperados en el proceso	X	
Revisar resultados de minería de datos en lo que concierne a criterios de éxito de negocio <i>Observación: Podría tener una checklist con los criterios de éxito de negocio, para revisar los resultados.</i>		X

3. Determinación de los próximos pasos

Tarea: Determinar los próximos pasos

Salida: Lista de acciones posibles

ACTIVIDAD	SI	NO
Analizar el potencial para el desarrollo de cada resultado		X
Estimar el potencial para la mejora de proceso actual	X	

Comprobar los recursos restantes para determinar si ellos permiten iteraciones de proceso adicionales (o si recursos adicionales pueden estar siendo disponibles)	X	
Recomendar continuar con las alternativas		X
<i>Observación: se podría tener un espacio para poner una nota, recomendando que se pueda hacer en los próximos pasos.</i>		
Refinar el plan de proceso	X	

Salida: Decisión

ACTIVIDAD	SI	NO
Clasificar las acciones posibles	X	
Seleccionar una de las acciones posibles	X	
Documentar las razones para la elección	X	

10.6 FASE 6

EVALUACION DESPLIEGUE

**DEIRO ENRIQUE ZUÑIGA
WILLIAM RAMIRO JOAQUI R.**



Presentado a: MAG. CARLOS ALBERTO COBOS LOZADA

**UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERIA ELECTRONICA Y COMUNICACIONES**

POPAYAN 2008

DESPLIEGUE

En la preparación para despliegue

Al ejecutar el programa en la autenticación se debería dar la opción al cliente de seleccionar un proyecto ya guardado, mediante la opción abrir, o al darle aceptar preguntarle que desea hacer, brindándole las opciones de abrir o crear un nuevo proyecto.

Lo mismo a la hora que aparece la interfaz de archivo se podría colocar opción de abrir un proyecto ya existente.

En la gestión de proyectos al guardar dar la opción de guardarlo en una dirección diferente, si ya existiese el proyecto dar opciones y no dejar guardar.

Colocar una opción de ayuda en donde se especifique el Crisp-DM, además distribuirlo de la forma como se gestiona un proyecto.

1. En la PREPARACION DEL DESPLIEGUE se sugiere:

Cambiar preparación por Planeación ya que planear es un término más amplio a la hora de llevar a cabo la gestión de un proyecto.

No aporta nada para nosotros a la hora de de la planeación del despliegue lo siguiente:

Realizar el informe del resumen de los resultados a desplegar:

....

Desarrollar y evaluar planes alternativos para el despliegue:

....

Se sugiere agregar unas ventanas para cada una presentando en ellas Items para ser llenados por las personas encargadas de la realización del despliegue de los resultados, así como para cada actividad del despliegue.

En las tablas decidir como se realizará la propagación de los datos a cada una de las partes interesadas:

Los campos de las tablas se presentan por fuera de la tabla.

Sugerencia

La parte introductoria seria conveniente establecer como a modo de ayuda, solo dejar la tabla de diligenciar.

A la hora de manipular la tabla habilitar opciones de navegabilidad por teclado de TAB.

Colocar la tabla de manera que no se puedan modificar sus campos, solo el contenido que se llena por el usuario.

La tabla colocarla de manera que se autoajuste al texto y no el texto a la tabla, ya que a la hora de llenarla ocupa las otras columnas aledañas.

2. REQUERIMIENTOS PARA EL PLAN DE MANTENIMIENTO Y SUPERVISION

No aporta nada para nosotros en el plan de mantenimientos y supervisión:

Definir como será supervisada la exactitud:

....

Establecer cuando ya no se deben utilizar el resultado o el modelo de minería:

....

Especificar si los objetivos del negocio usados en el modelo pueden cambiar en un cierto plazo:

....

Realizar el plan de mantenimiento y supervisión:

....

Se sugiere agregar unas ventanas para cada una, presentando en ellas Items para ser llenados por las personas encargadas de la realización del despliegue de los resultados, así como para cada actividad del despliegue.

3. En el reporte final

No aporta nada para nosotros en el plan de mantenimientos y supervisión:

Establecer que tan bien se han conseguido los objetivos de minería de datos:

....

Definir la estructura del informe y su contenido:

....

Especificar qué resultados se incluirán en el informe:

....

Se sugiere agregar unas ventanas para cada una, presentando en ellas Items para ser llenados por las personas encargadas de la realización del despliegue de los resultados, así como para cada actividad del despliegue.

Las tablas presentan los mismos problemas de manejo.

4. INFORMACION PARA LA DOCUMENTACION DE LA EXPERIENCIA

Falta la evaluación del proyecto, en el que se debe hacer:

- Valorar lo que salió bien y lo que salió mal, lo que fue hecho bien y lo que tiene que ser mejorado.

Se sugiere agregar unas ventanas para cada una de las actividades que se presentan, colocando en ellas los Items para ser llenados por las personas encargadas de la realización del despliegue de los resultados, así como para cada actividad del despliegue que se nombra.

ANEXO E. DISEÑO DE LAS EVALUCIONES UTILIZADAS EN LA REALIZACION DE LA PRUEBA BETA.

Test de Conocimiento

TEST Experimento CMIN

1. Conoce usted metodologías o procesos para el desarrollo de proyectos de minería de datos.

SI ____ NO ____

Si respondió SI escriba cuales

2. Para desarrollar un Proyecto de minería de Datos cree usted necesario una etapa de "Entendimiento del negocio".

SI ____ NO ____

Si respondió SI ¿Por qué cree usted que es necesaria?

3. ¿Que tareas realizaría en una etapa "Entendimiento del negocio" para un proyecto de minería de Datos?

4. ¿Suponiendo las siguientes dos (2) etapas para un proyecto de minería de datos, relaciona cada tarea presentada con la etapa en la que se deba realizar?

ETAPAS

A	Entendimiento de Los Datos
----------	-----------------------------------

B	Preparación de Los Datos
----------	---------------------------------

TAREAS

	Reporte de Limpieza de Datos
	Reporte de Calidad de los Datos
	Explorar Datos
	Construcción de Datos

5. Suponiendo una etapa de Modelado para Realizar un Proyecto de minería de Datos, Marque las tareas que realizaría en esta etapa.

TAREAS

	Seleccionar la Técnica de Modelado
	Evaluar los resultados
	Evaluar el modelo

	Selección de Datos
	Generar un diseño de Test del modelo

6. De las Sigüientes Etapas a presentar cuales piensa usted que deberían ser las últimas 2 Etapas de un proyecto de minería de Datos.

ETAPAS

	Entendimiento del Negocio
	Evaluación
	Modelado
	Despliegue
	Preparación de los Datos

Test Usabilidad

TEST DE USABILIDAD

1. Estructura de la aplicación

- 1.1. Organización estructural: distribución de elementos estructurales de la aplicación (eje. Barras de desplazamiento, zonas de contenido, botones, etc.) es buena.

En desacuerdo ___ ___ ___ ___ **De acuerdo**

- 1.2. Densidad estructural: la cantidad de elementos estructurales que se utilizan en la aplicación es excesiva.

En desacuerdo ___ ___ ___ ___ **De acuerdo**

- 1.3. Consistencia de la estructura: la distribución de los elementos estructurales se mantiene constante a lo largo de la aplicación.

En desacuerdo ___ ___ ___ ___ **De acuerdo**

2. Operación de la aplicación

- 2.1. Navegabilidad: el recorrido que se hace por el contenido de la aplicación es fácil.

En desacuerdo ___ ___ ___ ___ **De acuerdo**

- 2.2. Interactividad: la relación mutua entre el usuario y la aplicación es buena.

En desacuerdo ___ ___ ___ ___ **De acuerdo**

- 2.3. Accesibilidad: las acciones que solicita la aplicación son fáciles de ejecutar.

En desacuerdo ___ ___ ___ ___ **De acuerdo**

- 2.4. Sistema de indicación: se identifican fácilmente las figuras, las tablas, las zonas activas y el tipo de acción que se debe ejecutar.

En desacuerdo ___ ___ ___ ___ **De acuerdo**

2.5. Desempeño del sistema: la velocidad de funcionamiento de la aplicación, considerando el tipo e tarea que se exige, es buena.

En desacuerdo **De acuerdo**

2.6. Fiabilidad del sistema: hay demasiados errores durante la operación de la aplicación.

En desacuerdo **De acuerdo**

2.7. Consistencia de la aplicación: la ejecución de tareas (eje. Navegar por la aplicación, hacer clic en botones, seleccionar opciones etc.) sigue un estándar a lo largo de la aplicación.

En desacuerdo **De acuerdo**

3. Información al usuario.

3.1. Sistema de ayuda: las dudas del usuario se resuelven fácilmente.

En desacuerdo **De acuerdo**

3.2. Feedback (realimentación): la aplicación mantiene al usuario informado sobre las tareas en ejecución.

En desacuerdo **De acuerdo**

3.3. Búsqueda de información: los datos que busca el usuario son fáciles de encontrar.

En desacuerdo **De acuerdo**

4. Apariencia: la presentación del contenido (eje. El tipo y tamaño de fuente, e uso de color, disposición de los elementos según su significado etc.) es buena.

En desacuerdo **De acuerdo**

5. Intuición: los procedimientos de navegaron por la aplicaron o ejecución de las tareas asignadas se aprenden de forma prácticamente inmediata.

En desacuerdo **De acuerdo**

6. Contenido

6.1. Organización de contenido: la distribución del contenido de la aplicación (eje. Textos, imágenes, test, etc.) es buena.

En desacuerdo **De acuerdo**

6.2. Densidad de contenido: la información que se presentan en la aplicación es demasiado extensa.

En desacuerdo **De acuerdo**

6.3. Fiabilidad del contenido: no hay errores en la información que se presenta en la aplicación.

En desacuerdo **De acuerdo**

6.4. Comprensión del contenido: la información que se presenta en la aplicación es fácil de entender y memorizar.

En desacuerdo **De acuerdo**

7. Experiencia del usuario
 Imprescindible

Prescindible

Opinión general sobre la aplicación.
___ **Excelente**

Mala

— — — —

ANEXO F. ARTÍCULO

CMIN Herramienta CASE para el soporte de proyectos de Minería de Datos basado en CRISP-DM

Carlos Cobos

Profesor Asociado

Universidad del Cauca, Departamento de Sistemas

ccobos@unicauca.edu.co

Jhon Zuñiga

Estudiante Ingeniería de Sistemas

Universidad del Cauca

jzunigaparedes@unicauca.edu.co

Juan Guarín

Estudiante Ingeniería de Sistemas

Universidad del Cauca

jguarin@unicauca.edu.co

RESUMEN

Existen metodologías para desarrollar Proyectos de Minería de Datos, la más destacada es CRISP-DM que contempla desde la comprensión del negocio hasta la fase de despliegue. Teniendo en cuenta que las Herramientas CASE permiten sistematizar metodologías en este trabajo se creó una herramienta CASE integrada que se basa en una metodología estándar, presenta una guía útil al usuario (ofreciendo ayuda y soporte en cada paso del proyecto) en cada paso del Proyecto y que cuenta con un WORK FLOW de minería de datos donde posibilita realizar estas labores y permite adicionar dinámicamente técnicas de minería de datos. Se hablara de su modelo conceptual, descripción de las funcionalidades más importantes y la forma en que permite adicionar algoritmos de minería de datos a su WORK FLOW. Finalmente las conclusiones y el trabajo futuro a realizar.

PALABRAS CLAVE

Minería de datos, metodología CRISP-DM, Herramientas CASE, WORK FLOW en minería de datos, Reflexión, Interfaces Software

ABSTRACT

Exist Methodologies for developing data mining project, the most prominent is CRISP-DM this provided from the understanding of the business through to deployment. Given that the CASE tools enable systematic methodologies in this paper believe that an integrated CASE tool is based on a standard methodology, presents a useful guide to the user (by providing help and support in every step of the project) in each step of the Project and which has a data mining WORK FLOW which enables to perform these tasks and allows to dynamically add data mining techniques. Telling of their architecture, the most important features and how they can add data mining algorithms to their WORK FLOW. Finally conclusions and future work to be done.

KEYWORDS

Data Mining, CRISP_DM Methodology, CASE tool, Data Mining Work Flow, Reflection, Interface

1 INTRODUCCIÓN

Para la explotación de la Minería de Datos existen diferentes herramientas, tema central de este trabajo porque se han determinado inconvenientes en ellas [1][2][3], entre los más relevantes: Acceso (costo), Desarrollo de interfaz (facilidad de uso), Metodología (cual siguen), Extensibilidad (en cuanto a algoritmos de minería de datos) y Equipos (Trabajo en equipos).

Del listado de las herramientas más representativas, según MetaGroup [1] y kdnuggest [2], se valoraron Clementine [4], Insightful Miner [5], WEKA [6], CART [7], PolyAnalyst [8] y SAS [9] (ver Tabla 7 Valoración de herramientas de minería de datos.). Para ver más detalles del estudio comparativo, refiérase a [1], [2] y muy especialmente a [3].

Herramienta	Acceso	Desarrollo Interfaz	Metodología	Extensibilidad	Equipos
Clementine	Costoso	Sencillo	Parte de CRISP-DM	No, sólo con el nuevo release	No
Insightful Miner	Costoso	Regular	PROPIA	No, sólo con el nuevo release	No
WEKA	Libre	Complejo	PROPIA	Si, por un usuario experto	No
CART	Costoso	Regular	PROPIA	No, sólo con el nuevo release	No
PolyAnalyst	Costoso	Regular	PROPIA	No, sólo con el nuevo release	No
SAS	Costoso	Sencillo	SEMMA	No, sólo con el nuevo release	No

Tabla 7 Valoración de herramientas de minería de datos.

Con esta valoración se planteo el desarrollo de una herramienta que se base en una metodología estándar para dar soporte a proyectos de minería de datos y que cuente con un WORK FLOW de minería de datos con funcionalidad extensible.

2 Herramientas CASE

En la medida que las computadoras se utilizaban en diversas disciplinas para facilitar el trabajo, la demanda de software ha aumentado dramáticamente. Para cubrir esta demanda, se crearon metodologías que pretendían establecer un estándar de desarrollo, además se creó un soporte automatizado para estas, el cual se denominó "ingeniería del software asistida por computador" (CASE) [10]. Las herramientas CASE ayudan a reducir el tiempo empleado en el desarrollo de un sistema, lo que mantiene el costo estable y contribuye en su calidad. [11]. Además, permiten al analista documentar y modelar un sistema, desde la definición de requerimientos hasta el diseño, implementación y prueba[11]. Diferentes investigaciones y usuarios sustentan que:[11]:

- Los usuarios de las CASE pueden mejorar la eficiencia, usando mecanismos de corrección de errores en el diseño de un sistema y usando herramientas para crear los correspondientes diagramas, evitando con esto volver atrás para realizarlos.
- Las organizaciones se motivan por la adopción de las CASE debido a que estas herramientas se basan en una metodología estándar para el desarrollo de sistemas, cuentan con el repositorio de datos central que describe el diseño del sistema reforzando el uso de la metodología soportada y ayudan a mejorar la calidad y seguridad durante el proceso de ingeniería de sistemas.

3 Metodologías y Procesos para el Desarrollo de minería de Datos

Existen metodologías estándar para el proceso de minería de datos que pretenden facilitar la realización de nuevos proyectos con características similares, optimizar la planificación y dirección de los mismos, reducir su complejidad y permitir realizar un mejor seguimiento a estos [12], entre las que se destacan CRISP-DM (Cross – Industry Standard Process for Data Mining) [13] y SEMMA (Simple, Explore, Modify, Model, Assess) [14]. SEMMA se centra más en las características técnicas del desarrollo del proceso, mientras que CRISP-DM, mantiene como foco los objetivos empresariales del proyecto. Esta diferencia se establece ya desde la primera fase donde SEMMA comienza realizando un muestreo de datos, mientras que CRISP-DM comienza realizando un análisis del problema empresarial para su transformación en un problema técnico. Desde ese punto de vista se puede considerar que CRISP-DM está más cercana al concepto real de un proyecto de minería de datos, pudiendo ser integrada con una Metodología de Gestión de proyectos específica que completaría las tareas administrativas y técnicas, además CRISP-DM es de libre distribución sin ningún costo a diferencia de SEMMA [14].

3.1 CRISP-DM: Cross-Industry Standard Process for Data Mining

La metodología CRISP-DM comprende una jerarquía de cuatro niveles. En el primer nivel se encuentran las fases que se componen de diferentes tareas genéricas las cuales son del segundo nivel, estas están diseñadas de tal forma que cubren todas las posibles situaciones que se presentan en el proceso de minería de datos. En el tercer nivel están las tareas específicas que soportan los diferentes escenarios que puedan presentar las tareas genéricas, y en el cuarto nivel se encuentra la instancia del proceso, que describe las actividades específicas a realizar en un proyecto de minería de datos (ver Figura 1 Jerarquía de Fases y Tareas en CRISP-DM (Adaptado de [13])).

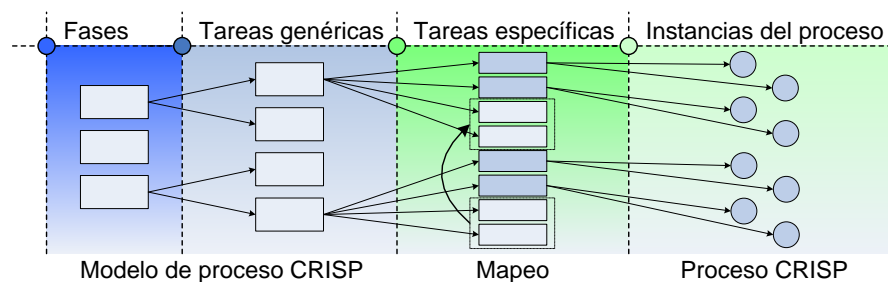


Figura 1 Jerarquía de Fases y Tareas en CRISP-DM (Adaptado de [13])

El proceso CRISP-DM define una estructura para proyectos de minería de datos y suministra la orientación para su ejecución, consta de un modelo de referencia y una guía de usuario.

Modelo de referencia [13]: Este documento da una visión general del ciclo de vida de un proyecto de minería de datos, contiene las fases con sus objetivos, las tareas, las relaciones entre estas y las instrucciones paso a paso que se deben llevar a cabo. Las fases definidas por el modelo de referencia son (ver Figura 2):

Guía de usuario [13]: La guía de usuario ofrece consejos más detallados, pistas por cada fase, y cada operación dentro de una fase, y ejemplifica cómo hacer un proyecto de minería de datos.

Esta guía de usuario es una excelente opción para desarrolladores de proyectos de minería de datos con poca experiencia y/o novatos.

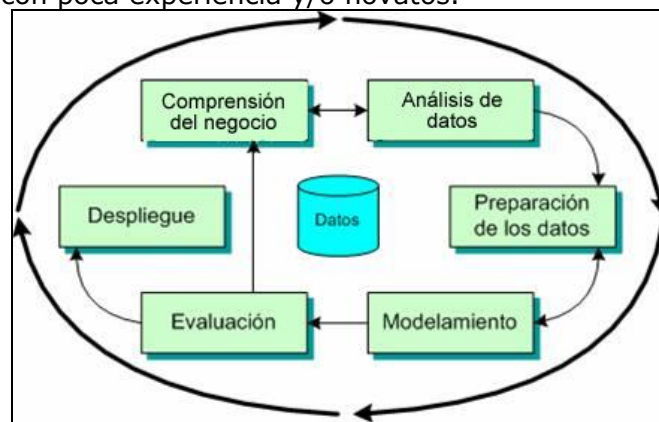


Figura 2 Fases del modelo de referencia CRISP-DM (Adaptado de [13])

4 MODELO CONCEPTUAL CMIN

Para comprender mejor el funcionamiento de CMIN, a continuación se presenta el modelo conceptual en el cual se representan los conceptos que hacen parte del sistema, así como las relaciones existentes entre estos, ver en la **Modelo Conceptual de CMIN 1.0**

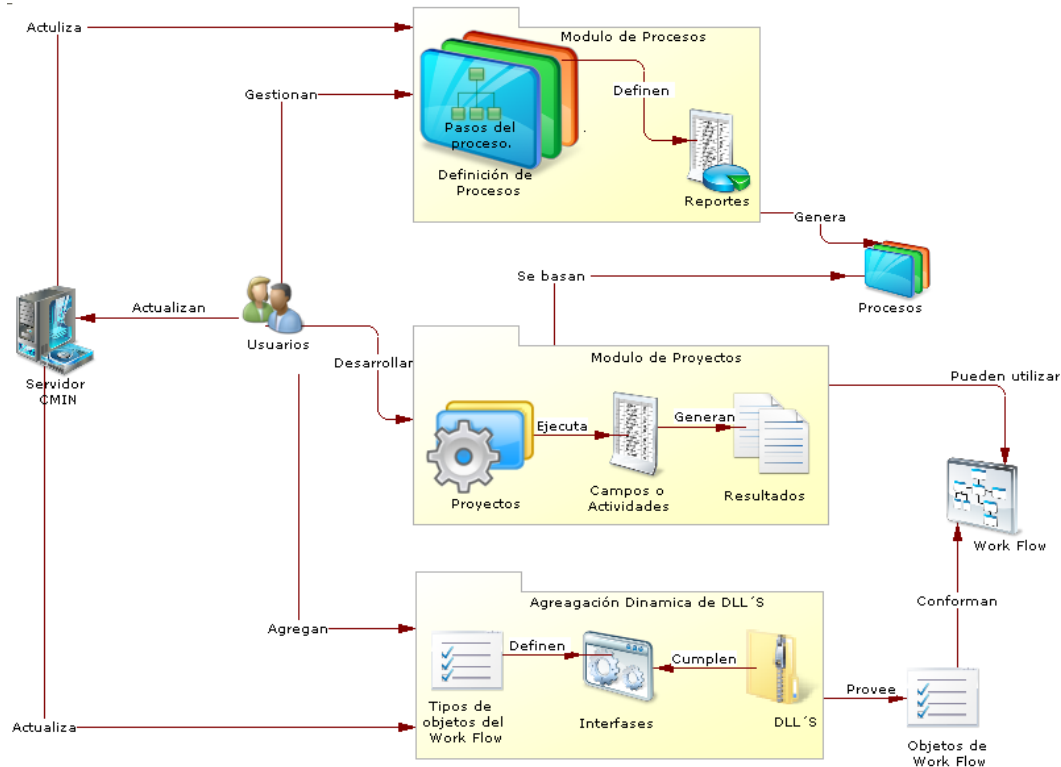


Figura 3 Modelo Conceptual de CMIN 1.0

A continuación se explica en forma breve aquellos conceptos y relaciones del modelo de CMIN que permiten percibir, identificar y describir su funcionamiento.

- **Usuarios:** Representa a las personas que pueden utilizar el sistema, los cuales pueden ser principiantes o expertos en minería de datos.
- **Modulo de Procesos:** Representa el modulo que permite la agregación de procesos o metodologías como CRISP-DM, definiendo la estructura, pasos y entregables propuestos por cada una. La **Definición de procesos** representa la acción de construir un proceso o metodología mediante la agregación y definición de sus pasos, campos o actividades que propone para el desarrollo de un proyecto de minería de datos. Los **Reportes** son los documentos o entregables que se deben arrojar durante un proyecto, los cuales servirán como soporte durante la ejecución del mismo.
- **Procesos:** Representa los procesos o metodologías que se han agregado a CMIN 1.0 y que servirán como base para los proyectos de minería a desarrollar.
- **Modulo de Proyectos:** Representa el modulo que permite la ejecución de un proyecto de minería de datos basado en uno de los procesos o metodologías agregado en el modulo de procesos. El concepto **Proyectos** representa el conjunto de proyectos que se han creado en CMIN 1.0 y están en curso o terminados. El concepto **Campos o Actividades** representan los campos que pertenecen a un paso, los cuales especifican las actividades que se deben realizar para cumplir con el fin del paso al que pertenecen. Los **Resultados** representan el resultado de la realización de una actividad, el cual puede ser: una sugerencia, un texto explicativo o una plantilla que se debe diligenciar.
- **Agregación dinámica de Librerías de Enlace Dinámico (DLL's):** Representa el modulo que permite la agregación de objetos que servirán para la ejecución

del Work Flow, por medio de DLL's con su funcionalidad. El concepto **Tipos de objetos del Fulo de Trabajo (Work Flow, WF)** representa el conjunto de tipos de objetos reconocidos por CMIN 1.0 para ser agregados y posteriormente ser utilizados por el WF ofrecido por CMIN 1.0. El concepto **Interfases** representa el conjunto de interfases que deben cumplir las DLL's para poder ser agregadas al conjunto de objetos que serán utilizados por el WF. Las **DLL's**: representan el conjunto de DLL's que poseen la implementación de los objetos del WF.

- **Objetos de WF**: Representa el conjunto de objetos que se han agregado a CMIN 1.0 y podrán ser utilizados por el WF, el cual puede crecer a medida que los usuarios realicen implementaciones nuevas de cualquiera de los tipos de objetos del WF especificados en CMIN 1.0.
- **Work Flow**: Representa la herramienta que ofrece CMIN 1.0 para que los usuarios realicen actividades propias de minería, y puedan crear modelos utilizando la tareas de minería propuestas por CMIN 1.0.
- **Servidor CMIN**: Representa el servidor que aloja nuevas definiciones de metodologías, así como nuevas implementaciones de objetos del WF por medio de DLL's, para que los usuarios actualicen CMIN 1.0 si así lo requieren ya que CMIN 1.0 se ejecuta independientemente de este servidor.

Con la combinación de cada uno de estos conceptos y sus relaciones, el presente modelo conceptual formula como es el flujo normal del funcionamiento de la herramienta CMIN 1.0 para el soporte en el desarrollo de proyectos de minería de datos.

5 CMIN: Sistematización de CRISP-DM

El principal objetivo de este trabajo fue Desarrollar una herramienta CASE integrada para soportar y orientar el desarrollo de proyectos de minería de datos, Basada en una Metodología estándar como lo es CRISP-DM. Este objetivo origino la Creación del Modulo de Gestión de Procesos, que permite Editar los Procesos del sistema, y de cada Proceso sus Pasos y de cada Paso sus Actividades o Campos. Este Modulo resuelve el Problema de Actualización de futuras versiones de CRISP-DM. En la actualidad CMIN tiene creado el Proceso CRISP-DM V1.0 el cual contiene todos los pasos y actividades planteadas en la metodología CRISP-DM versión 1.0. Este modulo posibilita la adición de nuevos procesos para cada versión de CRISP-DM.

En la Figura 4 se presenta como el Modulo Permite Editar los Procesos del Sistema y como ejemplo aparece el Ingreso del Proceso CRISP-DM V1.0 (botón **Guardar**), a continuación se presenta la edición de los Pasos del Proceso (botón **Pasos**).

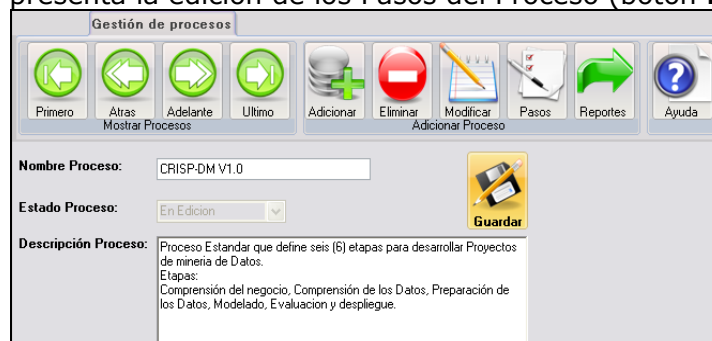


Figura 4 Edición de Procesos (Información Base)

La Figura 5 presenta como este modulo permite crear la estructura de la metodología o Proceso a editar, adicionando cada pasó propuesto en ella especificándole el tipo (Fase, Tarea Genérica, Tarea específica, etc.) de paso para cumplir con la jerarquía propuesta en la misma y una descripción que especifique que se debe hacer o lograr en ese pasó. Cada paso puede contener pasos internos (hijos) lo que lo hace posible. Además se aprecia el resultado de la Edición de los pasos del Proceso, presentando como ejemplo a CRISP-DM 1.0 incluido en CMIN.

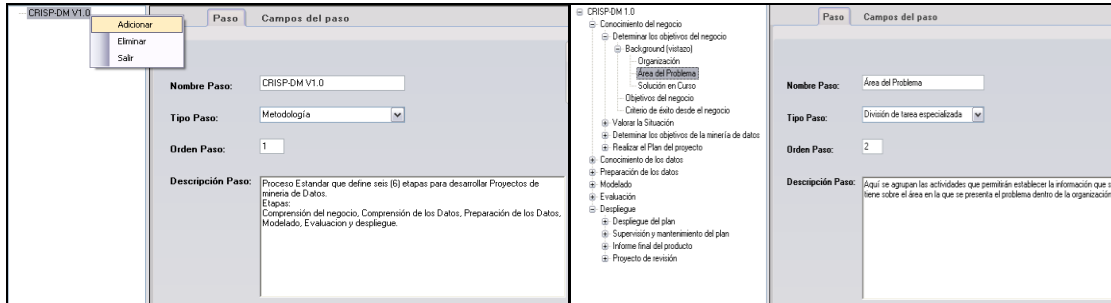


Figura 5 Edición de Pasos a partir del Paso Padre y CRISP-DM en CMIN como resultado.

Los Pasos contienen Actividades o Campos los cuales se editan para cada Paso en la pestaña **Campos del Paso**. En la **Edición de Campos del Paso**, se aprecia que cuando se adiciona un Campo o actividad a un Paso (botón **Adicionar**) entre la información requerida para el Campo o actividad se destaca la **Descripción** que contiene que se debe hacer si es una actividad o la sugerencia que presente la metodología en este paso, el **Tipo de Campo** que define si el Campo es una Actividad o sugerencia y **Utiliza Work Flow** que indica si para realizar la actividad o campo es necesario utilizar el WorkFlow de CMIN.

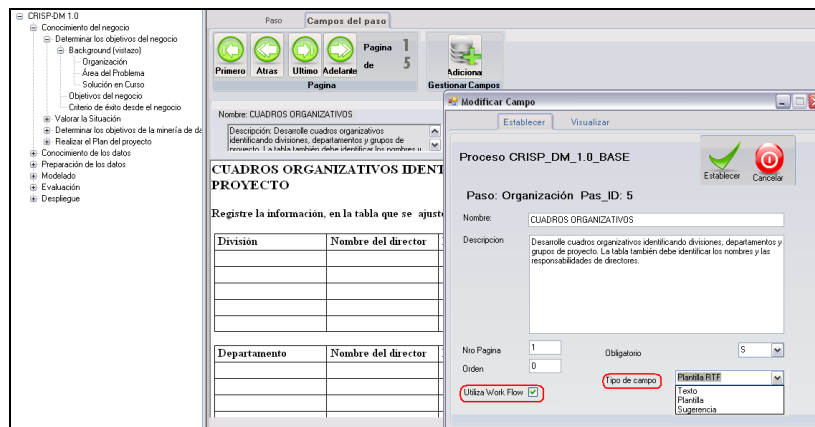


Figura 6 Edición de Campos del Paso.

Desarrollo de proyectos basados en metodologías

CMIN 1.0 permite desarrollar un proyecto de minería de datos basado en una metodología, para estos se hace que los proyectos heredaran toda la estructura de una metodología ingresada por el modulo de procesos anteriormente explicado, de tal manera que muchos proyectos se pudieran basar de una de las metodologías ingresadas.

En la se presenta como este Modulo Permite crear los Proyectos y como ejemplo aparece la creación del Proyecto "Proyecto de Análisis" (botón **Guardar**), a

continuación se puede comenzar el desarrollo del proyecto (botón **Pasos**) para el cual se debe cumplir toda la estructura del proceso BASE.

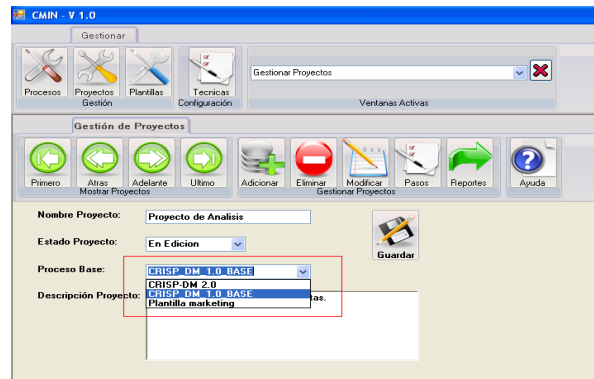


Figura 7 Modulo de gestión de proyectos de la Herramienta CMIN 1.0.

En la **Figura 8** se presenta el desarrollo de un proyecto, en el numeral **(1)** se puede observar la estructura de la metodología la cual es usuario la recorre a medida que esta desarrollando un proyecto de minería en CMIN, en el numeral **(2)** se observa la sección donde aparecen los campos o actividades a desarrollar pertenecientes al paso en el cual se encuentren, el numeral **(3)** se encuentra el botón que guarda la información resultante del campo o actividad, el numeral **(4)** se observa como se puede crear un ciclo en un proyecto característica que CMIN ofrece para hacer iteraciones de cualquier Paso, en el numeral **(5)** vemos como se visualizan los ciclos creados.

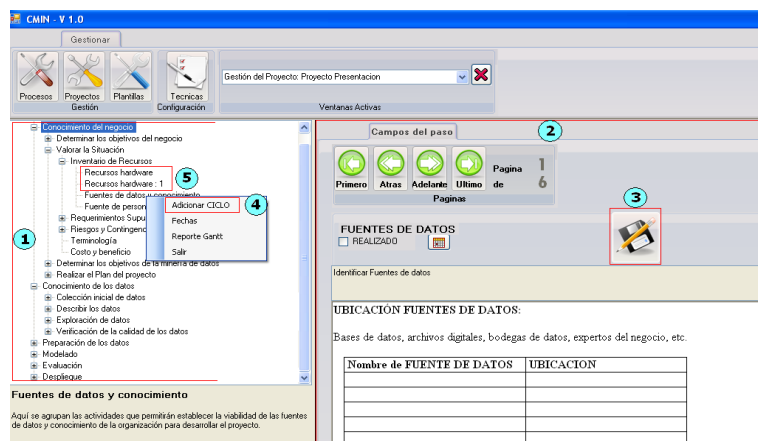


Figura 8 Desarrollo de proyectos basados en una metodología o proceso en la Herramienta CMIN

6 CMIN: Work Flow de minería de datos

La Herramienta CASE CMIN tiene varias características especiales, entre las que se encuentra la extensibilidad del conjunto de algoritmos que ofrece en su WORK FLOW de minería de datos. Este WORK FLOW es utilizado por los usuarios para realizar las labores de minería de datos en las Actividades de los Proyectos que así lo requieran. El Modulo de WORK FLOW en CMIN soporta el proceso de adición de algoritmos y la utilización de los mismos en el WORK FLOW.

Modulo de WORK FLOW

En la **Figura 9** WORK FLOW de minería en CMIN el numero (1) señala los **tipos de objetos** del WORK FLOW, el numero (2) un **objeto ofrecido** del tipo de objeto “Fuente de Datos” y el numero (3) el **objeto en ejecución** del WORK FLOW el cual es configurado y utilizado por el usuario ver WORK FLOW de minería en CMIN.

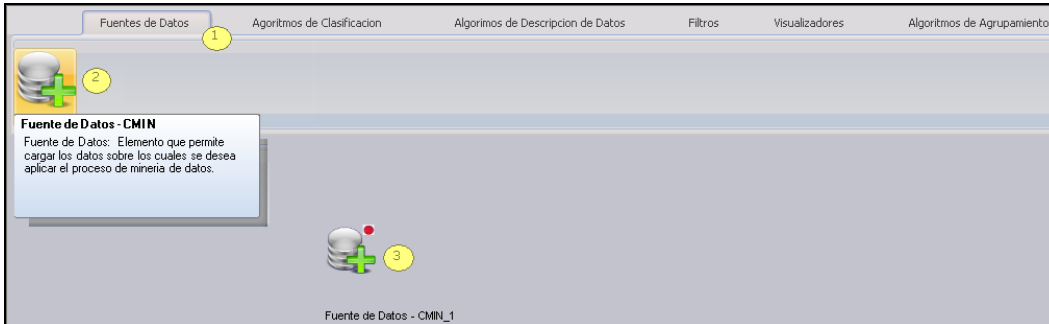


Figura 9 WORK FLOW de minería en CMIN

Tipos de Objetos del WORK FLOW y relaciones entre ellos: Los tipos de objetos en CMIN están definidos en dos grupos los basados en los tipos de Problemas de minería de datos y los basados en las herramientas necesarias para las etapas de comprensión y procesamiento de datos. Obteniendo el siguiente listado:

Tipos de Problemas de minería de datos	Tipos de Fases Comprensión y Procesamiento de Datos
<ul style="list-style-type: none"> • Algoritmos de Clasificación • Algoritmos de Descripción de Datos • Algoritmos de Regresión Lineal • Algoritmos de Agrupamiento 	<ul style="list-style-type: none"> • Fuentes de Datos • Filtros • Visualizadores

Tabla 8 Listado de Tipos de Objetos

Con el objetivo de permitir adicionar algoritmos u objetos a los tipos de objetos en tiempo de ejecución, para cada tipo de objeto del WORK FLOW se le definió una Interfaz de Software [15], que agrupe los métodos necesarios para su aplicación y otros métodos de Interacción con los demás Tipos del WORK FLOW. En tiempo de desarrollo para registrar (ver **Figura 10**) un tipo de objeto del WORK FLOW en CMIN es necesario que la Interfaz sea convertida a un Assembly mediante un Proyecto de Librería de Visual Studio [16].

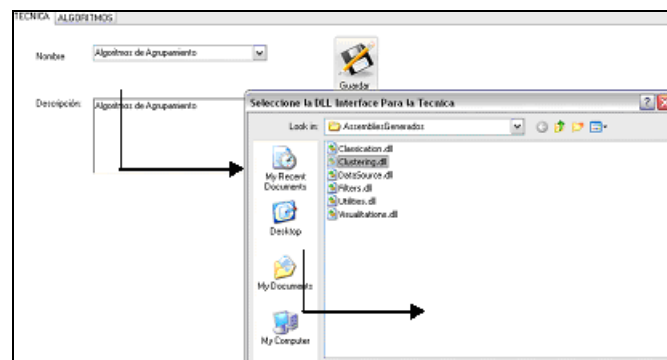


Figura 10 Edición de Tipos de objetos del WORK FLOW

La información del tipo de objeto es almacenado en la Base de Datos y el archivo “.dll” es copiado y almacenado en la carpeta local de CMIN **Assemblies_CMIN**. Después de Ingresar el tipo de objeto se le debe definir con quien puede establecer enlaces, es decir definir qué tipo de objeto puede entregarle información y a que tipo de objeto le puede entregar información (ver **Figura 11**).

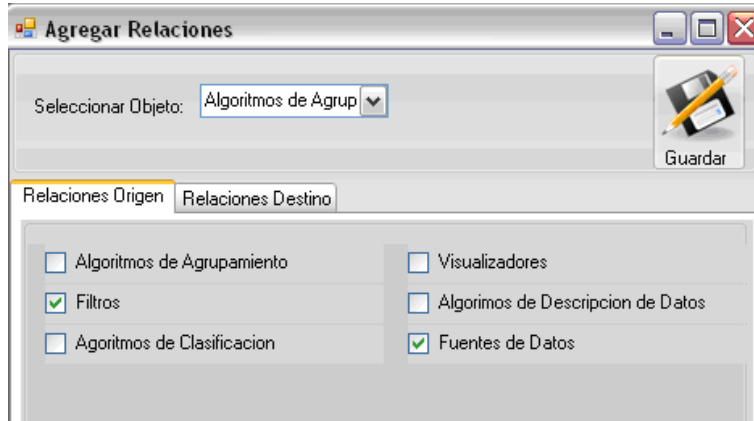


Figura 11 Edición de Relaciones de Tipos de Objetos del Work Flow de CMIN

Proceso de adición de un algoritmo a un Tipo de Objeto del WORK FLOW de CMIN: El proceso para adicionar un algoritmo o nuevo **objeto ofrecido** a un **tipo de objeto** de CMIN es el siguiente:

- El usuario Programador crea un Proyecto de Librería en Visual Studio[16] en cual debe adicionar como referencia la dll que se generó a partir de la interfaz de Software[15] creada para el Tipo de objeto, para el cual se va a crear el Nuevo objeto ver **Figura 12**

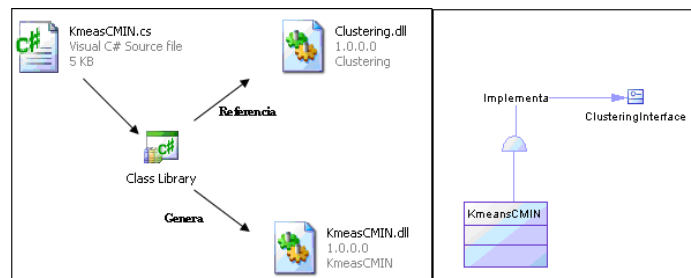


Figura 12 Generar Assembly y diagrama de Clases dentro del Proyecto de librería

- El usuario programa su algoritmo en el Proyecto de Librería, implementando la Interfaz definida por CMIN para el tipo de objeto al cual va a ser adicionado, generando una **nueva dll**, que contiene la funcionalidad del nuevo Algoritmo para CMIN. Las dll resultado de este proyecto son comprimidas en un .zip. (ver **Figura 13**)

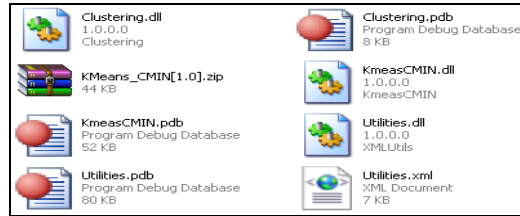


Figura 13 dll resultado del Proyecto de librería

- El usuario que desee agregar este nuevo algoritmo a CMIN utiliza este modulo, el cual solicita el .zip que contiene la dll del algoritmo ver **Figura 14**

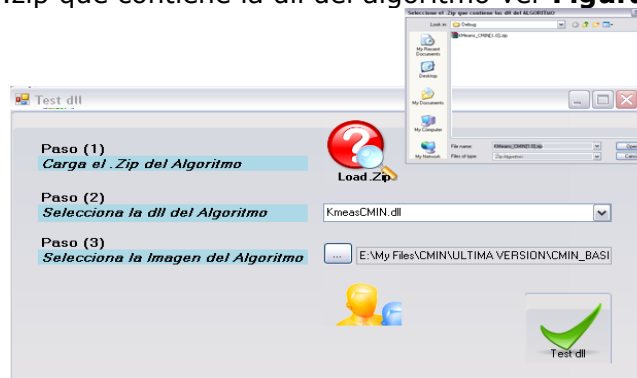


Figura 14 Adición de algoritmo a un tipo de objeto en CMIN.

- El modulo de WORK FLOW valida que ese nuevo algoritmo cumpla con la interfaz Software definida para el Tipo de objeto al que se dese ingresar. Utilizando la dll del tipo de objeto que tiene almacenada en la carpeta local de CMIN **Assemblies_CMIN** y la dll que viene en el .zip especificado, esta comparación se realiza utilizando reflexión (System.Reflection) [17] cargando los Assemblies y comparando los métodos. (ver **Figura 15**)

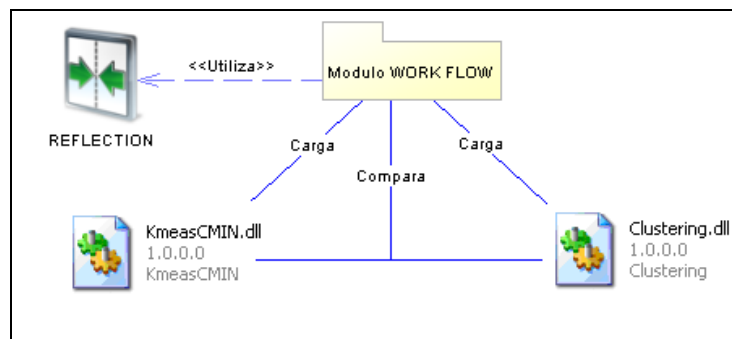


Figura 15 Proceso de validación de algoritmo en CMIN

- Si el nuevo algoritmo cumple con la Interfaz Software del Tipo de Objeto se registra en la Base de datos y los archivos del .zip son descomprimidos y almacenados en la carpeta local de CMIN **ALGORITMOS**, quedando listo para ser utilizado en el WORK FLOW de CMIN (ver **Figura 16**)

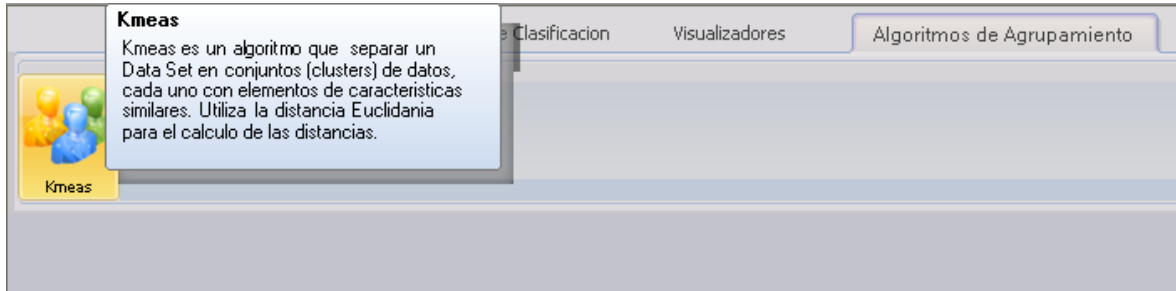


Figura 16 Algoritmo Nuevo listo para su utilización

Invocación de Métodos de Algoritmos en CMIN

Para la invocación de los métodos de las dll de los algoritmos, se debe tener en cuenta que CMIN almacena los Assembly o dll de los algoritmos en carpetas locales y que tiene también almacenados los Asemmbles de los tipos de Objetos. Estos tipos de objetos del WORK FLOW son estáticos y la parte dinámica son algoritmos u objetos de cada uno, los cuales pueden crecer en tiempo de ejecución. Con este precedente y destacando que nosotros como desarrolladores de CMIN creamos las interfaces de Software de cada tipo de Objeto, teniendo en cuenta métodos que permitieren la interacción de los algoritmos con el usuario y el Núcleo CMIN. Desarrollamos la programación del WORK FLOW basados en la información de las definiciones planteadas en las Interfaces Software (es decir programamos para los tipos de objetos) para la invocación de los métodos de los algoritmos de cada tipo de objeto, utilizando para esto REFLECTION (System.Reflection) [17] para cargar los Assembly o dll de cada algoritmo, y realizar la invocación de los métodos donde fuere necesario. Programando además la interacción de los objetos basados en las reglas o relaciones definidas en la **Figura 11**

CONCLUSIONES Y TRABAJO FUTURO

Conclusiones

CMIN es una Herramienta CASE integrada que orienta el desarrollo de los proyectos a través de Procesos, facilita la integración del Proceso o Metodología propuesta con el Proyecto y asegura el cumplimiento del Proceso en la ejecución del Proyecto.

CMIN una herramienta con funcionalidad extensible, que facilita el desarrollo en comunidad, con la razón de que cada nueva funcionalidad es programada por aparte (por los miembros de la comunidad), después es probada y evaluada por un grupo de expertos y finalmente es incluida y expandida a los demás miembros de la comunidad.

Mediante la información, detallada y apropiada del Proceso guía, que presenta CMIN en cada Paso del proyecto y la realización de actividades en esta CASE, posibilita que el usuario apropie el Proceso guía identificando la necesidad de cada Paso.

Trabajo Futuro

Implementar un componente de seguimiento a Proyectos basado en lo ya hecho que tenga en cuenta la administración de los recursos para cada actividad, de tal forma

que se puedan hacer reportes de costos en cada paso del Proyecto y tener integrado en CMIN una metodología de Gestión de Proyectos en conjunto con los Procesos de Minería de datos.

Crear una comunidad que permita un rápido crecimiento de los algoritmos que ofrece CMIN en la actualidad, para potenciar el uso del WORK FLOW.

REFERENCIAS

- [1] MetaGroup (METAspectrum Market Summary), (visitado 15 Junio de 2007), http://www.oracle.com/technology/products/bi/odm/pdf/odm_metaspectrum_1004.pdf
- [2] Kdnuggets (tools data mining), (visitado 2007, 15 febrero) http://www.kdnuggets.com/polls/2005/data_mining_tools.htm
- [3] Britos, P., Fernández, E., Ochoa, M., Merlino, H., Diez, E., y García, R. METODOLOGÍA DE SELECCIÓN DE HERRAMIENTAS DE EXPLOTACION DE DATOS. II Workshop de Ingeniería del Software y Bases de Datos. XI Congreso Argentino de Ciencias de la Computación (2005). Pág. 113-123.
- [4] SPSS Inc. (Clementine), (Visitado 2007, 14 Febrero). [Pagina de Información]. URL <http://www.spss.com/es/clementine/>
- [5] Insightful Corporation (Insightful Miner), (Visitado 2007, 14 Febrero). [Pagina de Información]. URL <http://www.insightful.com/products/iminer/default.asp>
- [6] University of Waikato, (Visitado 2007, Febrero 15). [Pagina de principal]. URL <http://www.cs.waikato.ac.nz/ml/weka/>
- [7] Salford System (CART), (Visitado 2007, 21 Febrero). [Pagina Información]. URL <http://www.salfordsystems.com/cart.php>
- [8] PolyAnalyst (tool data mining), (visitado 2007, 10 febrero), <http://www.megaputer.com>
- [9] SAS(Enterprise miner), (visitado 2007, 20 mayo) <http://www.sas.com/technologies/analytics/datamining/miner/>
- [10] INSTITUTO NACIONAL DE ESTADISTICA E INFORMATICA (Perú), Elaborado por la Sub-Jefatura de Informática. (Visitado 2006, 28 Noviembre). Herramientas CASE [Documento Informativo]. URL <http://www.inei.gob.pe/biblioineipub/bancopub/Inf/Lib5103/Libro.pdf>
- [11] Albizuri-Romero, M. B. 2000. A retrospective view of CASE tools adoption. *SIGSOFT Softw. Eng. Notes* 25, 2 (Mar. 2000), 46-50. DOI=<http://doi.acm.org/10.1145/346057.346071>
- [12] Gondar Nores, José Emilio. (Visitado 2007, 19 Marzo). Metodologías para la Realización de Proyectos de Data Mining. [Articulo en Inrernet]. URL <http://www.estadistico.com/arts.html?20040426>
- [13] Project CRISP-DM (CRISP-DM). (Visitado 2006, 01 Noviembre). [Pagina Principal]. URL <http://www.crisp-dm.org/>
- [14] SAS (SEMMA). (Visitado 2006, 10 Noviembre). [Pagina de Información]. URL <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>
- [15] Interface Statement (Visual Basic), Visual Basic Language Reference (visitado 2008, Noviembre) URL <http://msdn.microsoft.com/en-us/library/h9xt0sdd.aspx>
- [16] Mahech, Chand. 2000, Creating C# Class Library (DLL) Using Visual Studio .NET (visitado 2008, Marzo) URL <http://www.c-sharpcorner.com/UploadFile/mahesh/dll12222005064058AM/dll.aspx>
- [17] Reflection Overview, .Net FrameWork Developer's Guide (visitado 2008, Noviembre) URL <http://msdn.microsoft.com/en-us/library/f7ykdhsy.aspx>

