

Prometeo: Una Herramienta Para El Aprovechamiento De Metadatos De Base De Datos Relacionales

Marciszack, Marcelo
Maldonado, Calixto
Martinez Spessot, Cesar
Muñoz, Roberto
Navarro, Adrián
Peretti, Juan Pablo
Roggero, Luis

**Laboratorio de Investigación de Software/Dep. Ing. en Sistemas de Información/
Facultad Regional Córdoba
Universidad Tecnológica Nacional**

CONTEXTO

El presente trabajo es un informe de avance y describe estado del proyecto Prometeo según los logros alcanzados en el año 2008 por los integrantes del proyecto, iniciado en el año 2004 y que ha sido distinguido con la acreditación de las Secretarías de Ciencia y Técnica de la UTN en enero del 2009. Es un proyecto relacionado con TecnoDB, que es un motor de Base de datos relacional para ser ejecutado en dispositivos móviles, proyecto acreditado en sus dos fases, la última con finalización en el presente año. El objetivo del proyecto Prometeo es definir una heurística de explotación de meta datos y un software capaz de implementar esta metodología. Con este nombre fue presentado y aprobado en el programa de Grupos de Reciente Formación de la Agencia Córdoba Ciencia en el año 2007.

RESUMEN

Se presenta una síntesis del proyecto Prometeo y los avances logrados durante el año 2008. Se enuncian los antecedentes, características de la heurística usada y los objetivos del software, logros recientes y avances previstos en este año.

Palabras clave: *SQL, Generador de Consultas, QueryByExample, SQL autogenerado*

1. INTRODUCCIÓN

En presentaciones en WICC2005 [GAS05] y 35JAIIO [GAS06] se describía el proyecto Prometeo como “un sistema de software que se ejecute en múltiples plataformas de sistemas operativos, en principio, Windows 2000 y Red Hat Linux 7.1 y que sea capaz de generar sentencias SQL basados en los metadatos de los diccionarios de datos de Oracle y PostgreSQL: Consultas simples con SELECT, Consultas multi tablas de las tablas relacionadas con Constraints Referenciales, DML para insertar, borrar y actualizar columnas, DDL para creación de objetos

como vistas y tablas accesorias, modificación y borrado de objetos como Vistas, OQL o lenguaje de consulta de Objetos cuando la base de datos estudiada contenga esa opción”

Hubo cambios de integrantes por el egreso de los desarrolladores originales y por el ingreso de nuevos integrantes, provenientes del alumnado de la carrera de Ingeniería en Sistemas de Información.

El proyecto, que responde al nombre extenso de “Desarrollo de un método y una herramienta para el aprovechamiento de Metadatos de Base de Datos Relacionales” fue acreditado este año por la SeCyt del UTN, y ha logrado sólidos avances. El nombre extenso es habitualmente reemplazado por el nombre abreviado de Prometeo, que según Wikipedia [WIK09] “es el Titán amigo de los mortales, honrado principalmente por robar el fuego de los dioses en el tallo de una cañaheja, darlo a los humanos para su uso y ser castigado por este motivo.” La metáfora del nombre es por que este software ayuda a conocer y aprovechar la información almacenada en los metadatos y con la ayuda de unos criterios o Heurística permite construir consultas, ejecutarlas y guardarlas en tablas propias.

La propuesta inicial fue presentada en un paper que fue aceptado en las 34 JAIIO desarrolladas en Rosario en el mes de Septiembre del 2005, en el simposio ASIS [GAS05]

El hecho que permite obtener provecho de los metadatos es la capacidad de la sentencia SELECT de poder combinar el resultado de seleccionar el contenido de las columnas de las vistas del Diccionario con literales para poder escribir sentencias SELECT validas. Ejemplo:

Dada la vista del diccionario de datos Oracle [CON05] USER_TABLES, que contiene todas las

tablas del usuario conectado, se podría generar una sentencia select de todas las columnas de las tablas existentes con la siguiente sentencia:

```
SELECT 'select * from '||table_name||';'
FROM USER_TABLES;
```

La ejecución de esta consulta generará en la salida estándar de la herramienta usada:

```
select * from tabla1;
select * from tabla2;
```

Una sentencia SELECT por cada registro consultado de la vista del diccionario USER_TABLES.

Agregándole una línea de texto podríamos crear una sentencia SQL para crear una vista que guarde esta consulta como un objeto de la base. Ejemplo:

```
SELECT 'create or replace view '||table_name ||'_v
as select * from '||table_name||';'
FROM USER_TABLES;
```

Lograríamos esta sentencia valida SQL, basada en todas las columnas de la tabla 'PEPE' perteneciente al usuario actual:

```
create or replace view as PEPE_v as select * from
PEPE ;
```

Este funcionamiento del lenguaje SQL es la base del generador de sentencias SQL, basadas en los metadatos.

Sobre la Heurística tiene como objetivo el aprovechamiento de los metadatos residentes en motores de bases de datos comerciales como Postgresql y Oracle y en desarrollo, como TecnoDB. Este método o heurística guía a la generación de las sentencias útiles dentro de todas las combinaciones posibles y contiene 21 puntos o criterios a seguir y los resultados a obtener.

Logrando esto se espera poder generar sentencias SQL con algun grado de optimización [Cis02] que permitan ahorrar trabajo y errores comunes a los desarrolladores de aplicaciones en estas bases de datos. Estamos buscando mejorar y completar el método que viene desarrollandose desde su inicio y aumentar las funcionalidades de su interfaz, para que pueda definir qué consultas se deben generar para que asista al desarrollador a construir reportes, generar conjuntos de filas de pruebas de integridad y desarrollar una herramienta que la aplique par lograr que con el solo hecho de ejecutarla, obtener un subconjunto de sentencias útiles sin necesidad de conocer profundamente el modelo de datos objetivo. Una de las formas de mejora en el proyecto es diseñar el método para generar Java Beans, es decir,

extender la accion a construir codigo basado en JAVA, generando codigo reusable [WIK09b] [SUN09]

Prometeo tiene como objetivo también aportar al proyecto TecnoDB sirviendo como aporte al diseño del diccionario de datos, aún no finalizado en el desarrollo del motor de Bases de datos relacional.

Las sentencias que genera Prometeo son inicialmente SELECT, y CREATE VIEW. Se completará con las sentencias INSERT y CREATE TABLE, pudiendo aumentar esta capacidad a otras sentencias mas. Las sentencias INSERT van ser generadas para insertar conjuntos de datos de prueba de las tablas elegidas, respetando las restricciones de integridad referencial detectadas en el modelo[EMS08a] [EMS08b] [Dat97].

En imágenes una explicación de cómo este software, genera, muestra, ejecuta, muestra los resultados y grababa en sus propias tablas, las sentencias SQL obtenidas con las indicaciones del usuario. En la figura 1 se muestra un esquema del funcionamiento.



Fig. 1 Pasos en la generación de Sentencias

En la figura 2 el modelo de datos inicial, modificado durante 2008 al requerir almacenar mas información.

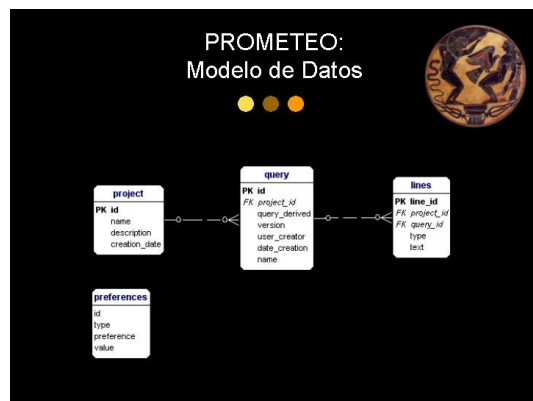


Fig. 2

La interfaz se conecta al diccionario de datos de la base de datos objetivo y guarda en la base de datos repositorio (que puede ser la misma) la información acerca de los objetos analizados en tablas propias. Permite ejecutar las consultas generadas o almacenadas y mostrar el resultado en la pantalla, para luego guardarla en el repositorio. Los lenguajes a utilizar en la construcción son PHP y JAVA para la interfaz y procedimientos y SQL para la comunicación con las bases de datos objetivo. Se realizó la comunicación con PostgreSQL y se está trabajando para realizar la conexión con MySQL.

También se avanzará en que la interfaz pueda recopilar información de documentación con los contenidos de Comentarios sobre tablas y columnas posibles de ser guardados en las vistas del diccionario de las bases de datos objetivo, por ejemplo, en Oracle [CON05] la vista ALL_TAB_COMMENTS y la ALL_COL_COMMENTS permitirán mejorar la documentación de los desarrollos que utilicen a Prometeo como herramienta.

Se están definiendo los casos de usos correspondientes a cada punto enunciado para poder diseñar la aplicación que lo resuelva y que sirva como base al proceso de testing [COC00] [LAR00] [JAC00].

2. LINEAS DE INVESTIGACION y DESARROLLO

Desarrollo de avances en la Heurística [SPE05] para el uso de los Metadatos en los diccionarios de las Bases de Datos Oracle, PostgreSQL, MySQL. [TOA08]

Diseño del modelo de datos de Metadatos para TecnoDB.

Diseño de Vistas para las distintas versiones de RDBMS usadas para lograr independencia en las distintas instalaciones.

Creación de Java Beans con la información de los Metadatos para construir automáticamente métodos de acceso, actualización a los datos de las tablas analizadas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Hasta el 2007 se programó un sistema web, desarrollado en PHP 5 y que obtuviera información de los metadatos de PostgreSQL [GAS05].

Durante el 2008, se trabajó sobre MySQL lográndose lo mismo que hasta el 2007 y además se pudo construir sentencias SELECT multitablas, es decir relacionaba tablas con constraints de clave foránea para escribir los Joins sobre las mismas tablas. Esto estaba previsto en la Heurística original pero no se había concretado aún [SPE07] [ALT08].

Durante el trabajo con MySQL se observó que el modelo de Datos objetivo de los metadatos de MySQL son manejados por el motor a través de la base de datos INFORMATION_SCHEMA, la misma está caracterizada por no contener tablas, sino que la información es manejada a través de vistas. En este Diagrama de Entidad-Relación se pueden observar las vistas más importantes (sobre todo las utilizadas por Prometeo) en forma de tablas y relaciones entre las mismas. [MYS08]

Se modificó esta pantalla con la idea de brindar la posibilidad al usuario de seleccionar el tipo de consulta a realizar. Se vio que sería importante lograr la pantalla de inicio debería mostrar una leyenda de bienvenida al usuario y darle la posibilidad de seleccionar una tarea entre varias, como por ejemplo, generar consultas, seleccionar consultas guardadas de acuerdo a un proyecto particular y ejecutar dicha consulta y/o ver todas las consultas guardadas junto a los proyectos a las cuales pertenecen. (Fig. 3)

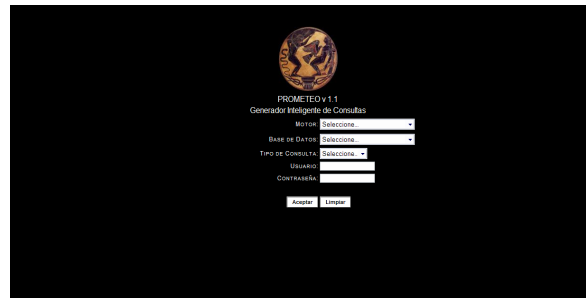


Fig 3 Pantalla de Inicio de la V1.1

La siguiente pantalla (Fig. 4) ofrece dos criterios para lograr los emparejamientos de columnas para las consultas multitablas, pero solo soporta uno solo de los dos criterios (Tablas referenciadas por esta tabla). Es decir, la funcionalidad está implementada para ese criterio solamente, por lo que debería desarrollarse la funcionalidad para el criterio restante.

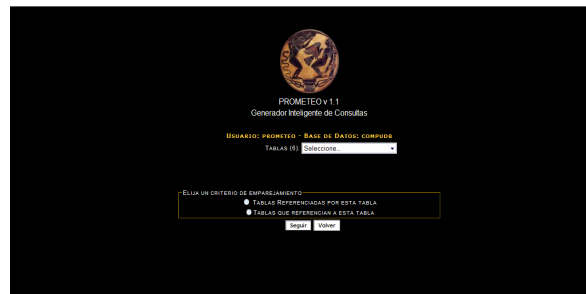


Fig. 4 Pantalla para generar consultas multitablas.

En la siguiente imagen (Fig 3) se muestra página con la consulta generada a partir de los datos seleccionados anteriormente. La funcionalidad de

los botones “Guardar Consulta” y “Ejecutar Consulta” esta en desarrollo.



Fig 3. Resultado de la Generación de consultas multitabla.

Son objetivos del trabajo durante este año:

- Extender la aplicación a otros motores comerciales como SQL Server y DB2 [DB208].
- Avanzar con nuevos productos del análisis de los metadatos como generar automáticamente Java Beans que contengan los métodos get y set de una tabla. [ISQ,08]
- Migrar la programación de la aplicación de PHP a Java. [QUE08]
- Generar en las tablas estudiadas set de datos de prueba de acuerdo a los criterios que se podran registrar como preferencias en la herramienta.

También se ha previsto para la próxima etapa del proyecto la inclusión de elementos de diseño, tales como estilos web para lograr una interfaz amigable con el usuario y desarrollar un asistente o Wizard para tener un punto de inicio y donde el usuario pueda seleccionar la/s tarea/s que interesan realizar y que se vayan desarrollando en la secuencia correcta.

Cabe señalar que del desarrollo anterior, solo se hizo una adaptación de lo existente para ser utilizado con MySQL, no se desarrolló de nuevo la funcionalidad para consultas simples, y que deben completarse ciertos aspectos que son heredados del desarrollo anterior.

4. FORMACION DE RECURSOS HUMANOS

Los participantes son tres alumnos de la UTN-FRC que están cursando los últimos años de la carrera de Ingeniería y su trabajo ha sido considerado dentro del ámbito de la materia Práctica Profesional Supervisada o PPS, durante el presente año.

Durante el año 2008 estuvieron trabajando como Becarios de la Secretaria de Alumnos y Extensión al desarrollar la tarea descripta.

Se hará transferencia de los aprendizajes adquiridos en algoritmos de mejora de performance, almacenamiento y procesamiento de consulta de datos a las cátedras de la Universidad Tecnológica Nacional, especialmente a Sistemas de Soporte de Decisión (5to año de la carrera de Ingeniería en

Sistemas de Información) donde se tratan temas avanzados de base de datos.

El trabajo esta relacionado con el grupo de desarrollo de TecnoDB de la UTN, con el objetivo de que Prometeo sea una herramienta de TecnoDB para mejorar sus prestaciones al contar con una interfaz amigable para realizar consultas SQL.

5. BIBLIOGRAFIA

[ALT08] www.Altova.com/DatabaseSpy visitado Mayo2008

[BUR99] Burgos M. (1999) “Tutorial de SQL”, <http://members.tripod.com/~MoisesRBB/sql.html>
Fecha de consulta: mayo 2005

[CAV00] Cavadini S. (2000) “Compiladores e Intérpretes”, <http://www.ucse.edu.ar/fina/compiladores/>
Fecha de consulta: agosto 2005

[CIS02] Cisterna M.(2002) “Métodos de Optimización de Consultas para el Lenguaje SQL”, <http://macine.epublish.cl/tesis/index.html>
Fecha de consulta: agosto 2005

[COC00] Alistair Cockburn “Writing Effective Use Cases, Addison-Wesley

[CON05] Documentación oficial Oracle Database, Concepts 10g Release 2- June 2005

[DAT97] Date C. J. - Darwen H. (1997) “A guide to the SQL Standard”, Addison Wesley, Reading – MA, Fourth Edition

[DB208] DB2 edicion express C <http://www-01.ibm.com/software/data/db2imstools/db2tools/db2wqt/> visitado mayo 2008

[DEE02] Deepak P. - Sandeep S. (24 de agosto 2002) "Valgrind HOWTO", <http://www.tldp.org/HOWTO/Valgrind-HOWTO/>
Fecha de consulta: desde enero a agosto de 2005

[EMS08a]<http://www.sqlmanager.net/products/mssql/datagenerator>

[EMS08b] <http://www.sqlmanager.net/?gclid=CNeW07SouZUCFQyenAod9gQCQA> visitado junio 2008

[GAS05] Gastañaga, Iris et al TecnoDB: Desarrollo de una metodología de aprovechamiento de Metadatos de los diccionarios de Datos de Bases de Datos Relacionales para lograr un generador de sentencias SQL JAIIO34 ASIS 2005 Argentine

Symposium on Information Systems Rosario,
Argentina - September 29-30, 2005

[GAS06] Gastañaga, Iris et al TecnoDB una Base de
Datos Relacional y Prometeo un metodo de
aprovechamiento de Metadatos y Generador de
Consultas WICC 2006

[ISQ,08] <http://java-source.net/open-source/sql-clients/isql-viewer> visitado Julio de 2008

[JAC00] Jacobson I. - Booch G. - Rumbaugh J.
(2000) “El Proceso unificado de Desarrollo de
Software”, Addison Wesley, Madrid – España

[Lar00] Craig Larman “Applying UML and
Patterns”, Addison- Wesley

[MYS,08]
<http://www.mysql.com/products/tools/query-browser/> visitado Mayo 2008

[ORA08]
<http://www.oracle.com/global/kr/download/seminar/2008/otn/s01.pdf> visitado Junio 2008

[QUE08] <http://www.quest.com/> visitado en Julio
2008

[QTO08] <http://www.querytool.com/> visitado en
julio 2008

[SPE07] Martinez Spessot et al, PROMETEO Un
método de explotación de Metadatos. JIDIS 07

[TOA08] <http://www.toadsoft.com/>

[UNI05a]<http://www.universia.com.ar/contenidos/universidades/index.html>
Fecha de consulta: agosto 2005

[UNI05b]“Universidades.org”,
<http://www.universidades.org/> - Fecha de consulta:
agosto 2005

[SUN09]
<http://java.sun.com/javase/technologies/desktop/javabeans/index.jsp>

[WIK09a]<http://es.wikipedia.org/wiki/Prometeo>
visitado Abril 2009

[WIK09b] Java Beans
<http://es.wikipedia.org/wiki/JavaBean> - Fecha de
consulta: desde diciembre 2008