

# Índice general

<b>1. INTRODUCCIÓN</b>	<b>5</b>
<b>2. TEMA</b>	<b>7</b>
2.1. Título . . . . .	7
2.2. Línea de Investigación . . . . .	7
2.3. Alcance y Delimitación . . . . .	7
<b>3. PROBLEMA OBJETO DE ESTUDIO</b>	<b>9</b>
3.1. Descripción del Problema . . . . .	9
3.2. Formulación del Problema . . . . .	10
<b>4. OBJETIVOS</b>	<b>11</b>
4.1. Objetivo General . . . . .	11
4.2. Objetivos Específicos . . . . .	11
<b>5. JUSTIFICACIÓN</b>	<b>13</b>
<b>6. MARCO TEORICO</b>	<b>14</b>
6.1. El Proceso de Descubrimiento de Conocimiento en Bases de Datos - DCBD . . . . .	14
6.2. Arquitecturas de Integración de las Herramientas DCBD con un SGBD . . . . .	15
6.3. Implementación de Herramientas DCBD débilmente acopladas con un SGBD . . . . .	16
6.4. Estado del Arte . . . . .	17
6.4.1. WEKA - Waikato Environment for Knowledge Analysis . . .	17
6.4.2. ADaM - Algorithm Development and Mining System . . . .	18
6.4.3. Orange - Data Mining Fruitful and Fun . . . . .	19
6.4.4. TANAGRA - A Free Software for Research and Academic Purposes . . . . .	21

6.4.5.	AlphaMiner . . . . .	23
6.4.6.	YALE - Yet Another Learning Environment . . . . .	25
<b>7.</b>	<b>DESARROLLO DEL PROYECTO</b>	<b>28</b>
7.1.	Análisis UML . . . . .	28
7.1.1.	Funciones . . . . .	28
7.1.2.	Diagramas de Casos de Uso . . . . .	31
7.1.3.	Diagramas de Secuencia . . . . .	36
7.2.	Diseño . . . . .	74
7.2.1.	Diagramas de Colaboración . . . . .	74
7.2.2.	Diagramas de Clase . . . . .	87
7.2.3.	Diagramas de Paquetes . . . . .	93
7.3.	Implementación . . . . .	97
7.3.1.	Arquitectura de TariyKDD . . . . .	97
7.3.2.	Descripción de clases . . . . .	99
7.3.3.	Casos de uso reales . . . . .	102
<b>8.</b>	<b>Conclusiones</b>	<b>143</b>

## **7.3. Implementación**

### **7.3.1. Arquitectura de TariyKDD**

Para el desarrollo de TariyKDD se utilizaron computadores con procesador AMD 64 bits, disco duro Serial ATA, útil al tomar los datos desde un repositorio y al momento de realizar pruebas de rendimiento de los algoritmos, ya que su velocidad de transferencia es de 150 MB/sg; además la RAM que se utilizó siempre fue superior a los 512 MB, ya que la Minería de Datos requiere grandes cantidades de memoria por el tamaño de los conjuntos de datos.

El sistema operativo sobre el cual se trabajó durante la implementación de TariyKDD es Fedora Core en su versión 3 y 5. El lenguaje de programación en el que está elaborado TariyKDD es Java 5.0, actualización 06.

Dentro del proceso de Descubrimiento de Conocimiento, TariyKDD comprende las etapas de Selección, Preprocesamiento, Minería de Datos y Visualización de Resultados. De esta forma la implementación de la herramienta se hizo a través de los siguientes módulos de software:

#### **Módulo de Conexión**

El Módulo de Conexión permite al usuario acceder a los conjuntos de datos a través de un Archivo Plano o una Base de Datos.

La opción Archivo Plano le permite al usuario seleccionar un conjunto de datos que se encuentra en disco, en un archivo de acceso aleatorio, el formato para el archivo debe ser ARFF[15], debido a que este es uno de los más conocidos y tiene una estructura que lo hace fácil de comprender por ser estándar (debido a su estructura de etiquetas).

En cuanto a la Conexión a Bases de Datos, TariyKDD puede conectarse con PostgreSQL a través de su manejador JDBC tipo 4. Este driver es el más eficiente ya que traduce de forma directa las peticiones del API Java al protocolo nativo del Sistema Gestor, con la ventaja de que resulta sencilla la migración a otro diferente, lo único que habría que hacer sería descargar el driver del fabricante adecuado.

#### **Almacenamiento de los datos en memoria**

Una de las principales dificultades dentro del proceso de descubrimiento de conocimiento es el uso adecuado de los recursos del sistema y en especial de la

memoria principal si tenemos en cuenta que se pretende trabajar con amplios volúmenes de datos. Ya sea cargando un conjunto de datos desde un archivo plano o directamente desde una conexión a un SGBD se espera organizar estos datos de una manera compacta con el objetivo de almacenar esta información en memoria principal evitando repetidas llamadas a disco lo que significa un aumento en los tiempos de ejecución de la herramienta.

Formatos tradicionales para el almacenamiento de transacciones, como el formato ARFF, trabajan con cabeceras donde se registran los diferentes campos o atributos del conjunto de datos seguidos de las transacciones como tal, separadas una de otra por cambios de línea donde cada atributo esta separado a su vez por comas. En conjuntos de datos discretizados cuyos atributos pueden tomar un rango determinado de valores, dentro de la parte en donde se almacenan los datos es común encontrar segmentos de transacciones que coinciden o incluso transacciones completas que se repiten.

Es posible aprovechar estas coincidencias dentro de una estructura de datos como un árbol N-Ario donde cada rama represente una posible transacción y donde las bifurcaciones dentro de esa rama representen segmentos compartidos con otras transacciones o inclusive transacciones que estén contenidas dentro de esa misma rama.

Para explicar de mejor manera esta propuesta consideremos el conjunto de datos representado en la siguiente tabla, hay que tener en cuenta que los items del conjunto de datos original son codificados para mejorar la administración de la memoria.

<b>T</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
1	1	1	3	4	5
2	1	1	3	4	6
3	1	2	3	6	7
4	2	3	4	1	3
5	1	2	3	6	7
6	2	3	4	6	5

Se puede ver que los cuatro primeros campos de las transacciones 1 y 2 son iguales por lo que pueden compartir nodos dentro del árbol N-Ario.

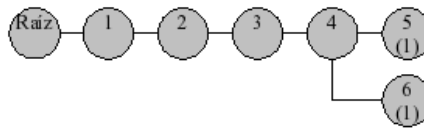


Figura 7.81: Árbol N-Ario

### 7.3.2. Descripción de clases

#### Paquete Utils

**Clase DataSet** En esta estructura los algoritmos de asociación almacenan los datos o items de forma comprimida, ocupando menos espacio en memoria. La estructura utilizada por DataSet es un árbol N-Ario que almacena los datos en cada nodo como tipo short. Lo particular de esta estructura es el aprovechamiento de la memoria principal, ya que en una sola rama almacena items de diferentes transacciones, controlando individualmente su número de apariciones.

**Clase FileManager** Esta clase gestiona todo lo relacionado con flujos de archivos, como por ejemplo crear un archivo plano, construir el diccionario de datos a partir de un archivo de acceso aleatorio, entre otras funciones.

**Clase BaseDatos** Esta clase gestiona todo lo relacionado con el manejo de las Bases de Datos, como la conexión, y la selección, de atributos.

**Clase NodeNoF** Esta clase representa un nodo básico del DataSet, este nodo no tiene soporte.

**Clase NodeF** Esta clase extiende a la clase NodeF y agrega el soporte a cada nodo del DataSet.

**Clase AvlTree** Los itemsets frecuentes generados por los algoritmos de asociación son almacenados en un árbol AVL balanceado, cuya estructura se encuentra en esta clase.

**Clase AvlNode** Es en si, un nodo del árbol AVL que almacena los itemsets frecuentes. Tiene un campo de tipo ItemSet en donde se guarda el dato que va en el nodo y tiene los punteros derecho e izquierdo a los demás nodos del árbol.

**Clase ItemSet** La clase ItemSet almacena un conjunto de items o itemsets en un vector así como su respectivo soporte.

**Clase Transaction** Esta clase gestiona todas las operaciones que deben hacerse sobre las transacciones. Como por ejemplo cargar las transacciones para los diferentes algoritmos y así como también realiza los diferentes ordenamientos de las transacciones, por item y por soporte.

### **Paquete Apriori**

**Clase Apriori** Esta clase implementa todos los métodos necesarios para ejecutar el algoritmo Apriori. Los parametros necesarios para comenzar el algoritmo son: un soporte de tipo short y un dataset (estructura de tipo árbol N-Ario en la cual los datos son comprimidos) y sobre el cual se realizan tantos recorridos como itemsets frecuentes existan.

### **Paquete EquipAsso**

**Clase EquipAsso** Para ejecutar el algoritmo EquipAsso los parametros necesarios son: un soporte de tipo short y un dataset (estructura de tipo árbol N-Ario en la cual los datos son comprimidos). Basicamente para obtener los itemsets frecuentes, lo primero que se debe hacer es recorrer el árbol N-Ario tomar cada una de sus transacciones, realizar todas sus combinaciones y ver cual de ellas pasa soporte y clasifica como itemset frecuente.

**Clase Combinations** Recibe como parametros el tipo y el itemset a combinar. El tipo es un número que indica hasta que profundidad se desea combinar el itemset en cuestion.

### **Paquete FPGrowth**

**Clase FPGrowth** El algoritmo FPGrowth tiene su propio árbol N-Ario para almacenar los datos. Recorre el árbol y toma cada una de sus ramas, a partir de estas construye los Patrones Condicionales Base, luego los Patrones Condicionales y a partir de estos determina cuales son los itemsets frecuentes, los cuales se almacenan en un árbol AVL balanceado.

**Clase FPGrowthNode** Clase que tiene la estructura del árbol N-Ario del algoritmo FPGrowth. Es decir tiene los punteros necesarios para armar un árbol N-Ario, tiene un puntero al hijo, al padre y al hermano.

**Clase BaseConditional** Clase que almacena los Patrones Condicionales Base a partir del árbol N-Ario de la clase FPGrowth.

**Clase BaseConditonals** Solo los nodos que pasan el soporte mínimo se consideran frecuentes, estos, tienen un puntero a cada uno de sus Patrones Condicionales Base, a partir de los cuales se obtienen los itemsets frecuentes.

### **Paquete MateBy**

**Clase MateBy** Así como los demás algoritmos, MateBy utiliza el dataset o estructura de tipo árbol N-Ario para comprimir los datos que se van a minar. A partir de estos datos se realizan combinaciones y se calcula su entropía y su ganancia. Las combinaciones con la mayor ganancia se almacenan en un árbol de reglas, conformando así los resultados de MateBy.

**Clase Entro** Agrupa los nodos del árbol de acuerdo a su padre o rama y determina quienes tienen la mayor ganancia, de acuerdo a esto se construye el árbol de reglas.

### 7.3.3. Casos de uso reales

#### Ingreso a la aplicación

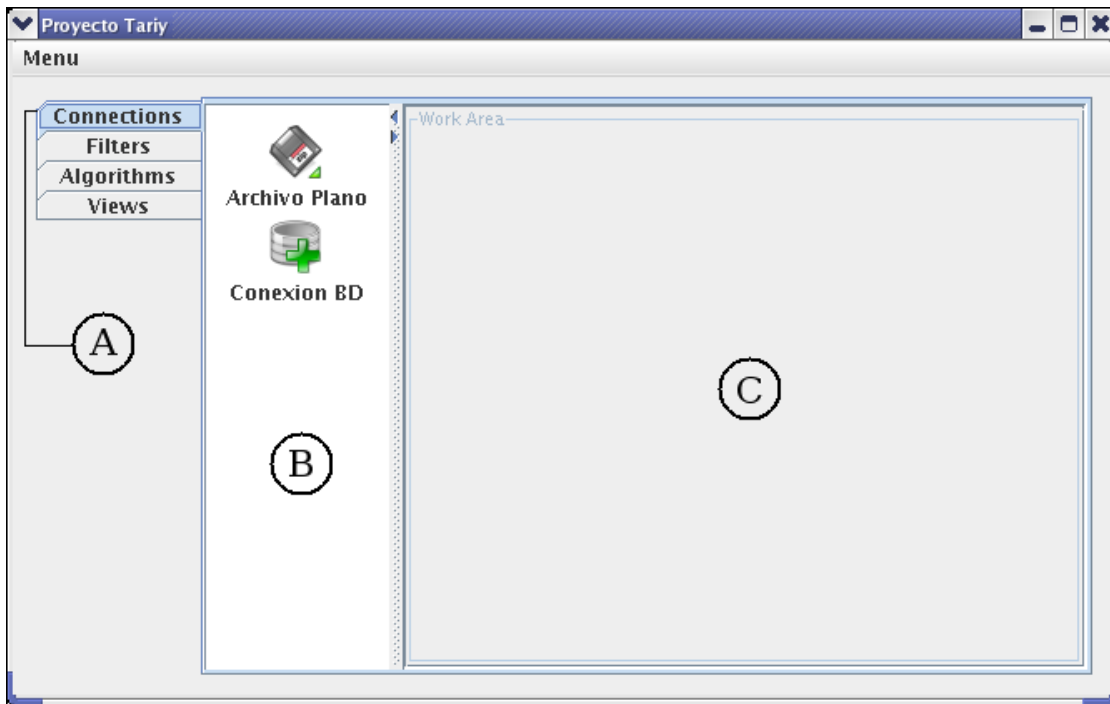


Figura 7.82: Ingreso a la aplicación

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario ejecuta la aplicación	2. La interfaz gráfica de la aplicación aparece como se muestra en la figura. A: Área de pestañas a través de las cuales se puede acceder a los diferentes módulos de la aplicación. Ej, el módulo por defecto es 'Connections'. B: Área en la que aparecen las opciones de cada módulo. Ej, las opciones del módulo 'Connections' son 'Archivo Plano' y 'Conexión DB'. C: Área de trabajo sobre la que se arman los proyectos de Minería de Datos



## Módulo filtros

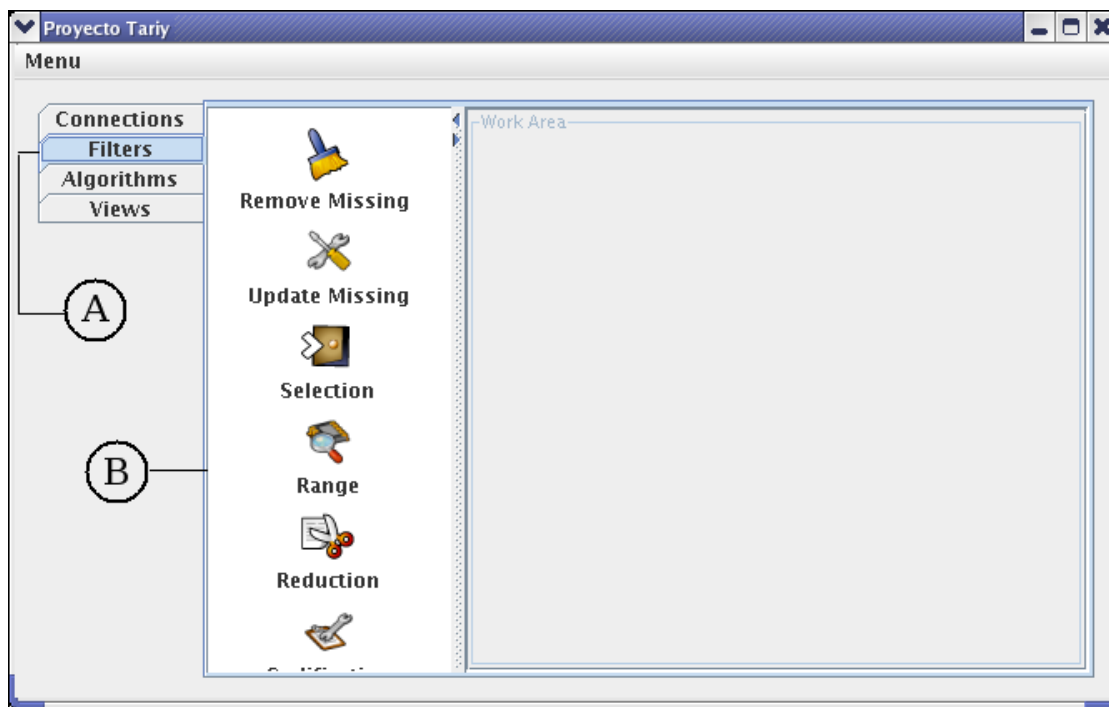


Figura 7.83: Módulo filtros

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click en la pestaña A 'Filtros'	2. Aparecen B Las opciones del módulo 'Filtros'

## Módulo algoritmos

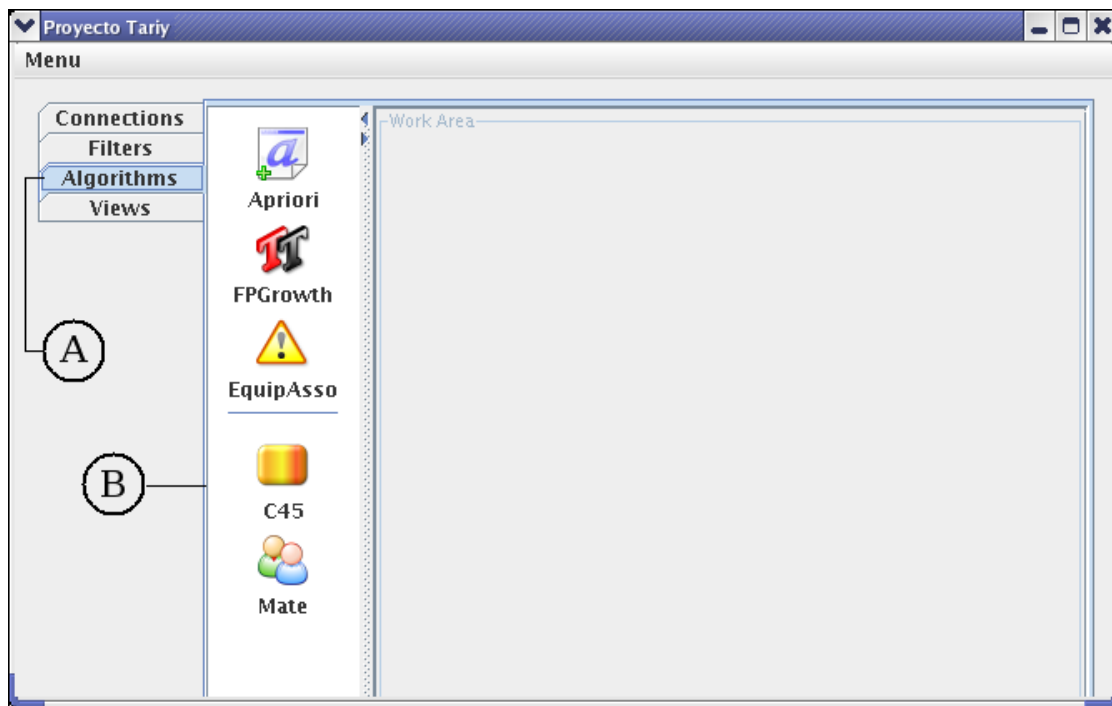


Figura 7.84: Módulo algoritmos

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click en la pestaña A 'Algoritmos'	2. Aparecen B Las opciones del módulo 'Algoritmos'

## Módulo visualización

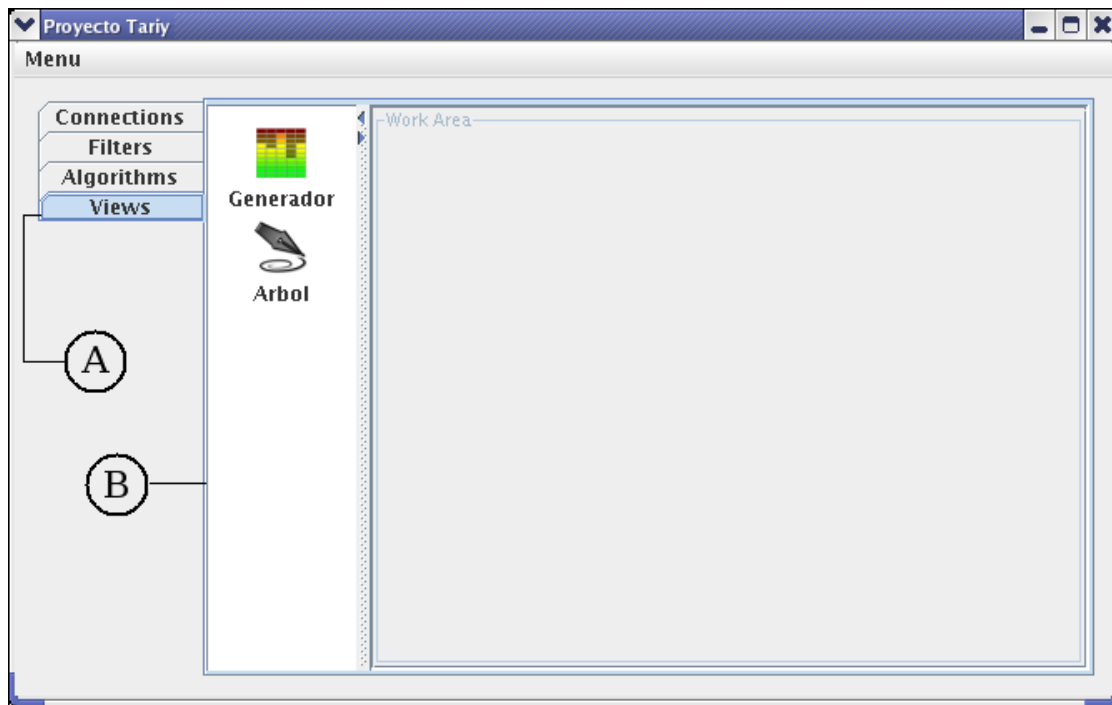


Figura 7.85: Módulo visualización

ACCIÓN $\frac{1}{2}$ DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click en la pestaña A 'Visualización'	2. Aparecen B Las opciones del módulo 'Visualización'

## Conexión a un archivo plano

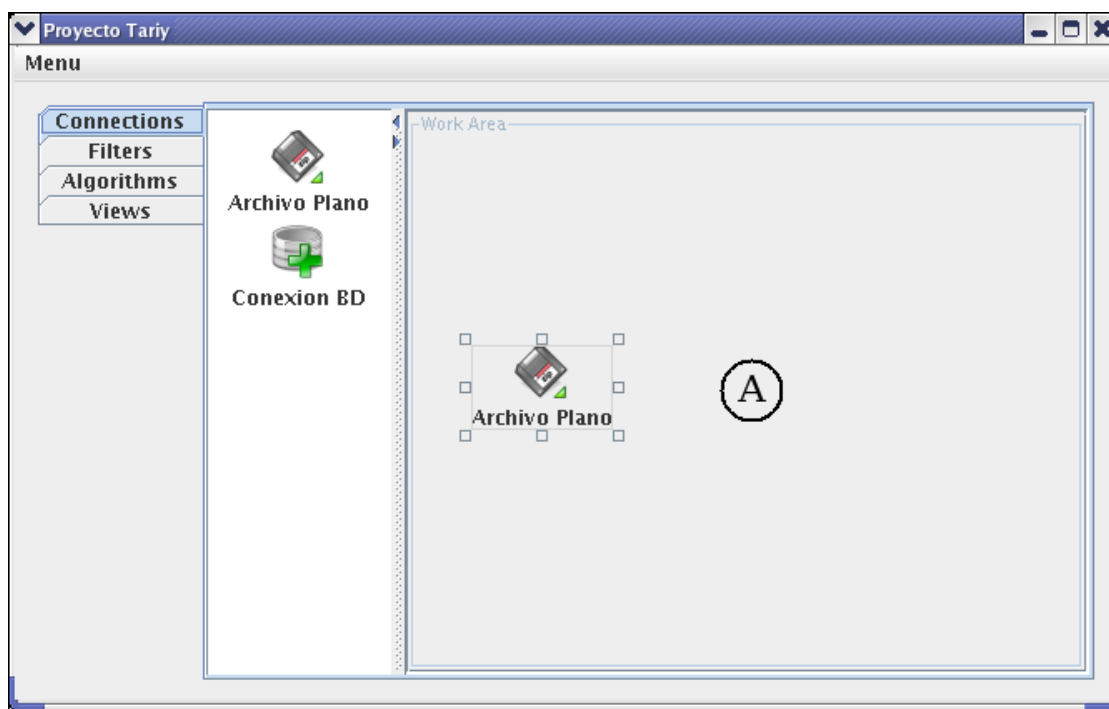


Figura 7.86: Conexión a un archivo plano

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click sobre el ícono 'Archivo de Texto'	2. El ícono 'Archivo de Texto' aparece sobre A: área de trabajo.

## Conexión a una base de datos

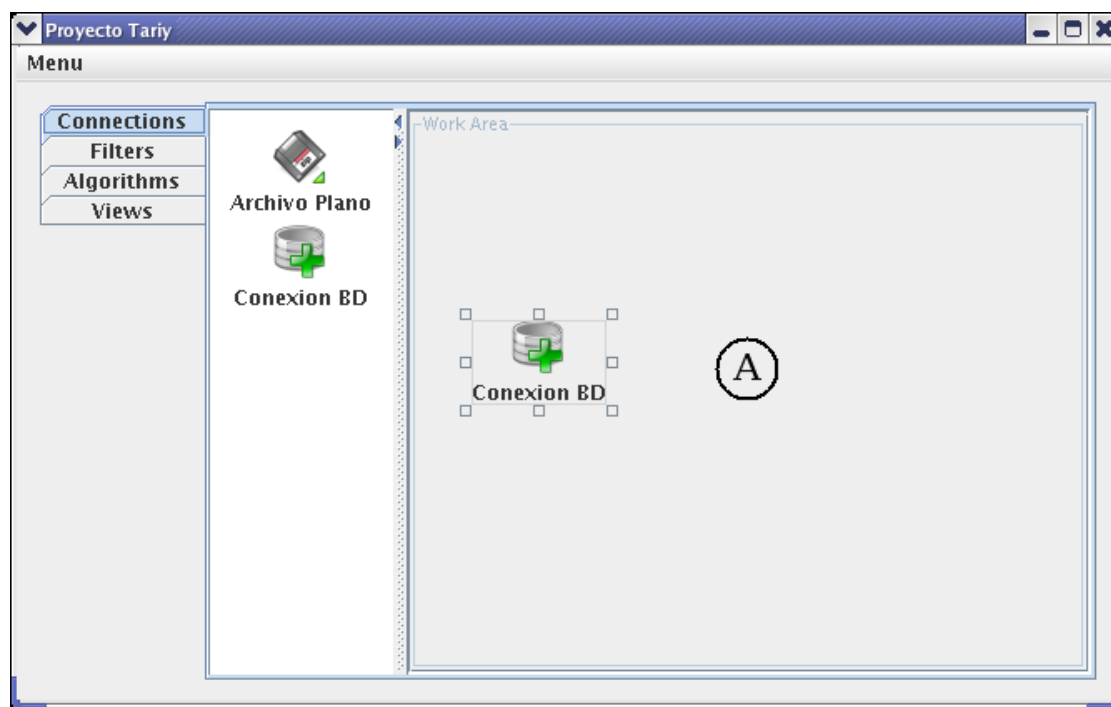


Figura 7.87: Conexión a una base de datos

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click sobre el ícono 'Conexión BD'	2. El ícono 'Conexión BD' aparece sobre A: área de trabajo.

## Menú emergente conexión BD

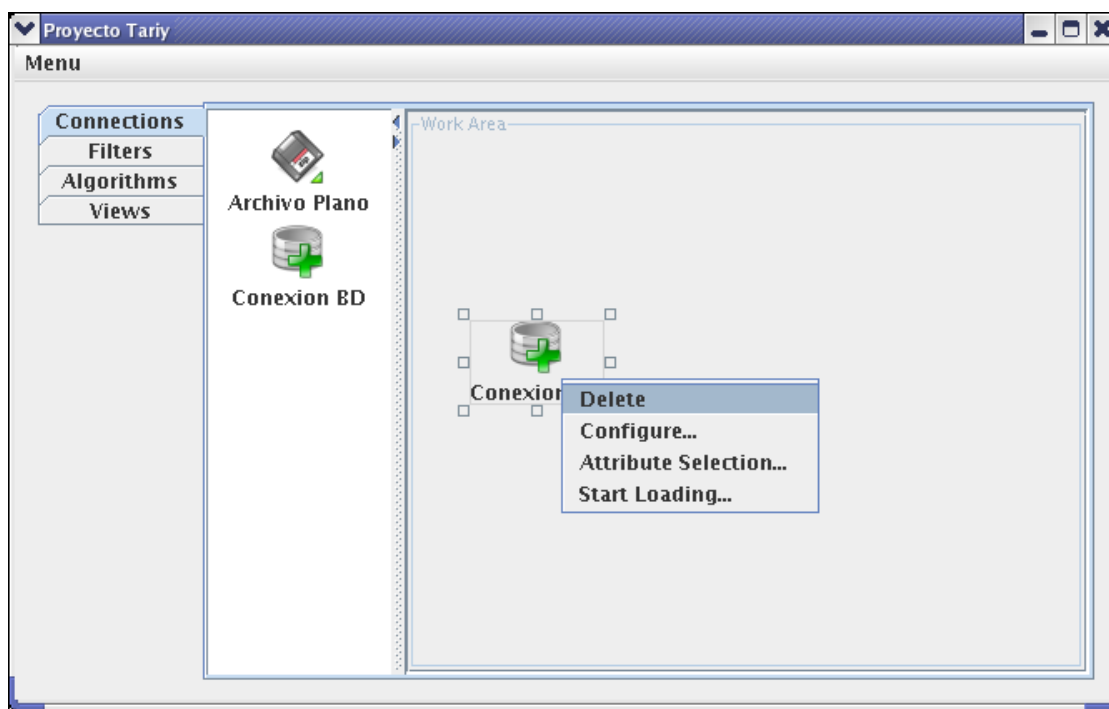


Figura 7.88: Menú emergente conexión BD

ACCIÓN $\frac{1}{2}$ DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el ícono 'Conexión BD'	2. Se despliega A: menú del ícono 'Conexión BD'. Las opciones son: 'Delete': usada para eliminar el ícono del área de trabajo. 'Configure': usada para configurar la conexión a una base de datos. 'Selección de atributos': usada para seleccionar de forma gráfica los datos que sería usados más adelante. 'Cargar': ejecuta el query que se generará en la selección de atributos

## Configuración conexión BD

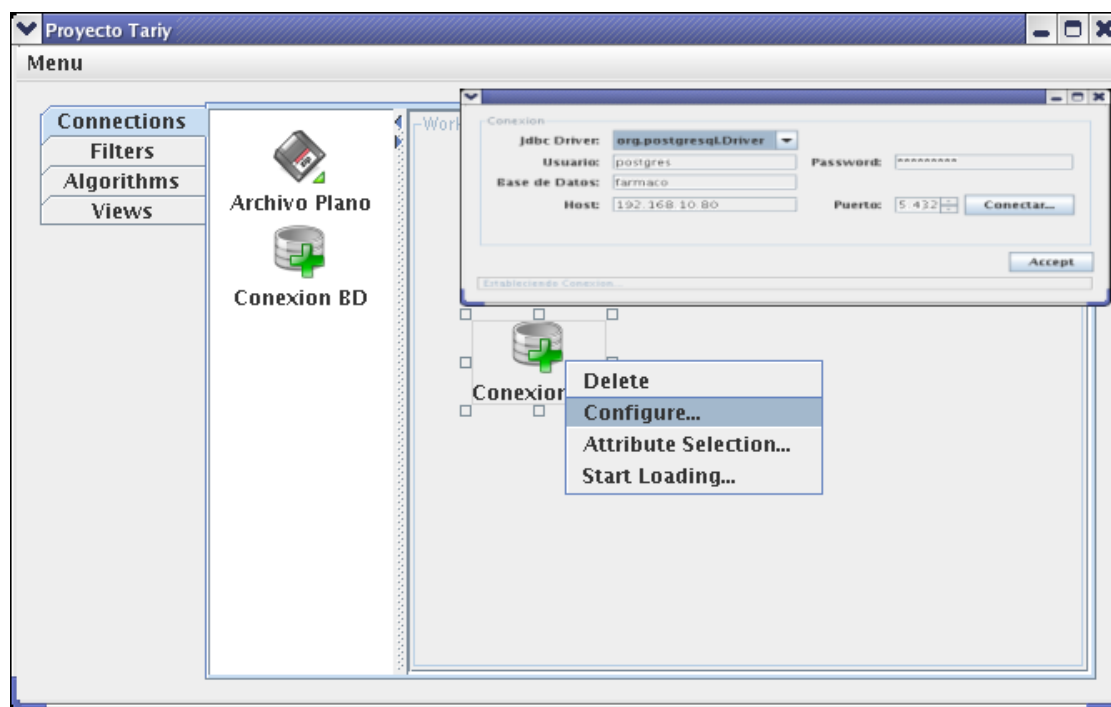


Figura 7.89: Configuración conexión BD

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el ícono 'Conexión BD' y selecciona la opción 'Configure'	2. Emerge una ventana de configuración de conexión a bases de datos

## Ventana de conexión BD

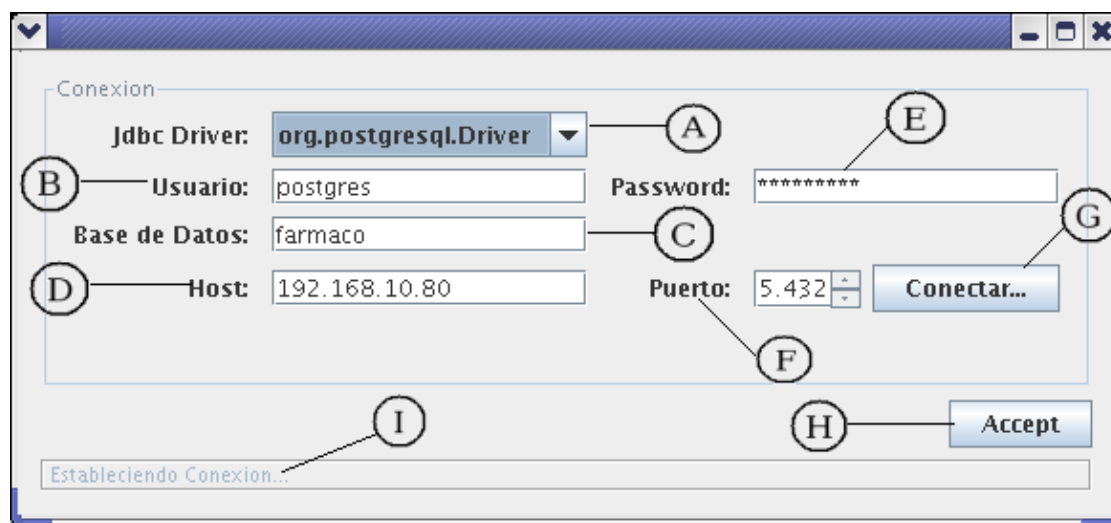


Figura 7.90: Ventana de conexión BD

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario desea configurar la conexión a una base de datos'	2. Las opciones de la ventana de configuración de conexión a bases de datos tiene los siguientes campos: A: Lista de controladores ODBC para varios tipos de bases de datos. B: Nombre del usuario de la base de datos. C: Nombre de la base de datos. D: Nombre del servidor. E: 'Password': clave de acceso a la base de datos. F: nmero del puerto utilizado para la comunicación con la base de datos. G: botón de conexión. H: botón para aceptar la conexión hecha. I: mensaje que indica el estado de la conexión.



## Selección de atributos

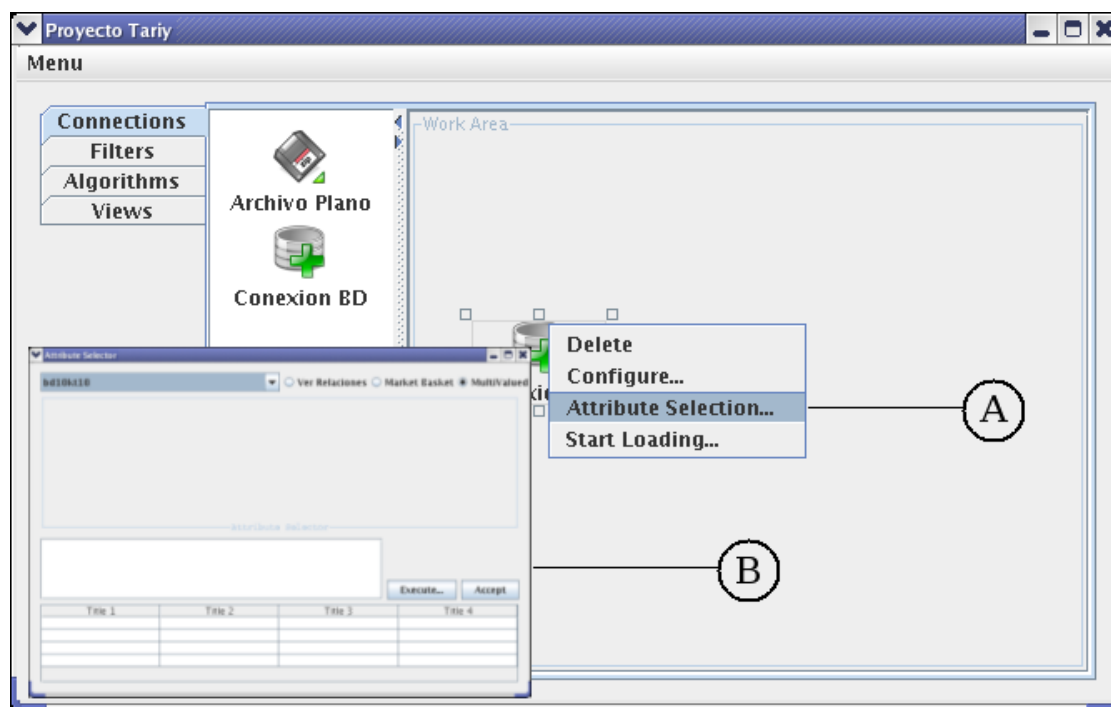


Figura 7.91: Selección de atributos

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el ícono 'Conexión BD' para hacer la selección de atributos	2. Aparece el menú emergente del ícono y se ejecuta la ventana de selección de atributos A.
3. El usuario hace click en la opción B: 'Selección de Atributos'	4. Aparece la ventana de selección de atributos B.

## Ventana selección de atributos

The screenshot shows the 'Attribute Selector' window with the following components:

- Top Bar:** A dropdown menu set to 'prueba' and three radio buttons: 'Ver Relaciones' (selected), 'Market Basket', and 'MultiValued'.
- Diagram Area:** A visual representation of the query. It shows two tables: 'bd10kt10' (with attributes 'tid' and 'item') and 'productos' (with attributes 'sq', 'item', and 'nom'). A line connects the 'item' attribute of 'bd10kt10' to the 'item' attribute of 'productos', labeled with the condition 'bd10kt10.item = productos.item'. Various letters (A, B, C, D, E, F, G, H, I, J, K) are placed around the diagram to indicate different parts or steps.
- SQL Editor:** A text area containing the following SQL query:
 

```
SELECT DISTINCT productos.nom, bd10kt10.tid, productos.item
FROM productos, bd10kt10
WHERE bd10kt10.item = productos.item
ORDER BY productos.nom
```
- Buttons:** 'Execute...' and 'Accept' buttons are located at the bottom right of the diagram area.
- Result Table:** A table with three columns: 'nom', 'tid', and 'item'. It contains five rows of data. A letter 'J' is placed next to the second row.
 

nom	tid	item
BUSCAPINA COMPOSITUN MF COMPRIMIDOS	534369	11074
BUSCAPINA COMPOSITUN MF COMPRIMIDOS	535372	11074
SIMILAC CON HIERRO	536422	2110
SUERO ORAL PEDIALYTE 30 NARANJA	539586	2002
SUERO ORAL PEDIALYTE 30 MANDARINA	541249	2142

Figura 7.92: Ventana selección de atributos

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario desea hacer la selección de atributos	2. Aparece la ventana de selección de atributos. A: lista desplegable de las tablas de la base de datos a la que se ha conectado. Al seleccionar una de ellas su representación gráfica aparecerá en el espacio de trabajo E. B: opción que permite ver las relaciones establecidas a través de la línea de conexión de atributos entre las tablas. C: esta opción es útil cuando se trabajan problemas de canasta de mercado. D: opción para trabajar tablas multivaluadas. F: línea que permite realizar las relaciones entre atributos de dos tablas. El resultado de la relación establecida se refleja en el query. G: Si se hace click sobre uno de los atributos aparece un ícono de verificación que indica los campos que serán mostrados al ejecutar el query. H: espacio en el que se crea el query. Es posible editarlo manualmente. I: botón de ejecución del query. J: tabla en la que se muestra el resultado de la ejecución del query. K: botón para aceptar las operaciones realizadas.

## Filtro Remove Missing

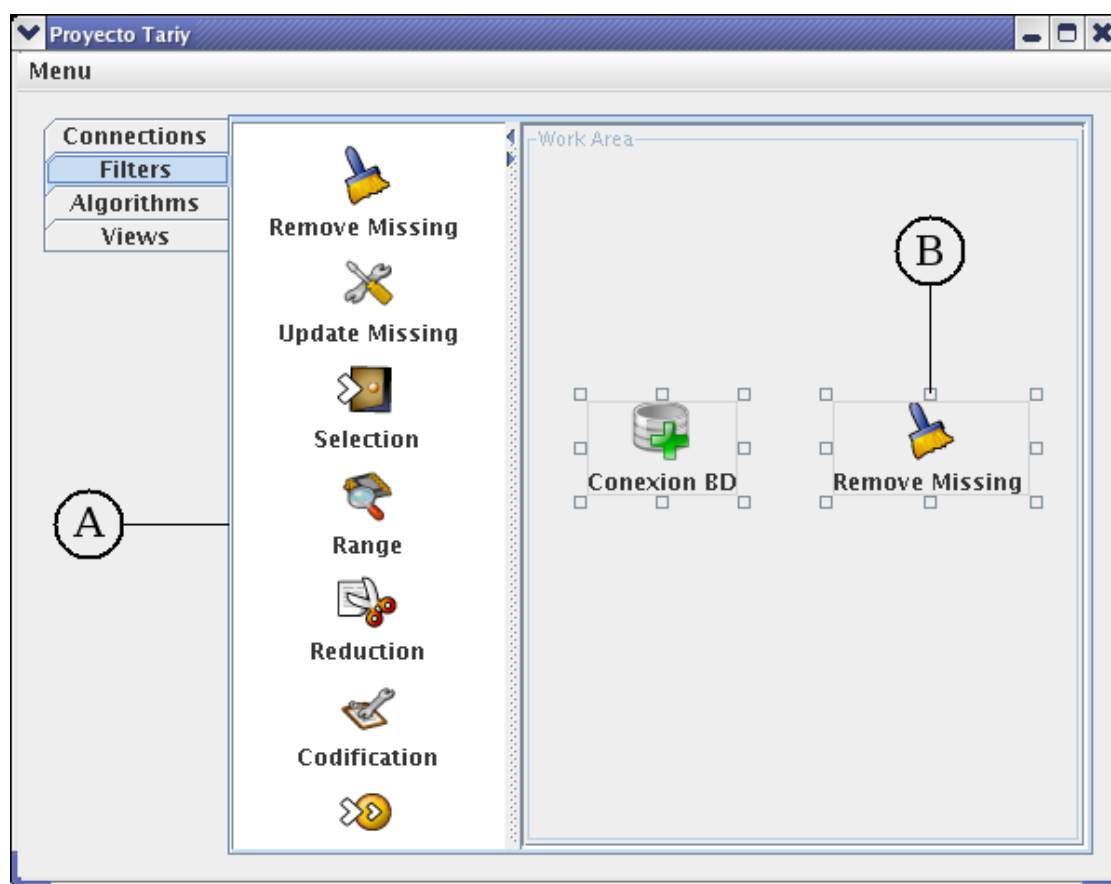


Figura 7.93: Filtro Remove Missing

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click sobre uno de los íconos del módulo A: 'Filtros'.	2. En el área de opciones del módulo aparecen los 9 íconos correspondientes a los filtros
3. El usuario hace click sobre uno de los íconos correspondientes a los filtros.	4. El ícono correspondiente aparece en el área de trabajo B.

## Conexión filtros a BD

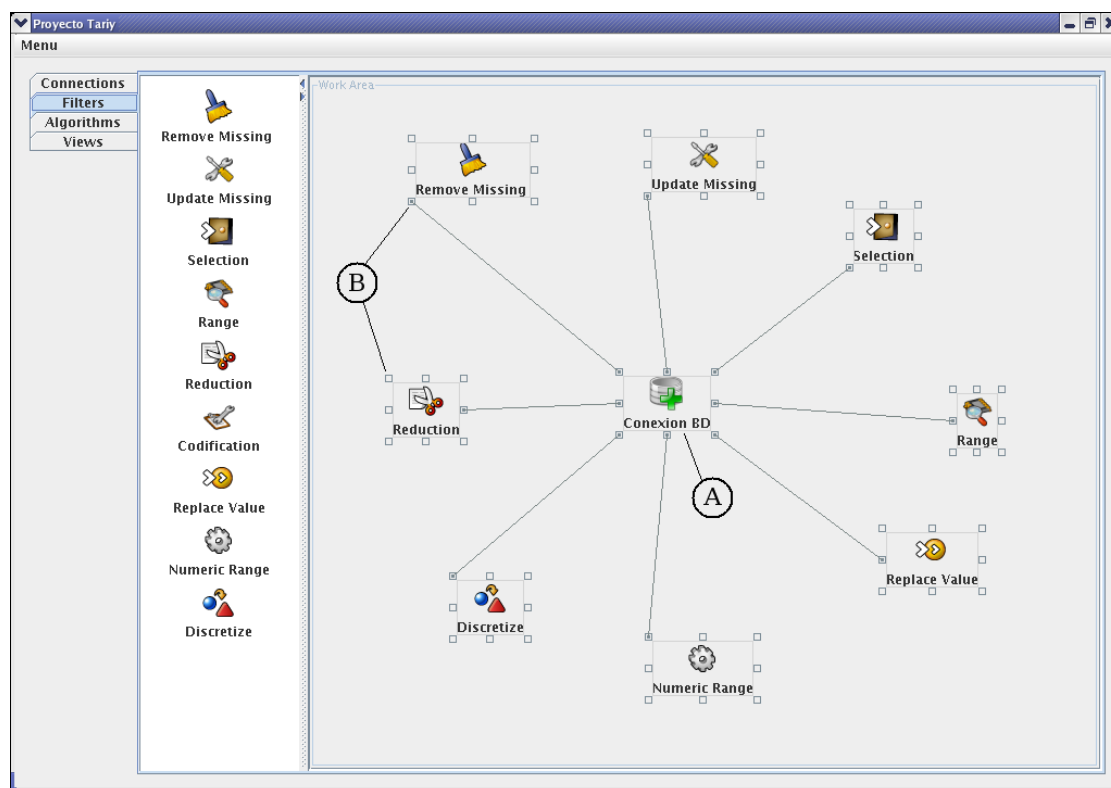


Figura 7.94: Conexión filtros a BD

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario conecta una base de datos a alguno o varios de de los filtros A.	2. Los íconos pueden ser conecta-dos por medio de una línea B.

## Menú emergente de filtros

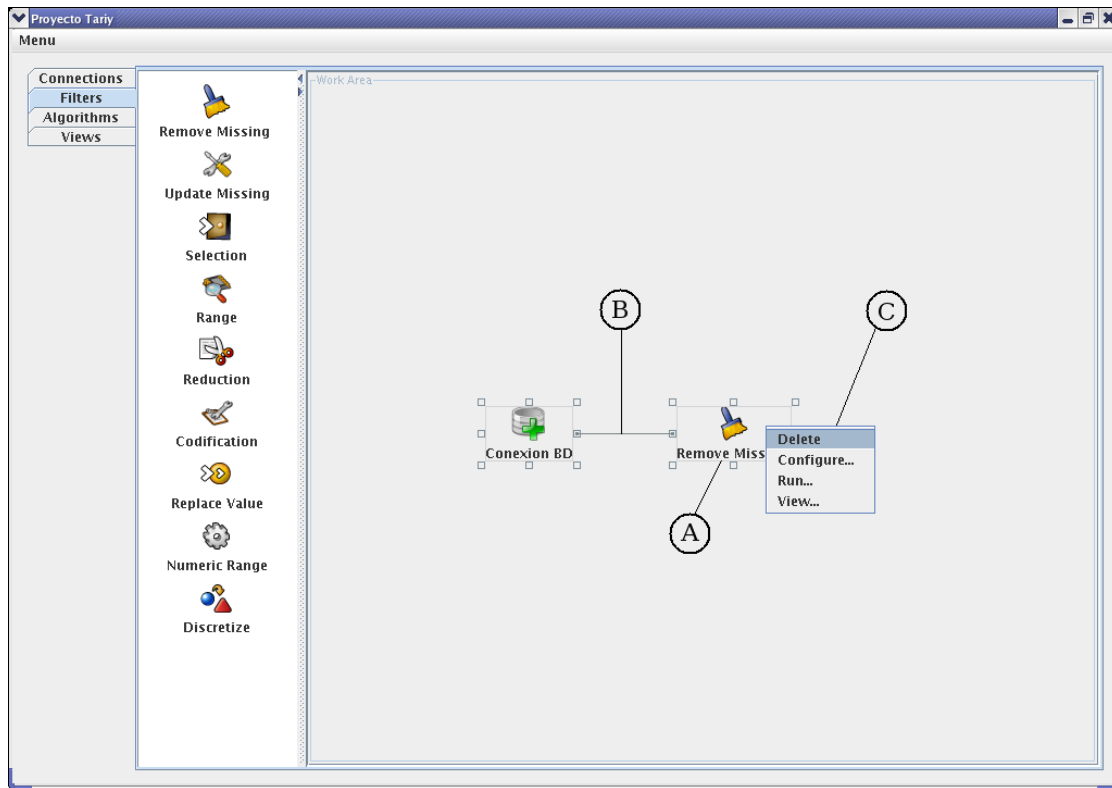


Figura 7.95: Menú emergente de filtros

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click sobre el ícono 'Remove Missing'.	2. El ícono aparece en el área de trabajo A.
3. El usuario conecta el filtro a la base de datos.	4. Aparece un hilo que conecta los íconos B.
5. EL usuario hace click derecho sobre el filtro.	6. Aparece el menu emergente del ícono C. La opción Delete, borra el filtro del área de trabajo. Este filtro no tiene ventana de configuración. La opción 'Run' ejecuta la aplicación del filtro. La opción 'View' muestra la ventana de visualización de datos que sería descrita en el siguiente caso de uso

## Visualización de datos filtrados



Figura 7.96: Visualización de datos filtrados

Ver Resultado de Eliminar Missing

Variables	Datos de Entrada	Datos Filtrados
	tid	item      nom
	457453	2142 SUERO ORAL PEDIALYTE 30 MAN...
	458534	2142 SUERO ORAL PEDIALYTE 30 MAN...
	462524	2143 SUERO ORAL PEDIALYTE 30 LIMON
	465453	8062 CREMA DIVINA COSMETICA
	467186	8062 CREMA DIVINA COSMETICA
	467523	8062 CREMA DIVINA COSMETICA
	467999	8062 CREMA DIVINA COSMETICA
	471704	8062 CREMA DIVINA COSMETICA
	473480	11074 BUSCAPINA COMPOSITUN MF CO...
	475686	8062 CREMA DIVINA COSMETICA
	478413	8062 CREMA DIVINA COSMETICA
	481201	2142 SUERO ORAL PEDIALYTE 30 MAN...
	482192	8062 CREMA DIVINA COSMETICA

Registros Eliminados : 0      Registros Actuales : 49

Figura 7.97: caso nueve



ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click sobre la opción 'View' del menu desplegable filtro en el área de trabajo .	2. Aparece la ventana de vizualización de datos filtrados y no filtrados. Los campos son, A: Variables o nombres de los campos de la tabla. B: Datos de entrada que son los datos que llegaron al filtro inicialmente. C: Datos filtrados que son el resultado de haber aplicado el filtro. D: nmero de registros eliminados al aplicar el filtro. E: Nmero de registros después de aplicar el filtro. En la figura 16 se ve la grilla sobre la que se muetran los datos en el caso 'Datos de entrada'

## Configuración filtro Update Missing

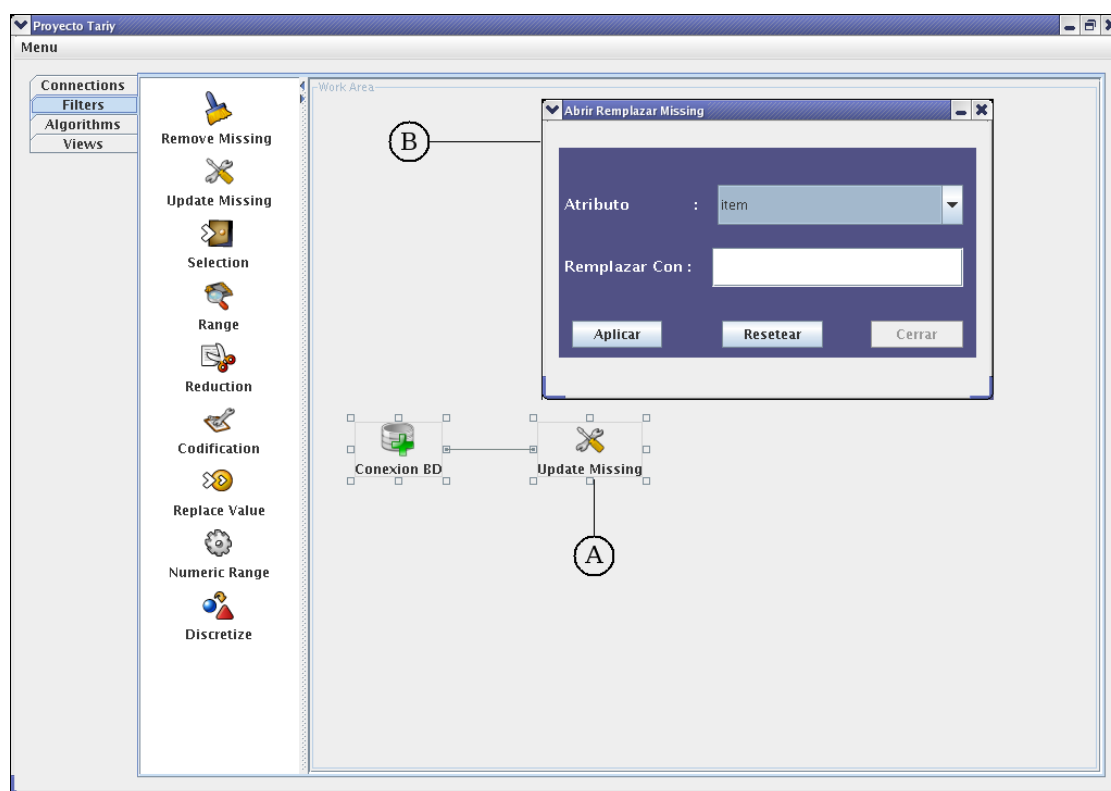


Figura 7.98: Configuración filtro Update Missing

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el filtro A y elige la opción 'Configuración'	2. Se muestra B la ventana de configuración correspondiente al filtro 'Update Missing'. Los campos son: Atributo, en el cual se escribe el nombre del atributo a buscar en el conjunto de datos. Reemplazar con, aqui se escribe el nuevo valor del atributo

## Configuración filtro Selection

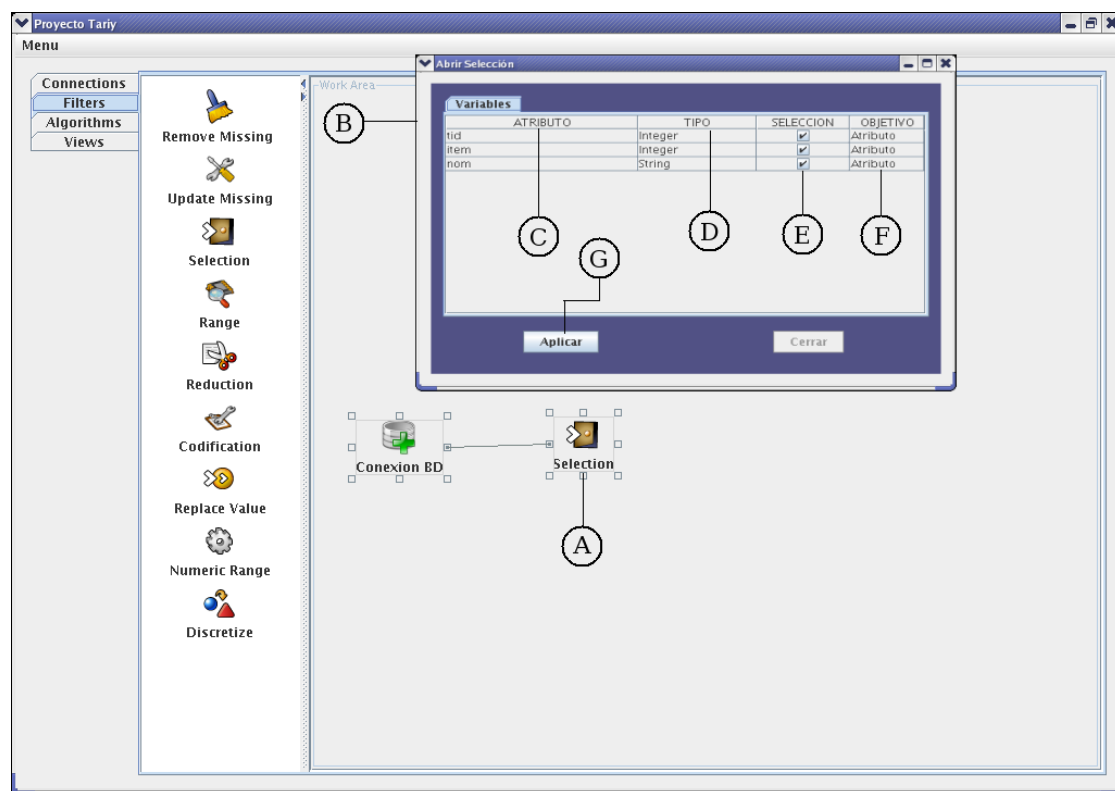


Figura 7.99: Configuración filtro Selection

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el filtro A y elige la opción 'Configuración'	2. Se muestra B la ventana de configuración correspondiente al filtro 'Selection'. Los campos son: C: Atributo, en esta grilla se muestran los nombres de los atributos seleccionados. D: Tipo, muestra el tipo de datos de los atributos. E: cajas de verificación para escoger los atributos a utilizar. F: es posible escoger un atributo clase haciendo click sobre estos campos. Esto es útil en experimentos de clasificación. G: el botón 'Aplicar' debe ser precionado para que el filtro sea aplicado.

## Configuración filtro Range

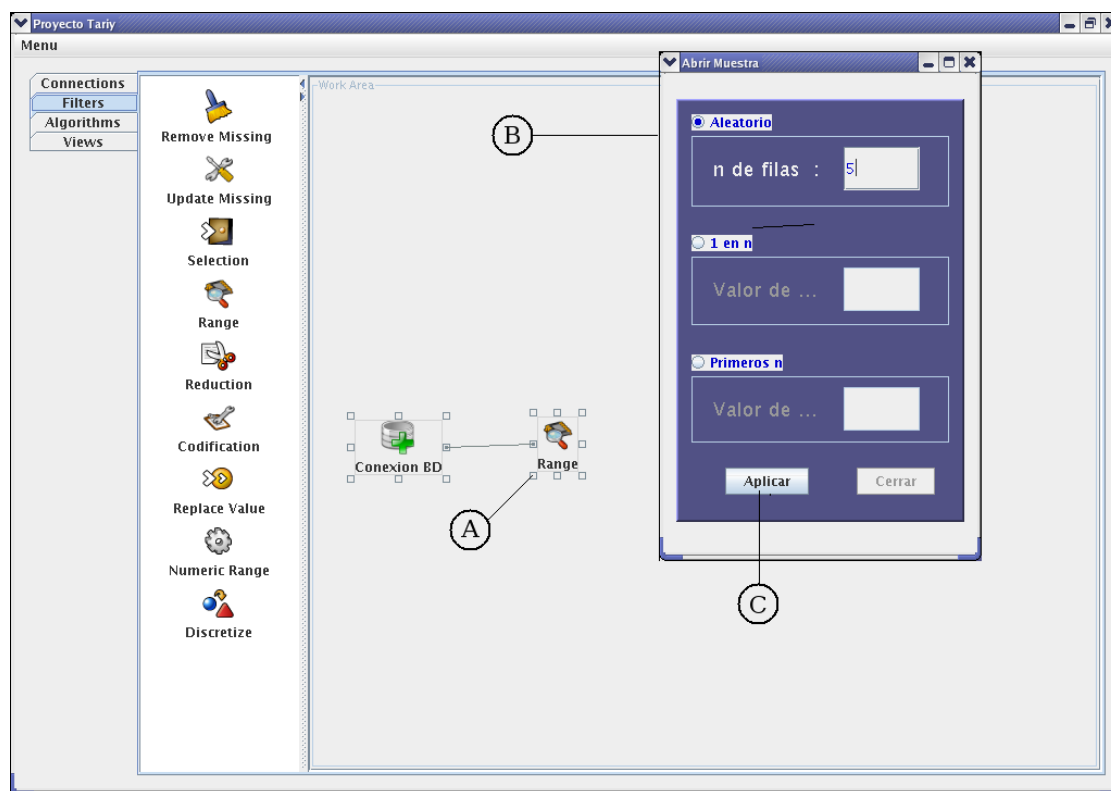


Figura 7.100: Configuración filtro Range

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el filtro A y elige la opción 'Configuración'	2. Se muestra B la ventana de configuración correspondiente al filtro 'Range'. Los campos son: <b>Aleatorio</b> , en donde se escribe el número <b>n</b> de filas que se desea sean escogidas aleatoriamente. <b>1 en n</b> , donde <b>n</b> es el periodo utilizado para seleccionar los datos a utilizar. <b>Primeros n</b> , donde <b>n</b> es el nmero campos a incluir en la selección a partir del primero.

## Configuración filtro Reduction

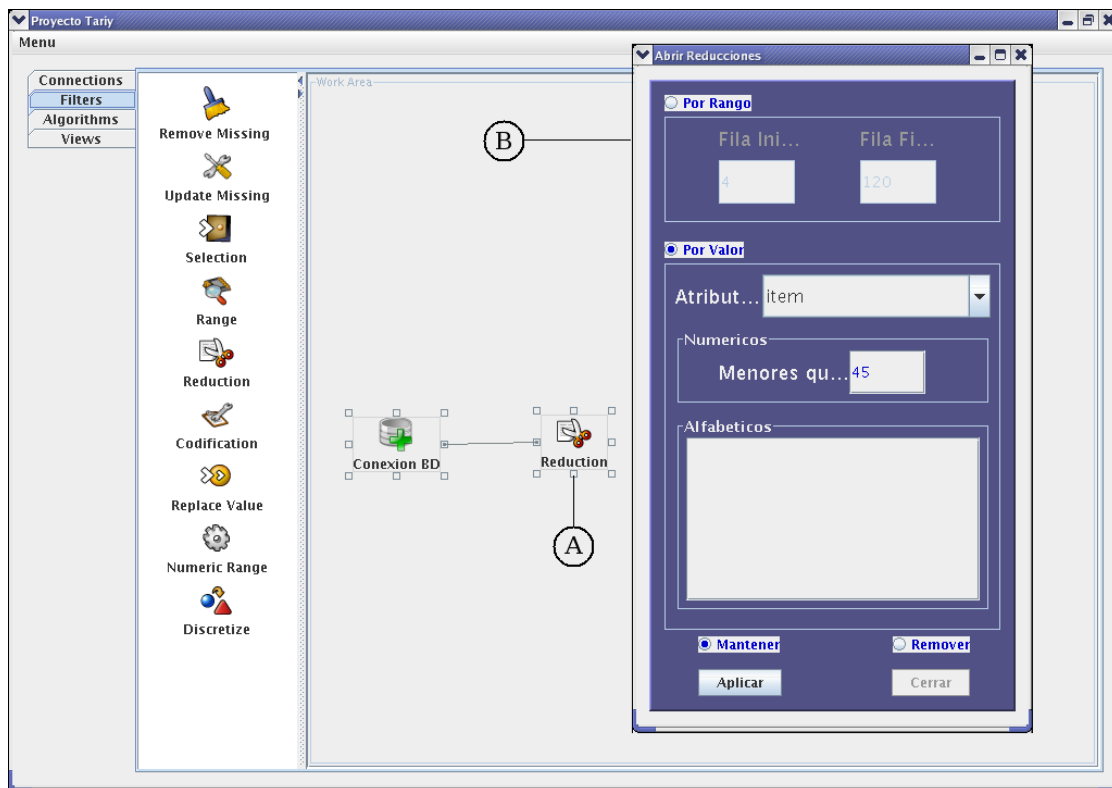


Figura 7.101: Configuración filtro Reduction

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el filtro A y elige la opción 'Configuración'	2. Se muestra la ventana B de configuración correspondiente al filtro 'Reduction'. Los campos son: <b>Por rango</b> , los campos son 'Fila inicial' donde se escribe la fila a partir de la cual inicia el rango y 'Fila final' que es el límite superior del rango. <b>Por Valor:</b> Se elige el nombre del atributo y luego en caso de que los valores a quitar sean numéricos en el campo 'Menores que' se especifica el número a partir del cual se hace la reducción. Si el atributo es alfabético se escribe su valor en el área de texto y en las casillas de selección se especifica si ese valor se desea 'Mantener' o 'Remover'. <b>Aplicar:</b> ejecuta el filtro.

## Configuración filtro Codification

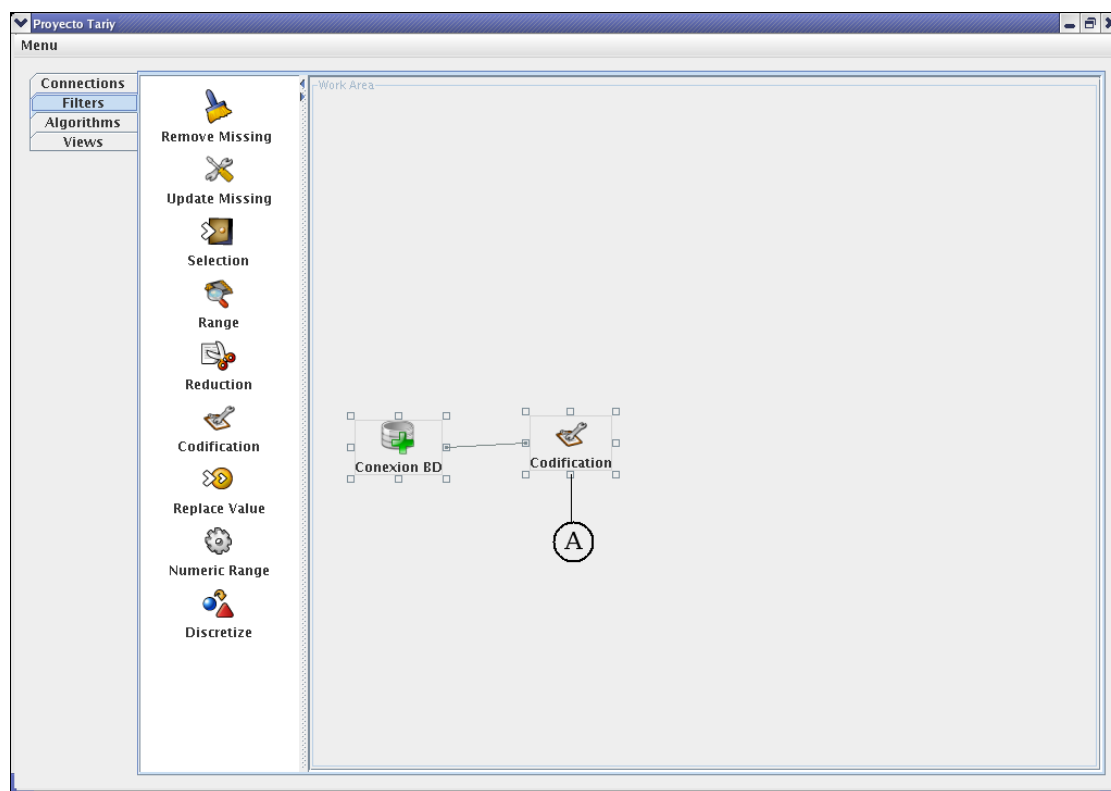


Figura 7.102: Configuración filtro Codification

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el filtro A y elige la opción 'Configuración'	2. Se muestra la ventana de configuración correspondiente al filtro 'Codification'. Este filtro no tiene ventana de configuración. Se aplica para asignar un número a valores alfabéticos



## Configuración filtro Replace Value

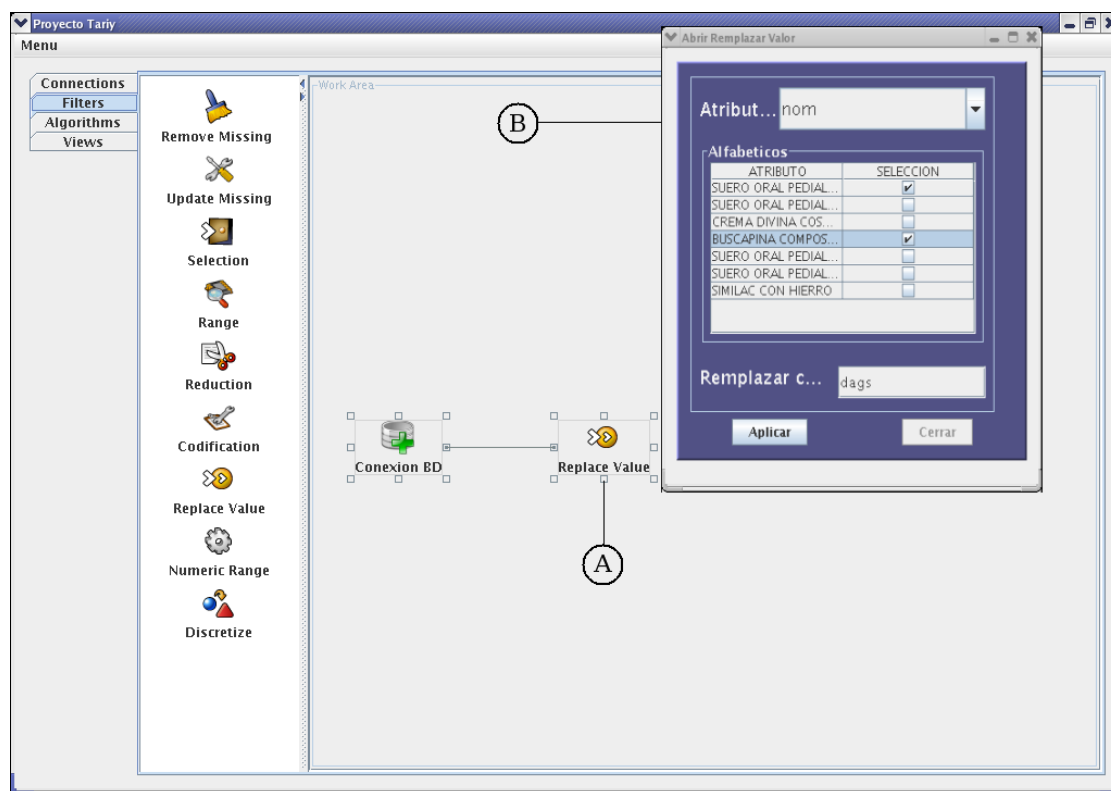


Figura 7.103: Configuración filtro Replace Value

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el filtro A y elige la opción 'Configuración'	2. Se muestra la ventana de configuración correspondiente al filtro 'Replace Value'. Los campos son: Atributo, en el cual se elige el nombre del atributo a buscar en el conjunto de datos. Reemplazar con, aquí se escribe el nuevo valor del atributo. <b>Aplicar</b> : ejecuta el filtro.

## Configuración filtro Numeric Range

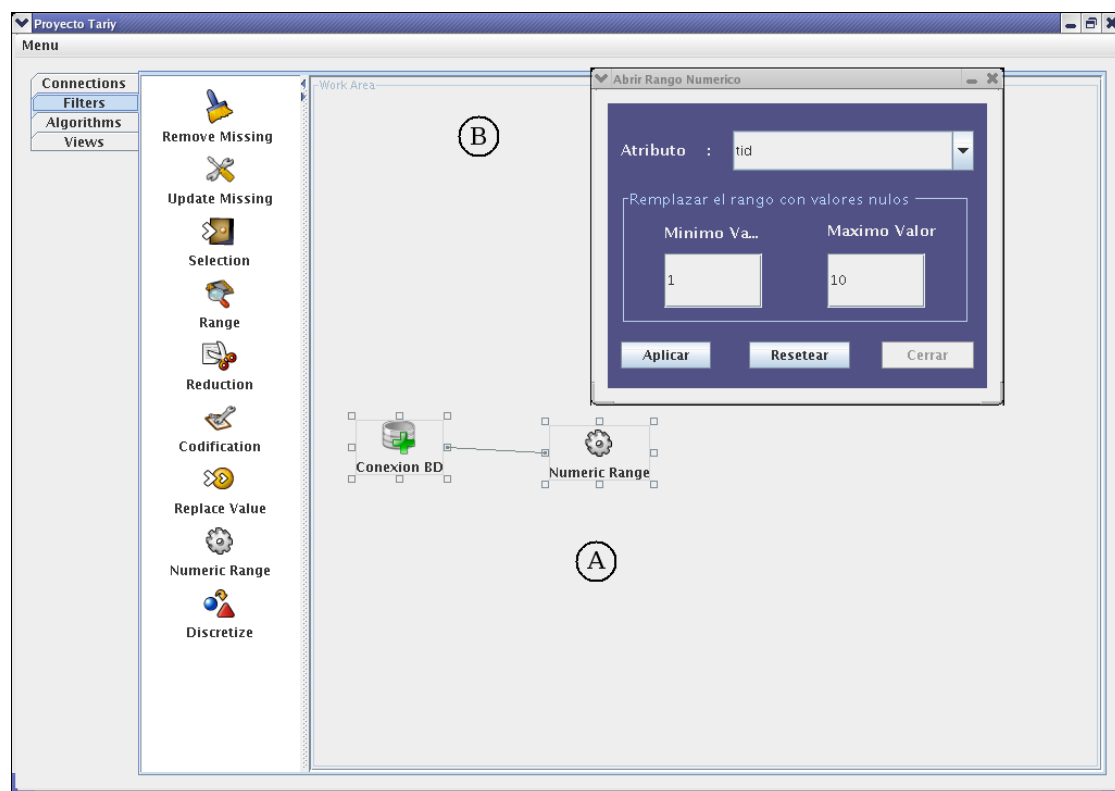


Figura 7.104: Configuración filtro Numeric Range

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el filtro A y elige la opción 'Configuración'	2. Se muestra la ventana B de configuración correspondiente al filtro 'Numeric Range'. Los campos son: <b>Atributo</b> , en el cual se escribe el nombre del atributo a discretizar de tipo numérico. <b>Reemplazar rango con valores nulos</b> : aquí es posible especificar un rango de datos que serán convertidos a nulos. <b>Mínimo valor</b> : límite inferior del rango. <b>Máximo valor</b> : límite superior del rango. <b>Aplicar</b> : ejecuta el filtro. <b>Resetear</b> : deja los campos en blanco

## Configuración filtro Discretize

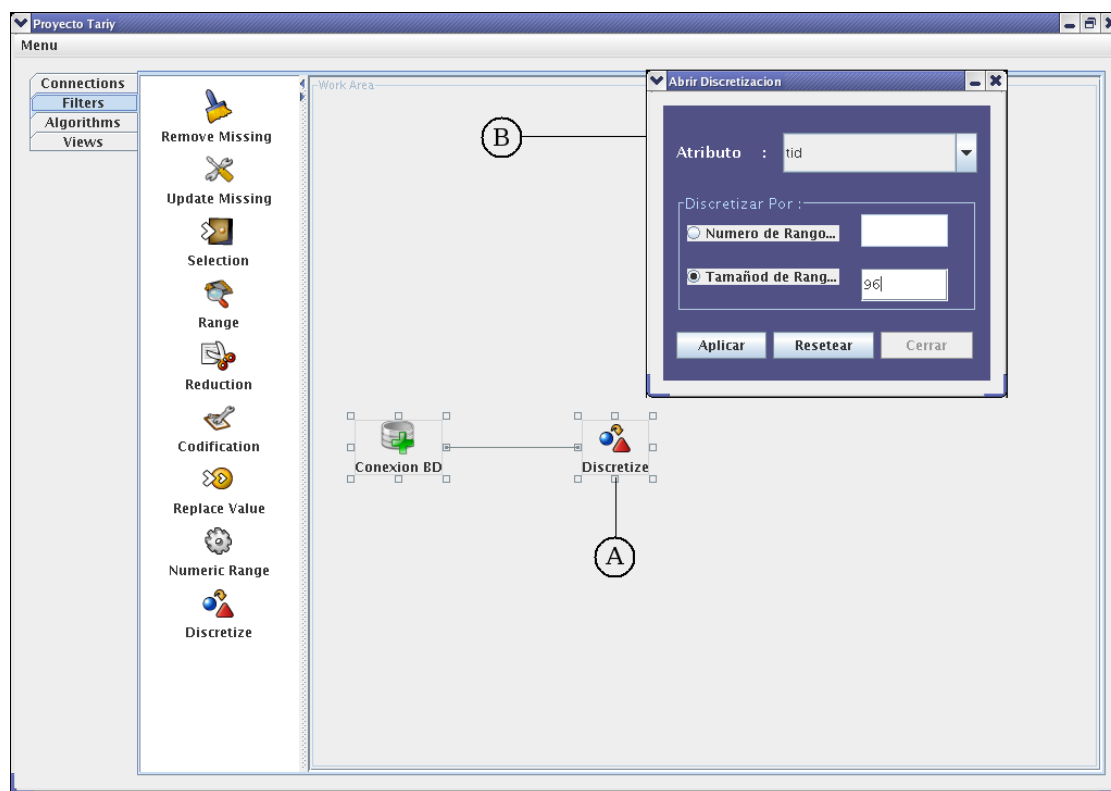


Figura 7.105: Configuración filtro Discretize

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre el filtro A y elige la opción 'Configuración'	2. Se muestra la ventana B de configuración correspondiente al filtro 'Discretize'. Los campos son: <b>Atributo</b> , en el cual se escribe el nombre del atributo a discretizar. <b>Discretizar por</b> : 'Número de rango': se puede establecer el número de rangos a crear. 'Tamaño del rango': se especifica el tamaño del rango <b>Aplicar</b> : ejecuta el filtro. <b>Resetear</b> : deja los campos en blanco

## Algoritmos

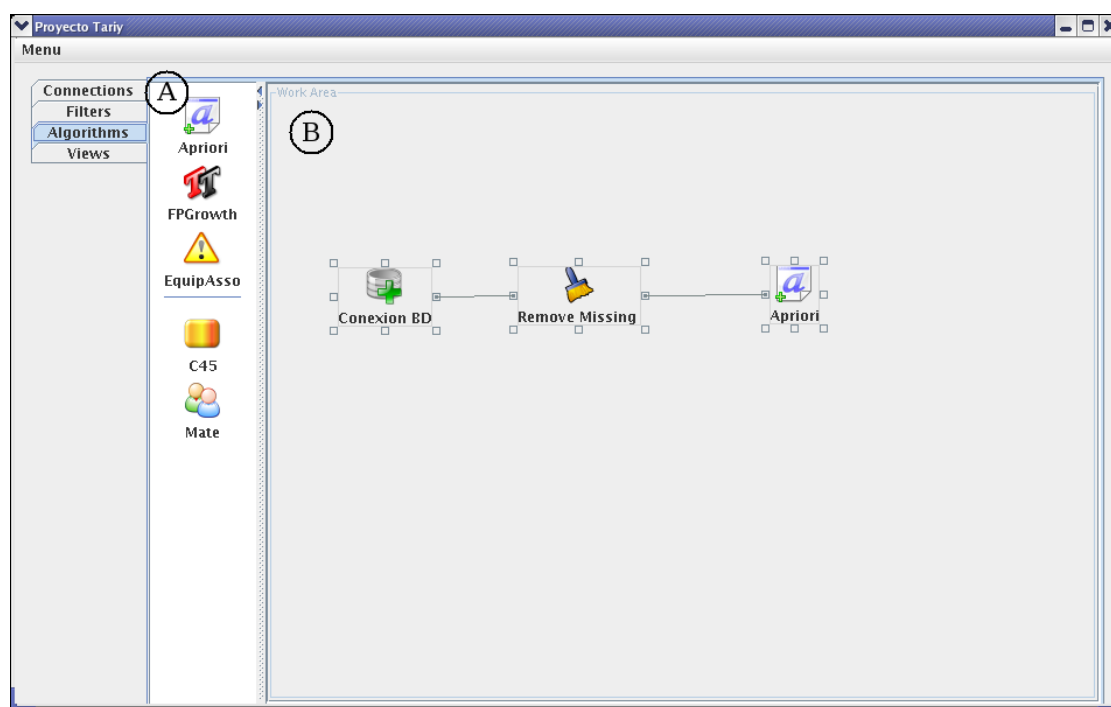


Figura 7.106: Algoritmo Apriori

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. Si el usuario quiere minar los datos con Apriori y presiona sobre A (Área de opciones), en el icono respectivo.	2. En B (Área de trabajo) aparece el icono del algoritmo Apriori.

## Opción Delete

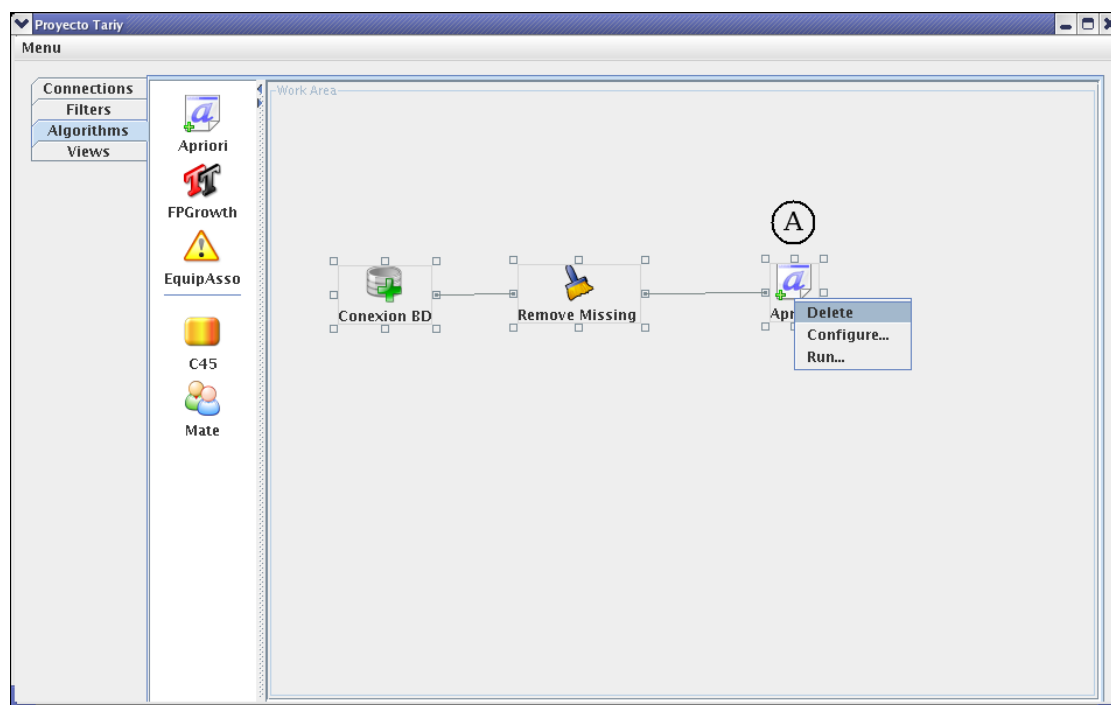


Figura 7.107: Opción Delete

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre A: el icono del algoritmo (cualquiera que este sea, Apriori, EquipAsso, FPGrowth, MateBy o C4.5) y elige la opción delete del menú de configuración.	2. El icono del algoritmo es borrado del área de trabajo.

## Opción Configure

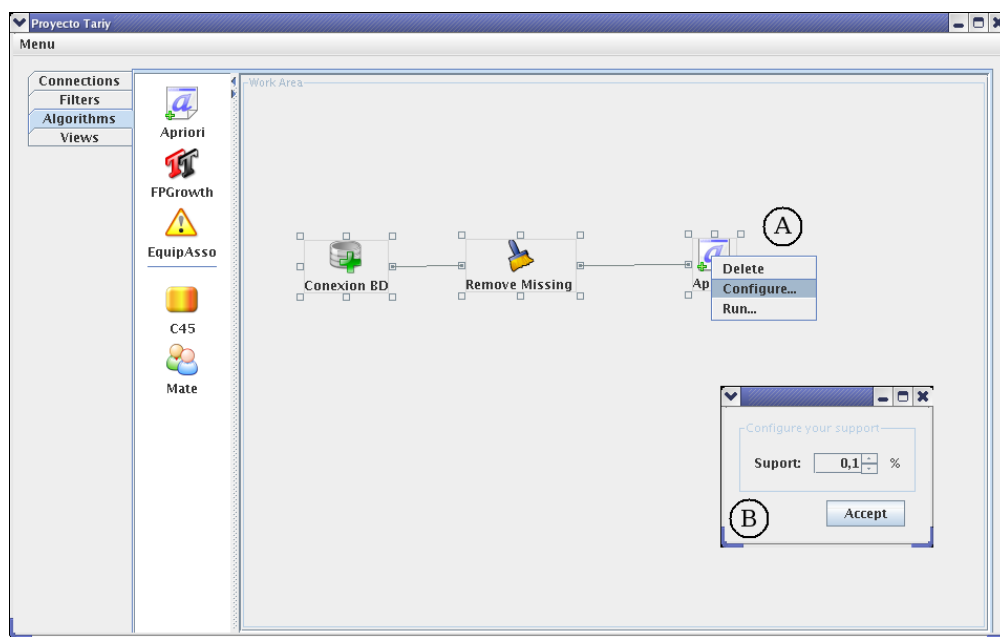


Figura 7.108: Opción Configure

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre A: el icono del algoritmo (cualquiera que este sea, Apriori, EquipAsso, FPGrowth, MateBy o C4.5) y elige configurar sus parametros.	2. Sobre el área de trabajo aparece una ventana B, para que el usuario configure el soporte del algoritmo.

## Opción Run

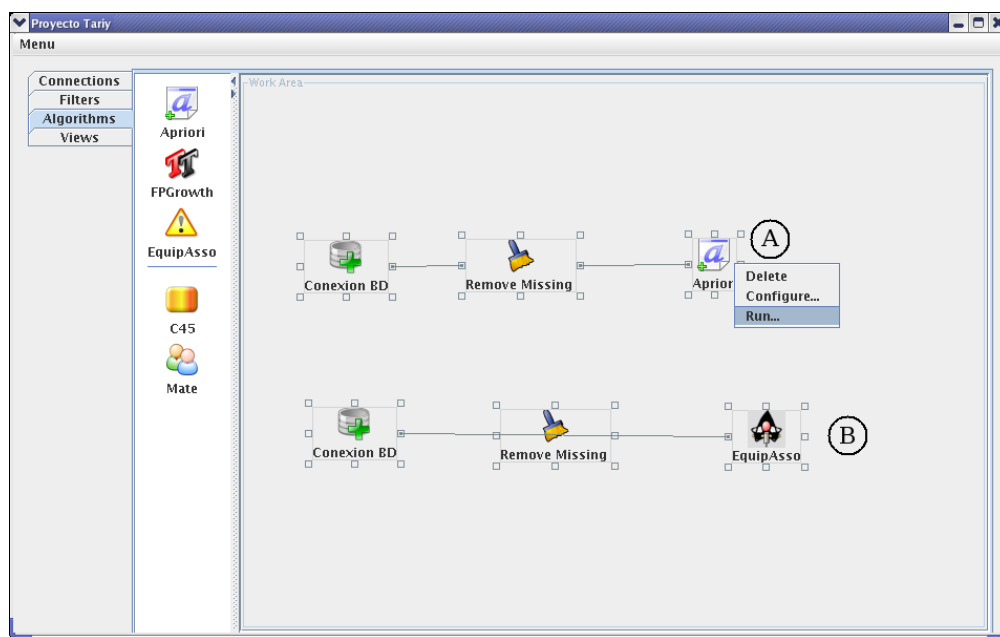


Figura 7.109: Opción Run

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre A: el icono del algoritmo (cualquiera que este sea, Apriori, EquipAsso, FPGrowth, MateBy o C4.5) y elige la opción run.	2. El icono del algoritmo cambia por una animación, así como se muestra en B.



## Algoritmo FPGrowth

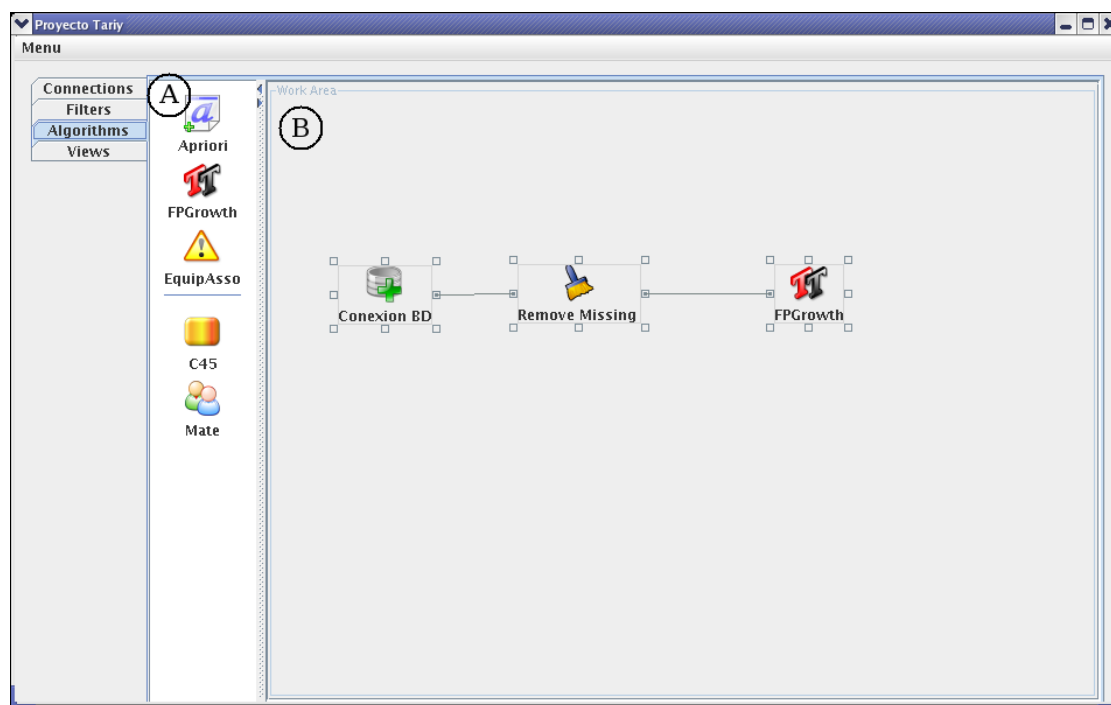


Figura 7.110: Algoritmo FPGrowth

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. Si el usuario quiere minar los datos con FPGrowth y presiona sobre A (Área de opciones), en el icono respectivo.	2. En B (Área de trabajo) aparece el icono del algoritmo FPGrowth.

## Algoritmo EquipAsso

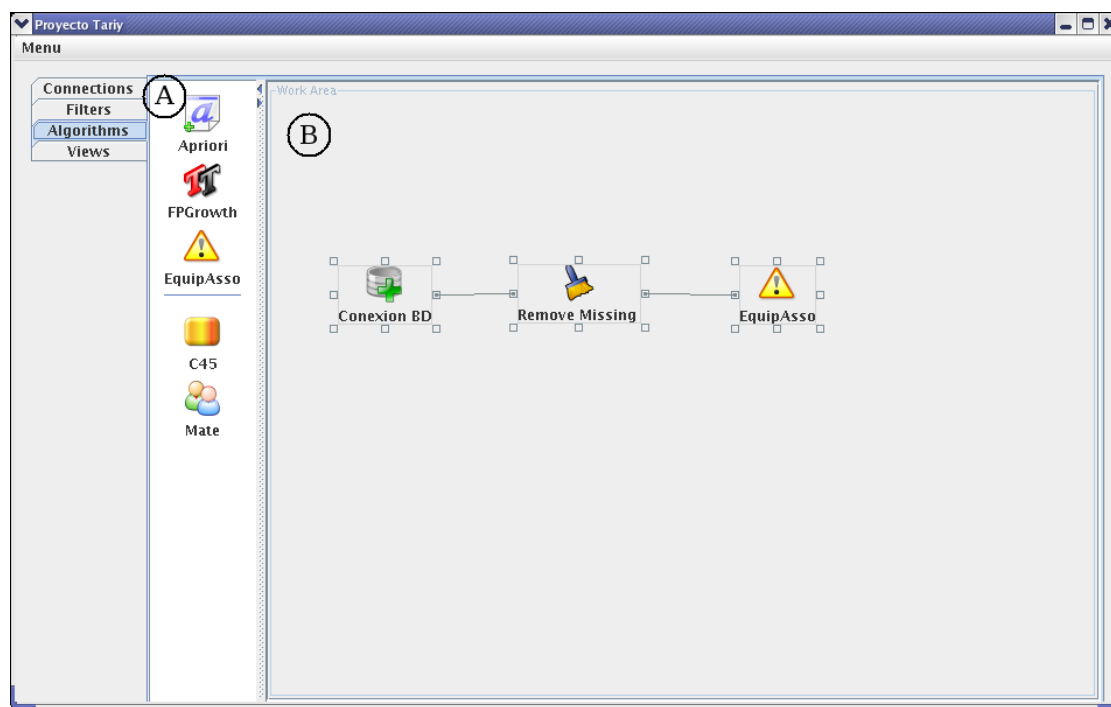


Figura 7.111: Algoritmo EquipAsso

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. Si el usuario quiere minar los datos con EquipAsso y presiona sobre A (Área de opciones), en el icono respectivo.	2. En B (Área de trabajo) aparece el icono del algoritmo EquipAsso.

## Algoritmo C4.5

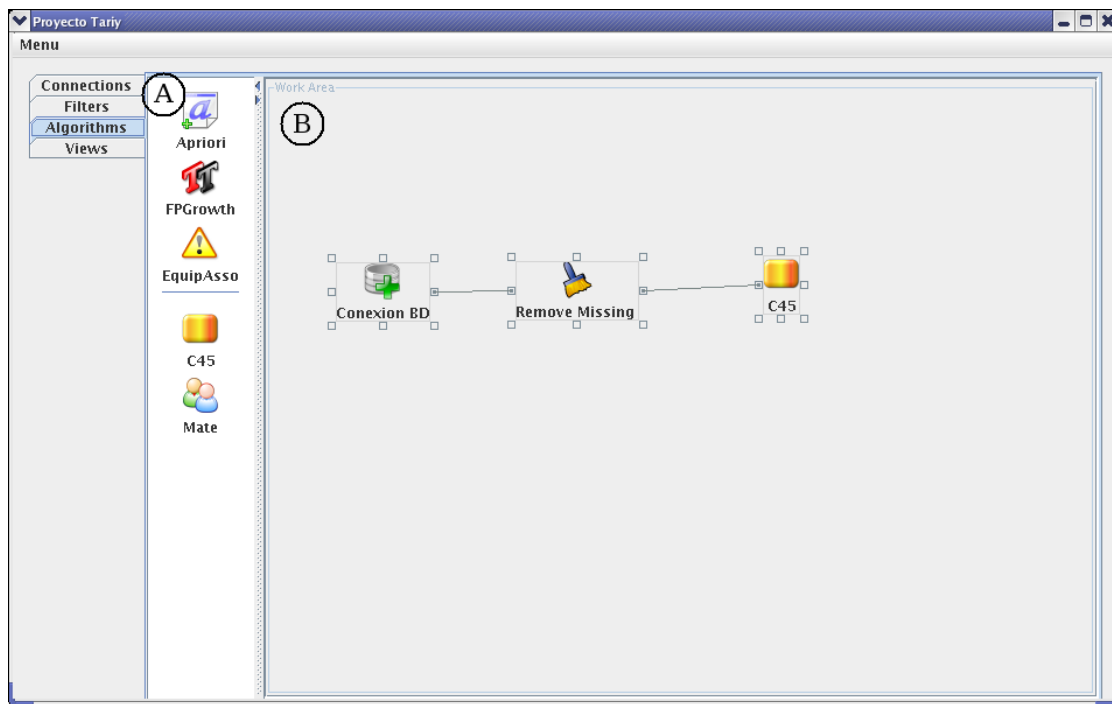


Figura 7.112: Algoritmo C4.5

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. Si el usuario quiere minar los datos con C4.5 y presiona sobre A (Área de opciones), en el icono respectivo.	2. En B (Área de trabajo) aparece el icono del algoritmo C4.5.

## Algoritmo Mate

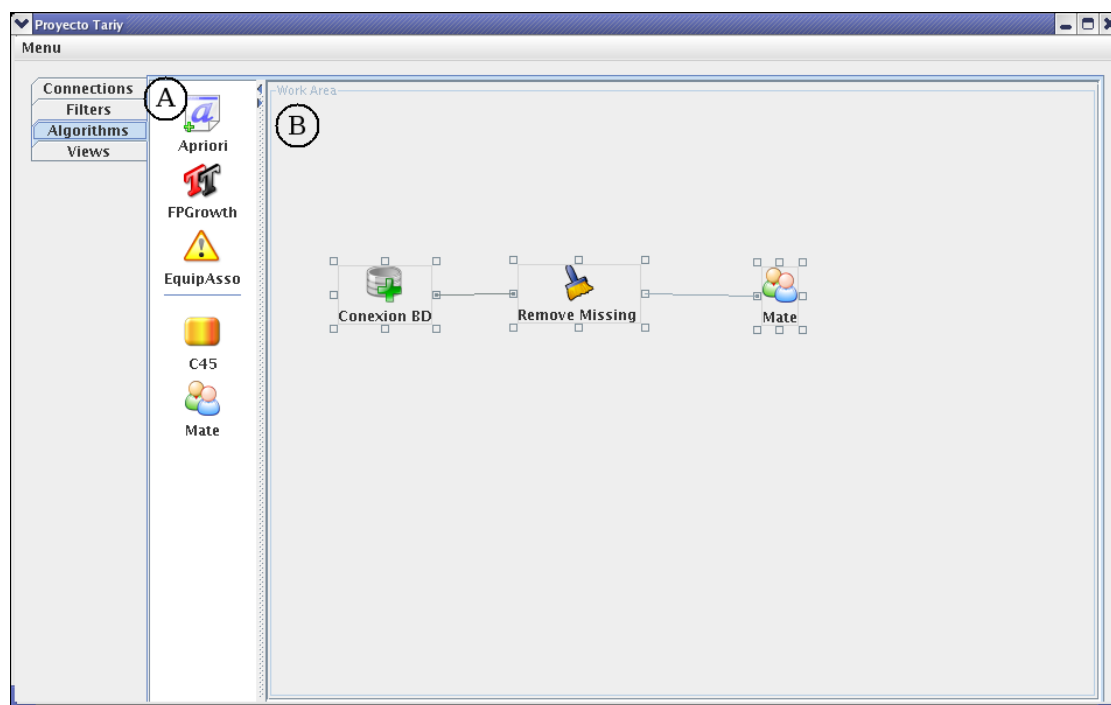


Figura 7.113: Algoritmo Mate

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. Si el usuario quiere minar los datos con Mate y presiona sobre A (Área de opciones), en el icono respectivo.	2. En B (Área de trabajo) aparece el icono del algoritmo Mate.

## Diagrama de Visualización

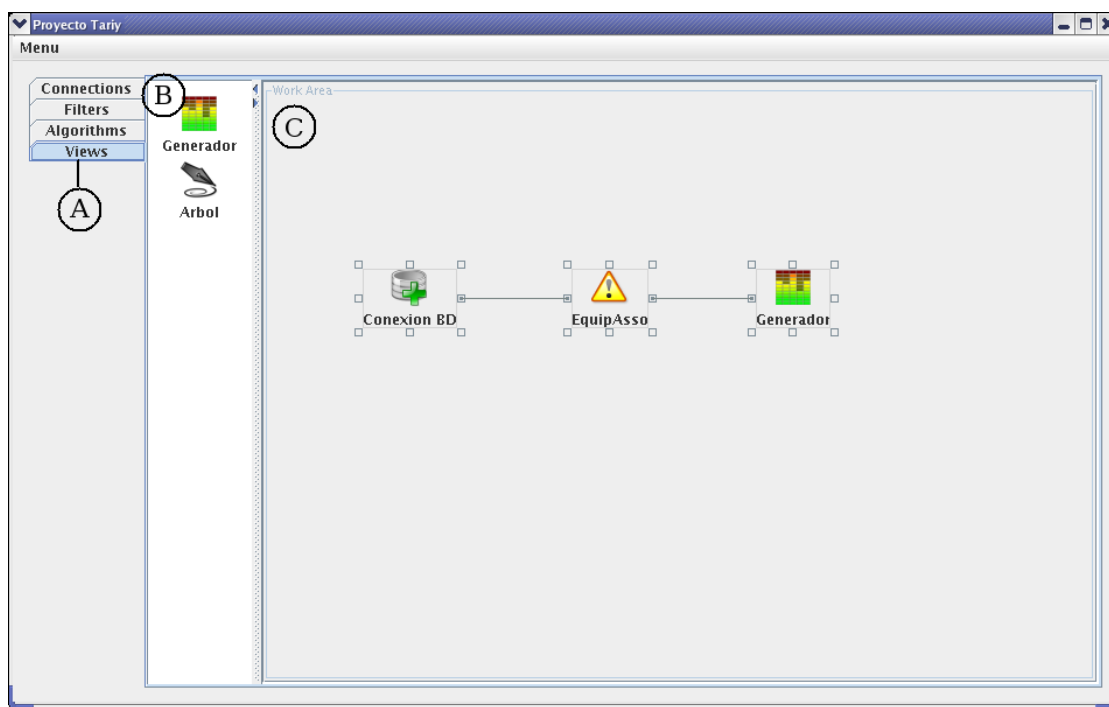


Figura 7.114: Diagrama de Visualización

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. Cuando el usuario ha construido una secuencia de Minería de Datos, con cualquiera de los algoritmos, en A se encuentra en la sección de vistas y en B (Área de opciones) ha hecho click en el icono generador.	2. Entonces en C (Área de trabajo) aparece el icono del generador, a través del cual el usuario puede acceder a las opciones de este modulo.

## Opción Delete

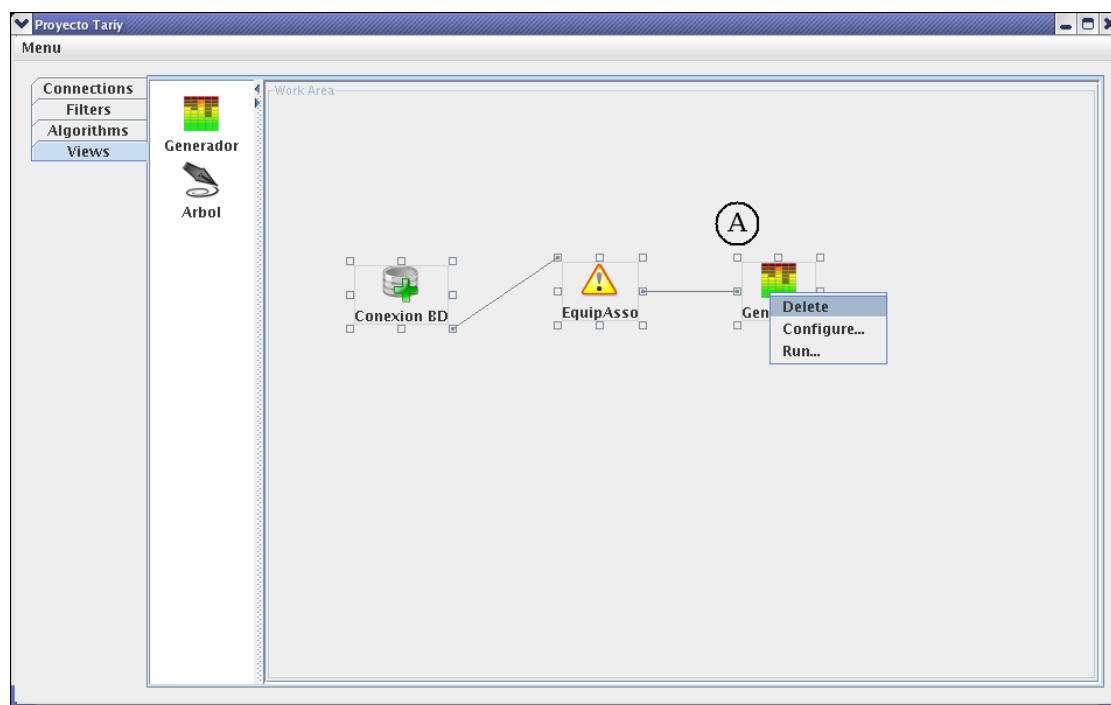


Figura 7.115: Opción Delete

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre A: el icono generador y elige la opción Delete.	2. El icono desaparece del área de trabajo, esperando un nuevo icono en la secuencia de Minería de Datos.

## Opción Configure

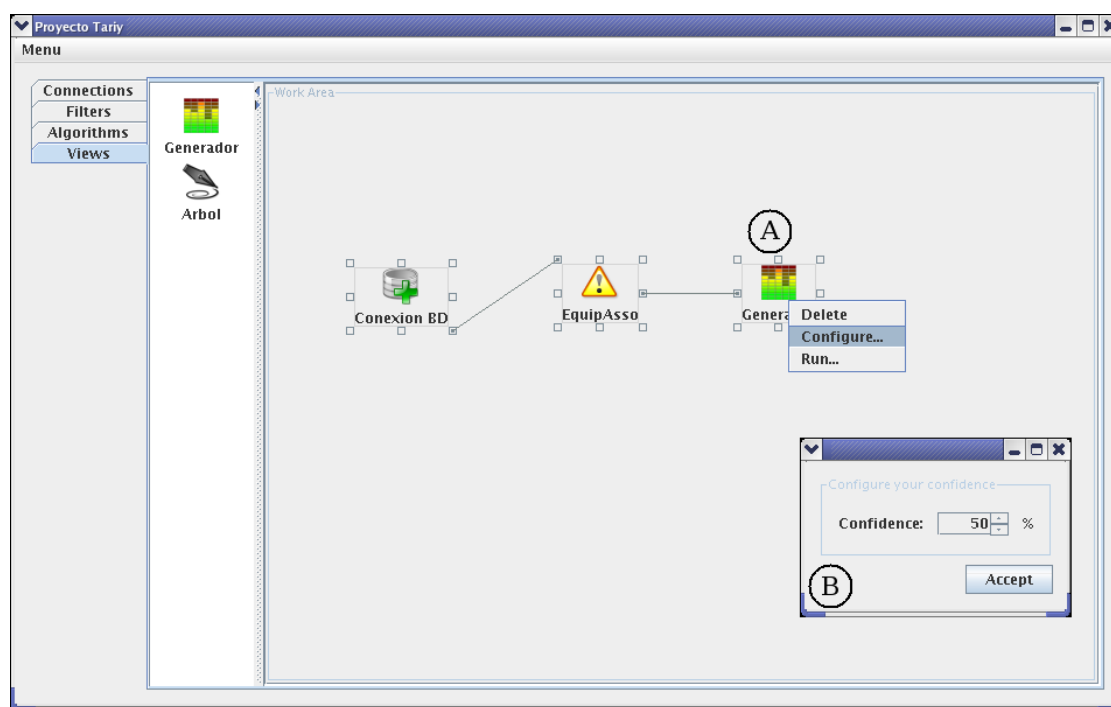


Figura 7.116: Opción Configure

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click sobre A: el icono del generador y elige configurar sus parametros.	2. Sobre el área de trabajo aparece una ventana B, para que el usuario configure la confianza con la cual se van a filtrar las reglas de asociación.

## Opción Run

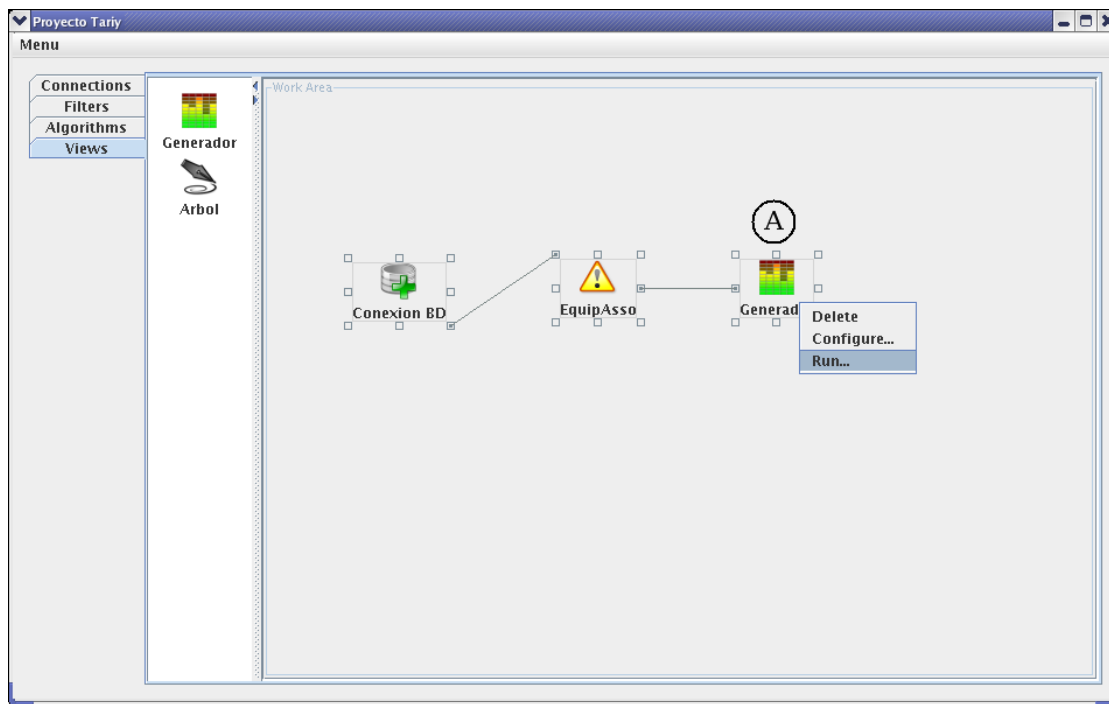


Figura 7.117: Opción Run

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario hace click derecho sobre A: el icono generador y elige la opción Run.	2. En el área de trabajo aparece una ventana B, con las reglas obtenidas a partir de los algoritmos de Minería de Datos (La cual se explica en la siguiente figura).



## Visor de Reglas

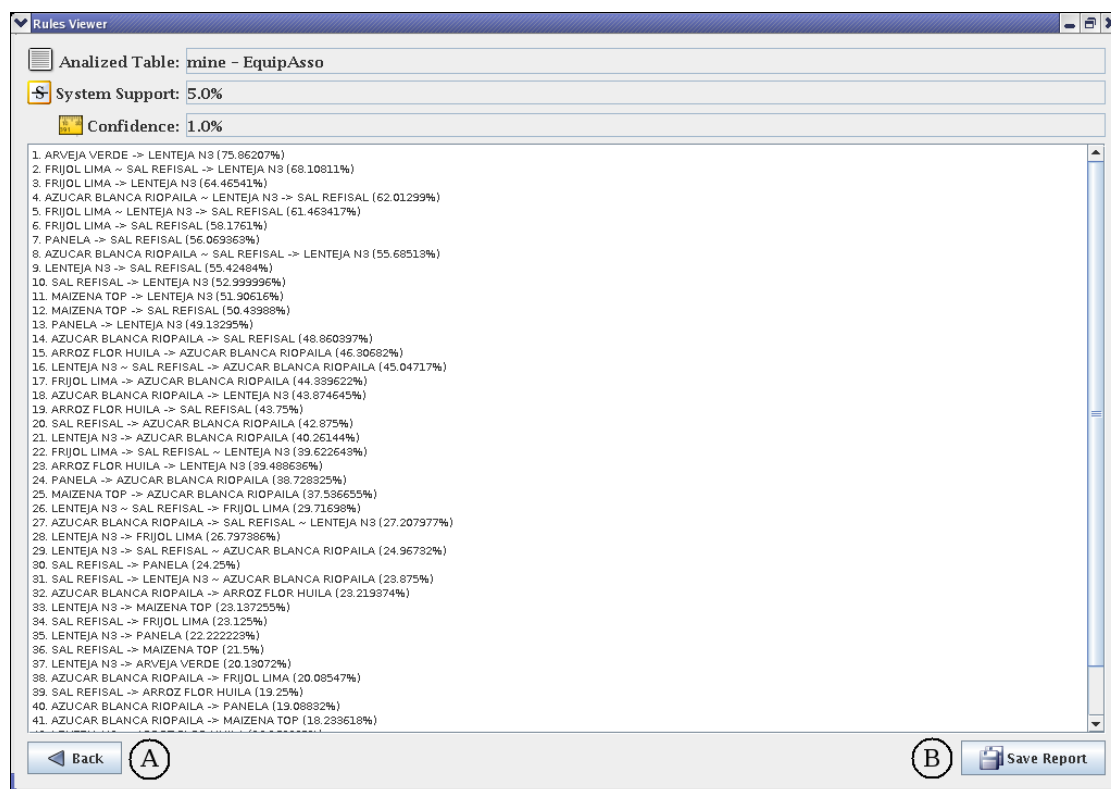


Figura 7.118: Visor de Reglas

ACCIÓN DEL ACTOR	RESPUESTA DEL SISTEMA
1. El usuario tiene la opción de hacer click en A, o en B.	2. Al hacer click en A, la ventana de reglas desaparece y si hace click en B el usuario tiene la opción de guardar el reporte de las reglas de asociación.

# Bibliografía

- [1] Université Lumière Lyon 2. Eric equipe de recherche en ingénierie des connaissances. <http://chirouble.univ-lyon2.fr>, 2006.
- [2] NASA National Aeronautics and Space Administration. Tropical cyclone windspeed indicator. <http://pm-esip.nsstc.nasa.gov/cyclone/>.
- [3] R. Agrawal, T. Imielinski, and A. Swami. *Mining Association Rules between Sets of Items in Large Databases*. ACM SIGMOD, 1993.
- [4] R. Agrawal, M. Mehta, J. Shafer, R. Srikant, A. Arning, and T. Bollinger. The quest data mining system. In *2<sup>nd</sup> Conference KDD y Data Mining*, Portland, Oregon, 1996.
- [5] R. Agrawal and K. Shim. Developing tightly-coupled data mining applications on a relational database system. In *The Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, 1996.
- [6] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *VLDB Conference*, Santiago, Chile, 1994.
- [7] R. Brachman and T. Anand. *The Process of Knowledge Discovery in Databases: A First Sketch, Workshop on Knowledge Discovery in Databases*. 1994.
- [8] S. Chaudhuri. Data mining and database systems: Where is the intersection? In *Bulletin of the Technical Committee on Data Engineering*, volume 21, Marzo 1998.
- [9] M. Chen, J. Han, and P. Yu. Data mining: An overview from database perspective. In *IEEE Transactions on Knowledge and Data Engineering*, 1996.
- [10] Quadrillion Corp. Q-yield. <http://www.quadrillion.com/qyield.shtm>, 2001.
- [11] IBM Corporation. Intelligent miner. <http://www-4.ibm.com/software/data/iminer>, 2001.

- [12] J. Demsar and B. Zupan. Orange: From experimental machine learning to interactive data mining. Technical report, Faculty of Computer and Information Science, University of Ljubljana, Slovenia, <http://www.ailab.si/orange/wp/orange.pdf>, 2004.
- [13] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery: An overview, in advances in knowledge discovery and data mining. In *AAAI Pres / The MIT Press*, 1996.
- [14] M. Goebel and L. Gruenwald. A survey of data mining and knowledge discovery software tools. In *SIGKDD Explorations*, volume 1 of 1, June 1999.
- [15] Waikato ML Group. Attribute-relation file format (arff). <http://www.cs.waikato.ac.nz/ml/weka/arff.html>.
- [16] Waikato ML Group. Collections of datasets. [http : //www.cs.waikato.ac.nz/ml/weka/index\\_datasets.html](http://www.cs.waikato.ac.nz/ml/weka/index_datasets.html).
- [17] Waikato ML Group. The waikato environment for knowledge analysis. <http://www.cs.waikato.ac.nz/ml/weka>.
- [18] J. Han, J. Chiang, S. Chee, J. Chen, Q. Chen, S. Cheng, W. Gong, M. Kamber, K. Kopferski, G. Liu, Y. Lu, N. Stefanovic, L. Winstone, B. Xia, O. Zaiane, S. Zhang, and H. Zhu. Dbminer: A system for data mining in relational databases and data warehouses. In *CASCON: Meeting of Minds*.
- [19] J. Han, Y. Fu, and S. Tang. Advances of the dblearn system for knowledge discovery in large databases. In *International Joint Conference on Artificial Intelligence IJCAI*, Montreal, Canada, 1995.
- [20] J. Han, Y. Fu, W. Wang, J. Chiang, K. Kopferski, D. Li, Y. Lu, A. Rajan, N. Stefanovic, B. Xia, and O. Zaiane. Dbminer: A system for mining knowledge in large relational databases. In *The second International Conference on Knowledge Discovery & Data Mining*.
- [21] J. Han, Y. Fu, W. Wang, J. Chiang, O. Zaiane, and K. Kopferski. *DBMiner: Interactive Mining of Multiple-Level Knowledge in Relational Databases*. ACM SIGMOD, Montreal, Canada, 1996.
- [22] J. Han and M. Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers, 2001.
- [23] J. Han and J. Pei. Mining frequent patterns by pattern-growth: Methodology and implications. In *SIGKDD Explorations*, volume 2:14-20, 2000.

- [24] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *ACM SIGMOD*, Dallas, TX, 2000.
- [25] T. Imielnski and H. Mannila. A database perspective on knowledge discovery. In *Communications of the ACM*.
- [26] RuleQuest Research Inc. C5.0. <http://www.rulequest.com>, 2001.
- [27] E-Business Technology Institute. E-business technology institute, the university of hong kong. <http://www.eti.hku.hk>, 2005.
- [28] Quinlan J.R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [29] Kdnuggets. <http://www.kdnuggets.com/software>, 2001.
- [30] C. Matheus, P. Chang, and G. Piatetsky-Shapiro. Systems for knowledge discovery in databases. In *IEEE Transactions on Knowledge and Data Engineering*, volume 5, 1993.
- [31] I Mierswa, M Wurst, R Klinkenberg, M Scholz, and T Euler. Yale: Rapid prototyping for complex data mining tasks. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06), 2006.
- [32] Faculty of Computer and Slovenia Information Science, University of Liubliana. Orange, fruitful and fun. <http://www.ailab.si/orange>, 2006.
- [33] Faculty of Computer and Slovenia Information Science, University of Liubliana. Orange's interface to mysql. <http://www.ailab.si/orange/doc/modules/orngMySQL.htm>, 2006.
- [34] Government of Hong Kong. Innovation and technology fund. <http://www.itf.gov.hk>, 2006.
- [35] Artificial Intelligence Unit of the University of Dortmund. Artificial intelligence unit of the university of dortmund. <http://www-ai.cs.uni-dortmund.de>, 2006.
- [36] G. Piatetsky-Shapiro, R. Brachman, and T. Khabaza. *An Overview of Issues in Developing Industrial Data Mining and Knowledge Discovery Applications*. 1996.
- [37] J.R. Quinlan. *Induction of decision trees. Machine Learning*. 1986.
- [38] R Rakotomalala. Tanagra. In *TANAGRA: a free software for research and academic purposes*, volume 2, pages 697–702. EGC'2005, 2005.

- [39] R. Rakotomalala. Tanagra project. <http://chirouble.univ-lyon2.fr/rico/tanagra/en/tanagra.html>, 2006.
- [40] Isoft S.A. Alice. [http://www.alice-soft.com/html/prod\\_alice.htm](http://www.alice-soft.com/html/prod_alice.htm), 2001.
- [41] S. Sarawagi, S. Thomas, and R. Agrawal. Integrating association rule mining with relational database systems: Alternatives and implications. In *ACM SIGMOD*, 1998.
- [42] SPSS. Clementine. <http://www.spss.com/clementine>, 2001.
- [43] Information Technology The University of Alabama in Huntsville and Systems Center. Adam 4.0.2 components. <http://datamining.itsc.uah.edu/adam/documentation.html>.
- [44] Information Technology The University of Alabama in Huntsville and Systems Center. Algorithm development and mining system. <http://datamining.itsc.uah.edu/adam/index.html>.
- [45] R. Timaran. Arquitecturas de integracion del proceso de descubrimiento de conocimiento con sistemas de gestion de bases de datos: un estado del arte, en revista ingeniera y competitividad. *Revista de Ingeniera y Competitividad, Universidad del Valle*, 3(2), Diciembre 2001.
- [46] R. Timaran. Descubrimiento de conocimiento en bases de datos: Una vision general. In *Primer Congreso Nacional de Investigacion y Tecnologa en Ingeniera de Sistemas*, Universidad del Quindo, Armenia, Octubre 2002.
- [47] R. Timaran. *Nuevas Primitivas SQL para el Descubrimiento de Conocimiento en Arquitecturas Fuertemente Acopladas con un Sistema Gestor de Bases de Datos*. PhD thesis, Universidad del Valle, 2005.
- [48] R. Timaran and M. Millan. Equipasso: an algorithm based on new relational algebraic operators for association rules discovery. In *Fourth IASTED International Conference on Computational Intelligence*, Calgary, Alberta, Canada, July 2005.
- [49] R. Timaran and M. Millan. Equipasso: un algoritmo para el descubrimiento de reglas de asociacion basado en operadores algebraicos. In *4a Conferencia Iberoamericana en Sistemas, Cibernetica e Informatica CICI 2005*, Orlando, Florida, EE.UU., Julio 2005.

- [50] R. Timaran, M. Millan, and F. Machuca. New algebraic operators and sql primitives for mining association rules. In *IASTED International Conference Neural Networks and Computational Intelligence*, Cancun, Mexico, 2003.
- [51] I. Waitten and F. Eibe. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. 2001.
- [52] YALE. Yale - yet another learning environment. <http://rapid-i.com>, 2006.