



VII Jornadas Iberoamericanas de Ingeniería de Software
e Ingeniería del Conocimiento
2008



TariyKDD: una Herramienta de Minería de Datos Débilmente Acoplada con un SGDB

**Grupo de Investigacion GRIAS - Línea KDD
Facultad de Ingeniería
Departamento de Ingeniería de Sistemas
Universidad de Nariño
Pasto - Colombia**



Autores



Ricardo Timarán Pereira, Ph.D.
Ing. Andrés O. Calderón Romero
Ing. Iván Ramírez Freyre
Ing. Fernando Guevara
Ing. Juan Carlos Alvarado

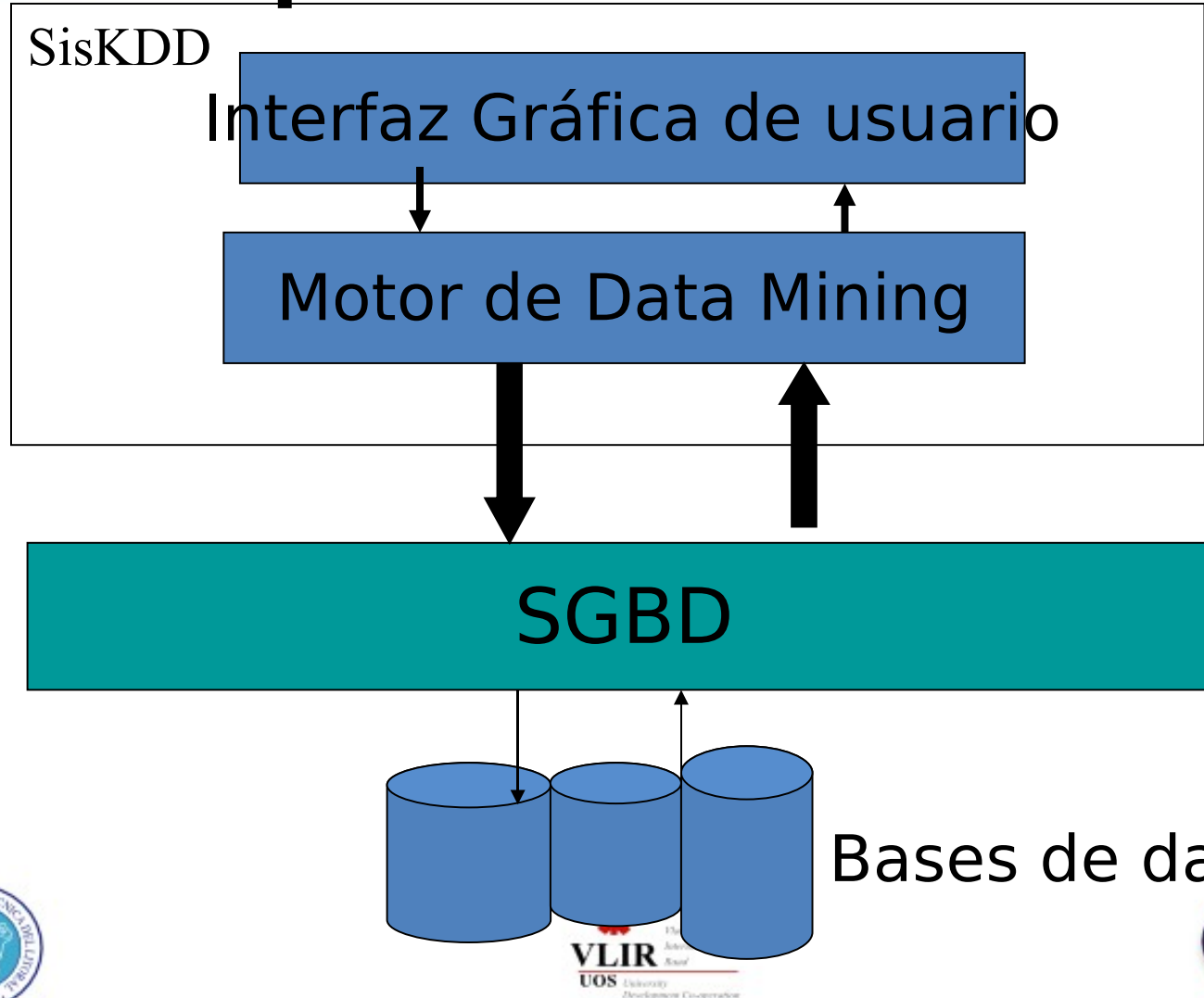


Agenda

- Arquitectura débilmente acoplada
- Arquitectura de TariyKDD
- Aspectos de diseño e implementación de TariyKDD
- Pruebas de funcionalidad
- Conclusiones y trabajos futuros



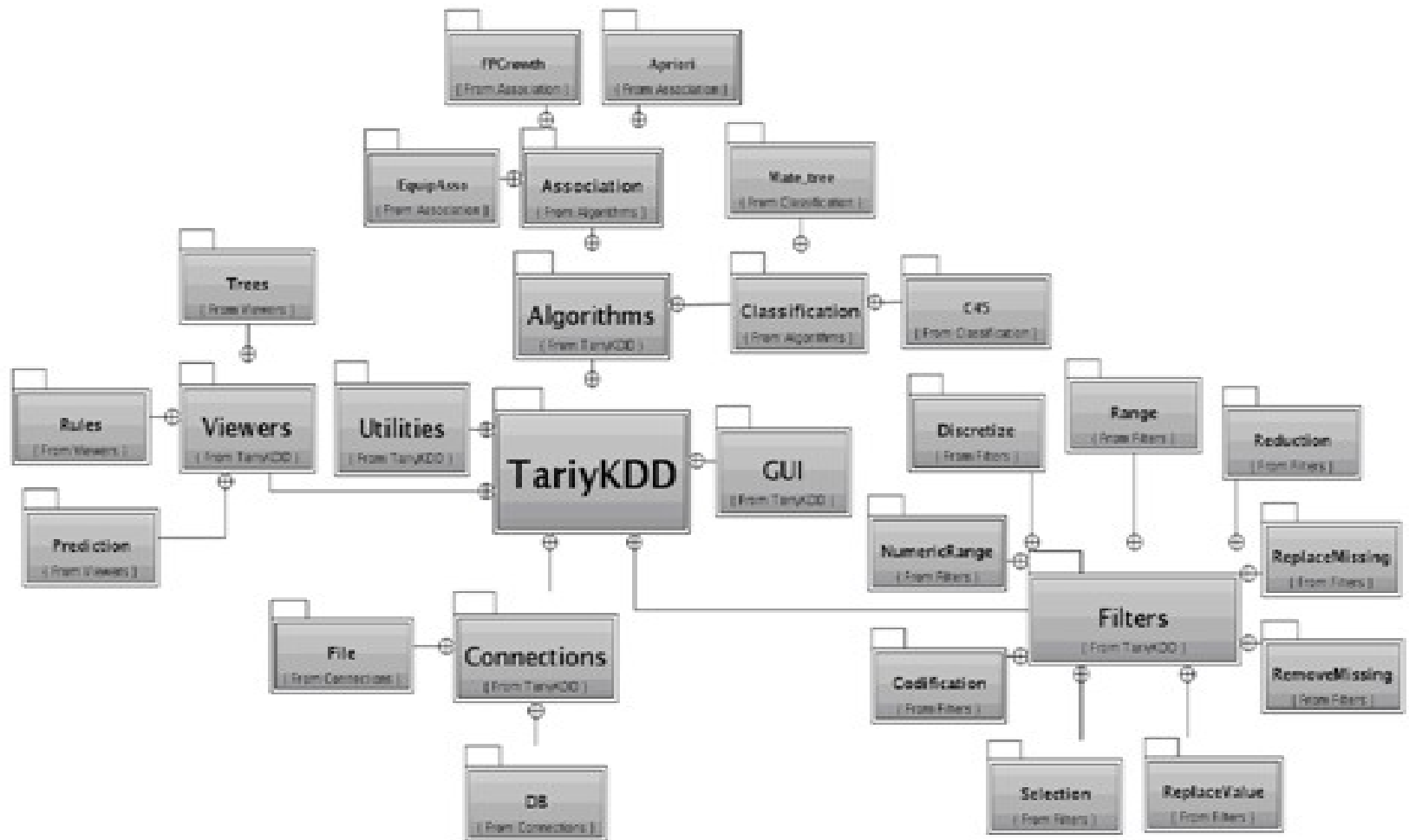
Arquitectura de un sistema de minería de datos débilmente acoplada con un SGBD



Arquitectura de TariyKDD



Diseño de TariyKDD



Aspectos de implementación

Software libre

- Sistema operativo Fedora Core
- Java 5.0
- Netbeans 5.5
- Subversion
- Gimp y KIconEdit
- PostgreSQL 8.1



Algoritmos implementados

- Tarea de Asociación:
 - A priori
 - Fp-Grow
 - EquipAsso
- Tarea de Clasificación
 - C4.5
 - Mate-tree



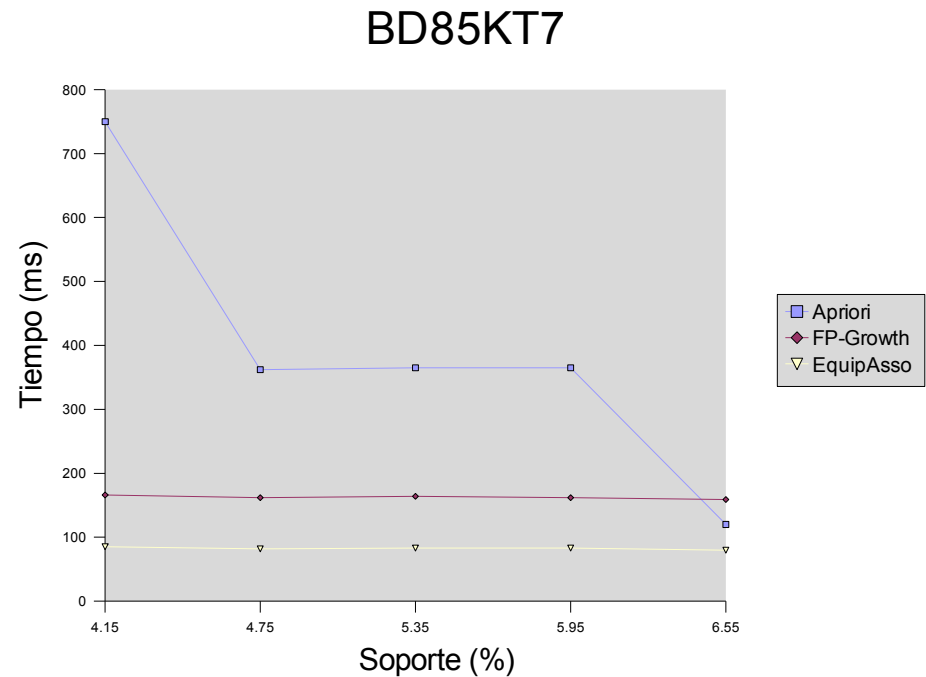
Pruebas y evaluación de rendimiento

- Equipo Utilizado
 - Procesador AMD 64 bits a 2Ghz
 - 1 Gb de RAM
 - DD ATA de 80GB a 150 Mb/sg
- Repositorios asociación
 - BD85KT7 con 555.123 registros, 85.692 transacciones y promedio de 7 items por transacción
 - BD40KT5 con 197.337 registros, 40.256 transacciones y 5 items por transacción
- Repositorios clasificación
 - UDENAR.dat con 20.328 registros y 22 atributos.

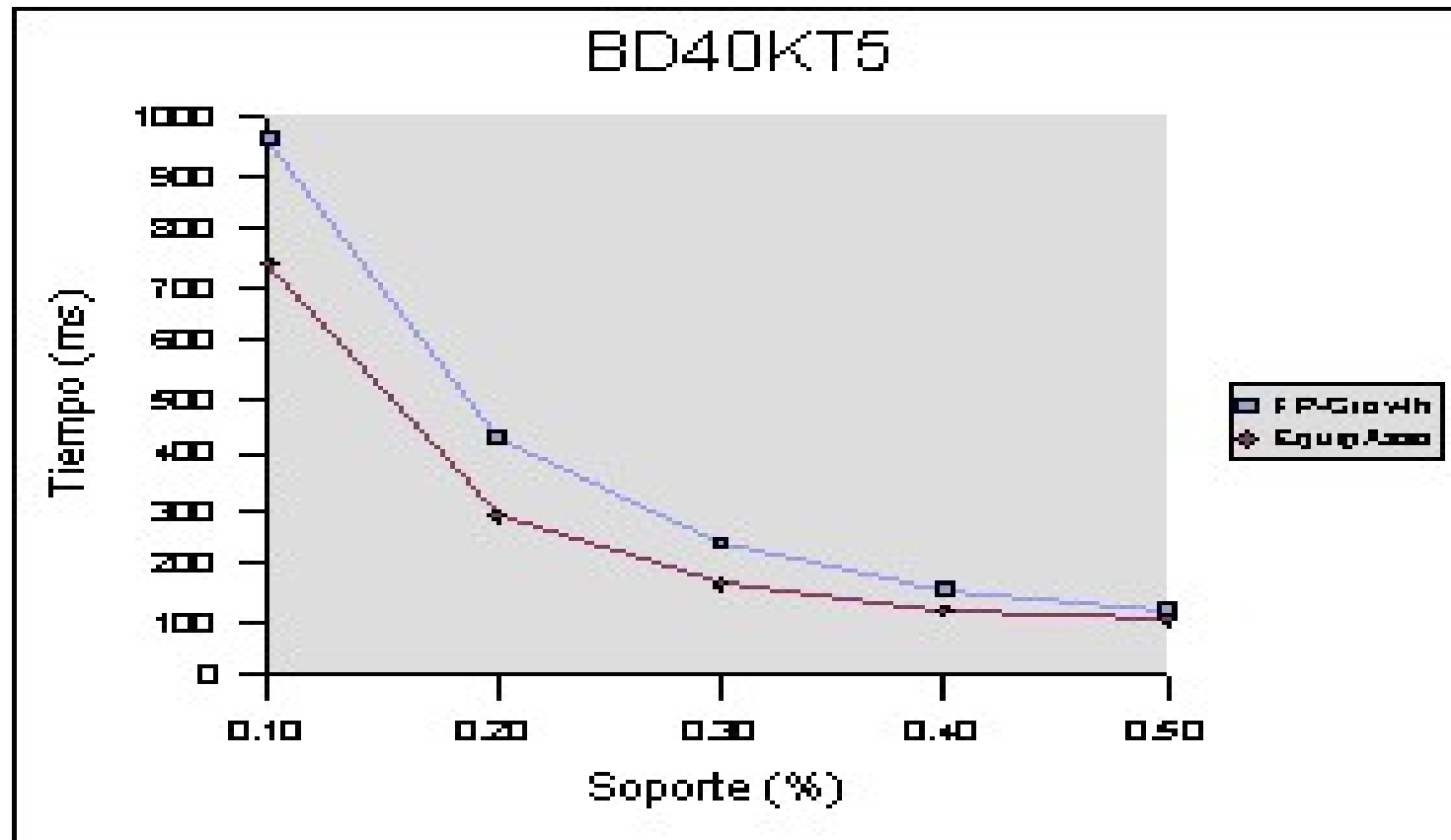


Resultados para BD85KT7

BD85KT7			
Soporte (%)	Tiempo (ms)		
	A priori	FP-Growth	EquipAsso
4.15	750	166	85
4.75	362	162	82
5.35	365	164	83
5.95	365	162	83
6.55	120	159	80

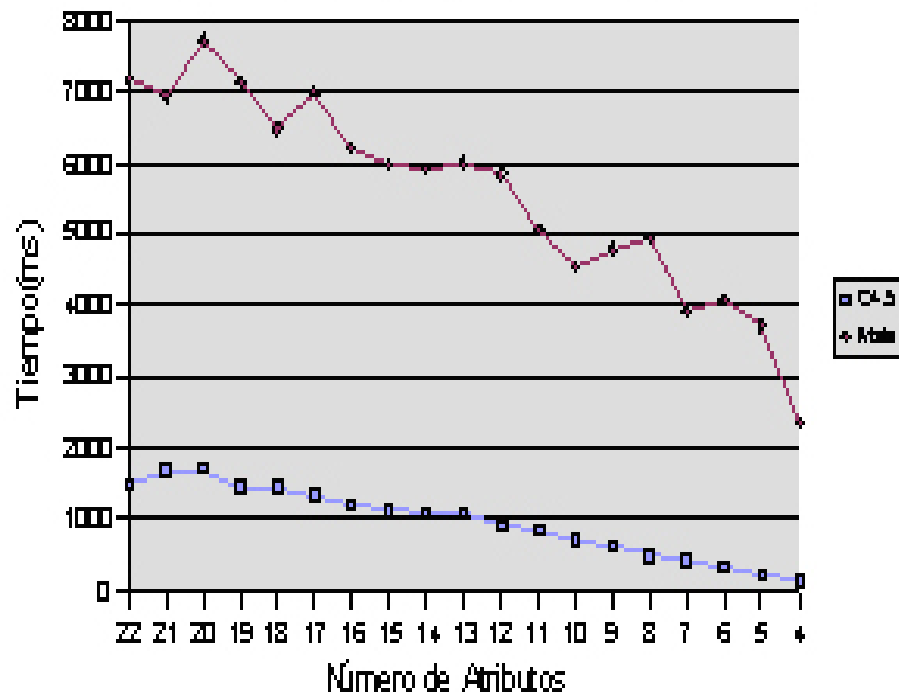


Resultados para BD40KT5

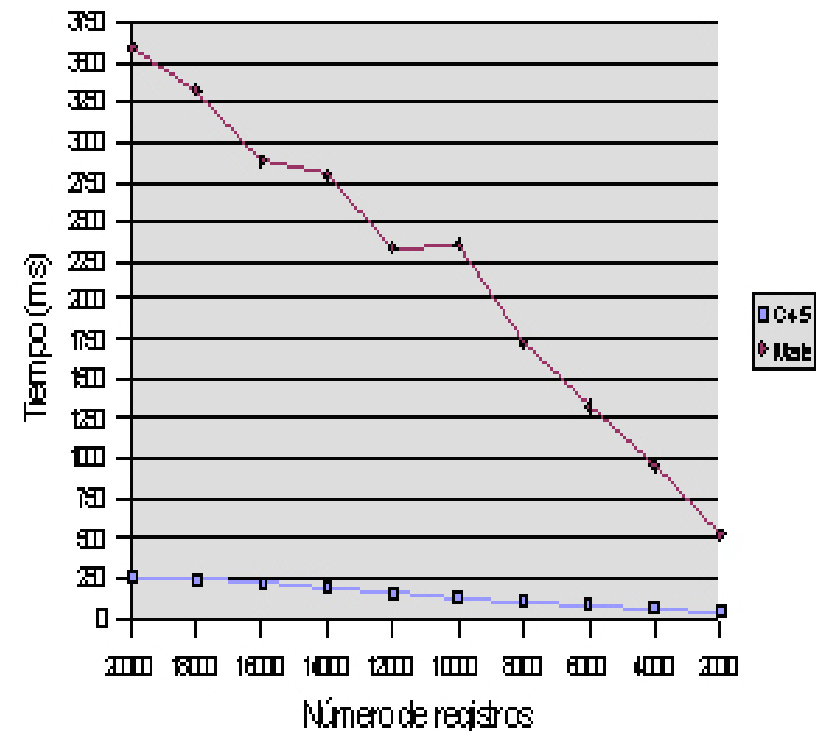


Resultados para Udenar.dat

Reduciendo número de atributos



Reduciendo número de registros



Conclusiones

- TariyKDD es una herramienta para el descubrimiento de reglas de asociación y clasificación.
- Soporta las etapas del proceso KDD: selección, limpieza, transformación, minería de datos y visualización.
- Las diferentes pruebas de funcionalidad realizadas demuestran que TariyKDD es fiable.
- TariyKDD es una herramienta de software libre, bajo licencia GPL v2.0.
- Disponible en <http://tariykdd.berlios.de>



Trabajos futuros

- Implementar otras tareas y algoritmos de minería de datos en TariyKDD.
- Implementar gráficos estadísticos a los conjuntos de datos cargados para obtener una información inicial de sus características.
- Utilizar TariyKDD en proyectos reales de minería de datos con empresas colombianas, especialmente con las PYMES.



Gracias



Ricardo Timarán Pereira, Ph.D.

Ing. Andrés O. Calderón Romero, Ing. Iván Ramírez Freyre,
Ing. Fernando Guevara, Ing. Juan Carlos Alvarado

Grupo de Investigacion GRIAS - Línea KDD
Facultad de Ingeniería
Departamento de Ingeniería de Sistemas
Universidad de Nariño
Pasto- Colombia

