

3rd Workshop on MSc dissertations and PhD thesis in Artificial Intelligence (WTDIA'2006)

5th Best MSc dissertation / PhD thesis contest (CTDIA'2006)

CTDIA'2006 is the 5th best MSc Dissertation/ PhD Thesis contest in Artificial Intelligence. Its main goal is to award the three best academic work developed at universities from Ibero-American countries in the last two years.

WTDIA'2006 is the 3rd Workshop on Msc Dissertation and Phd Thesis in Artificial Intelligence. It provides an opportunity for Ph.D. and Ms.C. students to present their on going research with a panel of established researchers in artificial intelligence.

CTDIA'2006 and WTDIA'2006 are co-located with the International Joint Conference SBIA/IBERAMIA /SBRN'2006.

For CTDIA'2006 we had 30 submissions (10 Ph.D. Thesis and 20 M.Sc. Dissertations). The contest was carried out in two phases. In the first phase, all papers were peer-reviewed by three referees of the area and 15 were selected for the second phase of evaluation (5 Ph.D. Thesis and 10 M.Sc. Dissertations). In the second phase, two Award Program Committees (APC), one for Ph.D. Thesis and the other for M.Sc. Dissertations, were responsible for selecting the three best Thesis/Dissertation in each category. To the fairness of this process, in each category, each member of the APC re-evaluated all papers selected in the first phase.

For WTDIA'2006 we had 39 papers (14 Ph.D. Thesis and 25 M.Sc. Dissertations). All papers were peer-reviewed by at least two referees of the area. As a result of this selection process, 22 papers (11 Ph.D. Thesis and 11 M.Sc. Dissertations) were accepted for presentation at the Workshop.

The CTDIA'2006 and WTDIA'2006 chairs would like to thank the authors for submitting their work and the Award and Program Committee members, as well as the other reviewers, for their dedication in the reviewing process.

Chairs:

José Augusto Baranauskas (FFCLRP/USP, Brazil)
Maria Carolina Monard (ICMC/USP, Brazil)

Agentes y Bases de Datos Activas: Un modelo formal de propósito general y una aplicación al registro y consulta de datos y metadatos científicos ambientales

Propuesta Doctoral

Virginia Padilla Sifontes

Estudiante Doctoral del Centro de Investigación y Proyectos en Simulación y Modelos

Universidad de Los Andes, Mérida, Venezuela.

vpadilla@uneg.edu.ve

Tutor:

Dr. Jacinto Dávila

jacinto@ula.ve

Centro de Investigación y Proyectos en Simulación y Modelos

Universidad de Los Andes, Mérida, Venezuela.

Resumen

Este es un proyecto de investigación en modelado de bases de datos y conocimiento donde nos proponemos desarrollar un modelo formal basado en Agentes que describa las facilidades de un Sistema de Base de Datos Activa de propósito general. El modelo debe manejar datos complejos, específicamente datos y metadatos sobre sistemas ambientales o ecosistemas y debe servir como soporte para el modelaje de la Biocomplejidad. Con el modelo formal esperamos también poder estudiar la relación entre las Bases de Datos Activa y los sistemas basados en ontologías que caracterizan a la Web Semántica.

Este trabajo está enmarcado en el Proyecto que llevan adelante Venezuela y Estados Unidos de América, para el Modelaje de la Biocomplejidad. Las Universidades participantes son la Universidad de Los Andes (Mérida, Venezuela) y Universidad Nacional Experimental de Guayana (Ciudad Guayana, Venezuela) y la Universidad de North Texas (Texas, EEUU) bajo el auspicio de NSF's Internacional Program (National Science Foundation) y el Fonacit.

La fecha estimada de culminación de este trabajo de investigación es el año 2009

Palabras Claves: Base datos activas, Agentes, Ontologías, metadatos científicos

Introducción. Nos proponemos desarrollar un modelo formal basado en Agentes que describa las facilidades de un Sistema de Base de Datos Activa de propósito general. Como modelo formal, queremos referirnos a una descripción lógica matemática de una base de datos con agentes caracterizando la dinámica que la convierte en una base de datos activa. Un modelo así podría servir, no solamente para describir y comparar sistemas manejadores de bases de datos activas ya existentes, sino para especificar

futuros sistemas, incluyendo algunos que aprovechen la tecnología de agentes para prestar servicios inéditos como los que requieren grandes almacenes de conocimiento, los simuladores de sistemas complejos y los sistemas basados en ontologías. De hecho, nos proponemos evaluar el modelo implementando prototipos funcionales de almacenes de datos y metadatos de relativa complejidad. Mas aún, esos prototipos funcionales implementarían variantes regulares del modelo general que pueden servir como mecanismos de evaluación (*benchmarking*) de diversos sistemas manejadores de bases de datos.

Por esta última razón, este proyecto se inscribe en el dominio de los estudios de Biocomplejidad, como área de aplicación. El modelo debe facilitar la caracterización de datos complejos, específicamente datos y metadatos ambientales y debe servir como soporte para el modelaje de la Biocomplejidad¹. La biocomplejidad ambiental [33] se refiere a la interacción compleja entre los sistemas vivientes y su ambiente, reflejando las propiedades que surgen de esta interacción, haciendo énfasis en la riqueza de sistemas ecológicos y su capacidad por la adaptación y la auto-organización. Es un dominio de conocimiento que plantea claros desafíos para las bases de datos en razón de los volúmenes típicos de datos que se deben manipular, pero también por la variedad de relaciones, transformaciones y consultas que experimentan esos datos en su cultivo y uso regular en la investigación científica.

A Continuación, precisamos la definición de Base de Datos Activa a partir de las referencias más conocidas en la literatura y, seguidamente, realizamos una breve introducción a la llamada teoría de Agentes de la Inteligencia Artificial. Bases de Datos Activa y Agentes Inteligentes se convierten así en los conceptos fundamentales en este proyecto.

Bases de Datos Activas. En [14] se define un Sistema de Base de Datos Activa como uno que proporciona funcionalidad adicional para especificar reglas activas. Esquemáticamente, un Sistema de Base de Datos Activa:

- Es un Sistema de Base de Datos
- Proporciona la capacidad de *monitorear* y *reaccionar* a circunstancias específicas que son relevantes a una aplicación.

El primer punto nos indica que el Sistema de Base de Datos Activa posee un Sistema Manejador de Base de Datos y la propia Base de Datos. Un *Sistema Manejador de Bases de Datos* es un sistema de software cuyo propósito es definir, construir, mantener y operar, de manera confiable y eficiente, bases de datos multi-usuarios

El segundo punto se refiere a la reactividad del sistema. La reactividad como una propiedad de sistemas puede ser enfocada, según [12], desde dos aspectos:

- **Reactividad de la Arquitectura:** capacidad del sistema de procesar las entradas en tiempo real y dar una respuesta oportuna. Procesamiento de entrada en tiempo real significa que el sistema debe

¹ Este proyecto está enmarcado en el Proyecto Nacional de Modelaje de la Biocomplejidad que llevan adelante la Universidad de Los Andes (Mérida, Venezuela) y Universidad Nacional Experimental de Guayana (Ciudad Guayana, Venezuela) y la Universidad de North Texas (Texas, EEUU) bajo el auspicio de NSF's Internacional Program (National Science Foundation) y el Fonacit.

permitir la suspensión del procesamiento normal a favor de una revisión periódica de las entradas, y el subsiguiente re-inicio del procesamiento, presumiblemente desde el punto donde fue parado.

- **Reactividad basada en el conocimiento:** un segundo aspecto de la reactividad está íntimamente enlazada a la estructura de conocimiento en la base de conocimiento del sistema. Si el conocimiento del sistema está altamente optimizado y compilado juntos, por ejemplo, a las reglas condición acción, el sistema puede reaccionar a entradas en un mínimo tiempo con una gran oportunidad de éxito debido a su respuesta oportuna (*timely response*). En una configuración extrema, estas reglas condición acción son guardadas en una tabla o en un arreglo que directamente relaciona las entradas de los sensores a las acciones dictadas a los efectores.

Típicamente, el comportamiento activo de una base de datos se describe usando reglas, las cuales tienen tres componentes: un evento, una condición y una acción. Una regla con tales componentes es conocida como reglas eventos-condición-acción o reglas ECA. El significado de estas reglas es: "cuando un **evento** ocurre, chequea la **condición** y si ésta condición se cumple, ejecuta la **acción**". Una regla permanece dormida hasta que se monitorea la ocurrencia de un evento, entonces la regla se 'dispara'. Las condiciones de la regla disparada son evaluadas, y si son todas verdaderas, la regla se agrega a un *módulo de conflicto*, en el cual se almacenan las reglas disparadas que son luego accedidas por el supervisor, quien selecciona la acción de la regla para su ejecución.

Las características de estructura y comportamiento de un Sistema de Bases de Datos Activa puede clasificarse, según [19], en tres áreas: modelo de conocimiento, modelo de ejecución, y modelo de administración.

El modelo de conocimiento representa la vista sintáctica de las reglas activas como es vista por el programador de las reglas. Esto tiene tres principales facetas, según [19]:

- *lenguaje de evento:* una notación con la que se especifican las condiciones que disparan una regla.
- *lenguaje de condición:* una notación usada para expresar restricciones adicionales las cuales pueden ser satisfechas por la base de datos después que la regla disparada es añadida al conjunto de conflicto.
- *Lenguaje de acción:* una notación usada para especificar el efecto que la regla debe tener cuando es ejecutada.

Como ya anunciamos, nos proponemos hacer un modelo formal de propósito general para un Sistema de Bases de Datos Activa, y estas notaciones que especifican eventos, condición y acción trataremos de integrarlas en un solo lenguaje con una sintaxis común y una semántica común. Un único lenguaje nos proporcionaría, de inmediato, la posibilidad de comparar modelos particulares que describan sistemas particulares, apelando al modelo general implícito en la semántica del lenguaje. Esto, desde luego, supone que hayamos podido crear ese modelo general de referencia para el lenguaje integrado, incorporando, en el lenguaje, todos los conceptos y constructos necesarios para describir todos los tipos y variantes de bases de datos activas existentes. Este es, desde luego, el mayor desafío.

Por su parte, el modelo de ejecución describe como las reglas interactúan en el contexto de todo el sistema de base de datos. Esto tiene los siguientes aspectos:

- *Granularidad de la transacción:* la naturaleza de los enlaces entre la ocurrencia de los eventos y la activación de las reglas – es posible que un evento individual ocurra cuando se dispare una regla, o que una colección de ocurrencia de un evento dispare una regla.
- *Modo acoplamiento:* la relación temporal y causal entre el disparo y la ejecución – por ejemplo, es posible que la acción de la regla sea ejecutada tan pronto como es posible después de la evaluación de la condición de la regla (inmediato), o en un punto retardado, tal como al final de la transacción (diferido).
- *Esquema de prioridades:* un ordenamiento especificando cuales reglas son consideradas primero cuando varias han sido disparada al mismo tiempo (ejemplo: cuando el conjunto de conflicto tiene más de una regla).
- *Políticas de efecto Nulo:* los cuales permiten cambios internos en los estados intermedios a una transacción para que sea ignorada por las reglas del sistema.

Un error común en los modelos formales, como muchas especificaciones basadas en lógicas modales, es relegar el equivalente al modelo de ejecución del sistema a un segundo plano, concentrándose únicamente en la estructura conceptual, sin mayor reparo por la operacionalidad del modelo. En este proyecto, trataremos de aprovechar las experiencias en lógica computacional que elaboran la estructura conceptual tratando de responder, al mismo tiempo, preguntas fundamentales acerca de la operacionalidad del modelo.

Finalmente, el modelo de administración comprende propiedades de las reglas bases tomadas como un todo, también como las operaciones para modificar las reglas:

- *Operaciones con reglas:* algunas bases de datos activas permiten que las reglas sean creadas o removidas, activadas o desactivadas, durante la ejecución.
- *Terminación:* sí o no la ejecución de una regla dada debe alcanzar un estado final.
- *Confluencia:* sí (en general no determinístico) el conjunto de reglas define un estado final único dado cualquier contexto de iniciación
- *Modelo de Equivalencia:* una noción de equivalencia semántica entre conjuntos de reglas, el cual es un prerrequisito para la optimización, ya que permite evaluar y decantar reglas equivalentes.

En [19] se evalúan diferentes propuestas de Sistemas de Gestión de Base de Datos Activa en términos de esos tres modelos y se describen los propósitos específicos para la descripción formal del comportamiento activo, por lo cual ese trabajo es el predecesor principal del proyecto que acá se describe. Sin embargo, al igual que otros esfuerzos que explicamos a continuación, ese trabajo no alcanza a integrar, en un todo coherente, los tres modelos que se describieron. Aún en este punto inicial del proyecto, tenemos razones para creer que un modelo basado en agentes podría proveer ese espacio integrado coherente.

Otros intentos por caracterizar las Bases de Datos Activas son:

- **Starburst in notación semántica.** El modelo denotacional del sistema de reglas activas Starburst [26] establece como puede especificarse formalmente el modelo de ejecución de un sistema de base de