

VII.4 SISTEMAS EVOLUTIVOS BASADOS EN CONOCIMIENTO

*Fernando Galindo Soria**

Los Sistemas Basados en Conocimiento son herramientas más generales que existen ya que por ejemplo un Sistema de Biblioteca Automatizada, un Sistema Experto o un Sistema Manejador de Base de Datos pueden verse como casos particulares de este tipo de sistemas. Sin embargo lo común es que se construyan sistemas particulares y no se ataque el caso general, por lo que, en este trabajo encontraremos como se puede construir una herramienta general para manejo de conocimiento soportado por un Sistema Evolutivo[1][2].

Como primer punto partimos de que en un Sistema Basado en Conocimiento. se cuenta con diferentes mecanismos para representarlo y manejarlo. Incluyendo:

- 1) *Directorios de datos*
- 2) *Archivos de datos*
- 3) *Bases de datos*
- 4) *Bases de conocimiento*
- 5) *Conjuntos de reglas de inferencia.*

y dependiendo del tipo de sistema, los medios de manejo de conocimiento se pueden Integrar de diferentes formas por ejemplo, en los sistemas de base de datos es común encontrar un directorio de datos interrelacionado con la base de datos, los sistemas expertos o de toma de decisiones integran al menos una base de conocimientos y un conjunto de reglas de inferencia, los

* Fernando Galindo Soria escribió este trabajo estando en la Escuela Superior de Computo del IPN, lo presentó como Conferencia Magistral y publicó en las memorias del Simposio Internacional de Computación, organizado por el CENAC del IPN, en la Cd. de México en Noviembre de 1994.

sistemas de biblioteca automatizada están formados al menos por una combinación de bancos de información (donde están almacenados los documentos) y directorios de datos (que permiten acceder directamente a elementos específicos de los bancos de información).

I ARQUITECTURA DE SISTEMAS EVOLUTIVOS BASADOS EN CONOCIMIENTO

La forma como se interrelacionan los elementos también depende de la aplicación, formando diferentes tipos de arquitectura. en particular en este trabajo partiremos de una arquitectura por capas. en la cual se tienen los diferentes componentes de un sistema basado en conocimiento colocados en capas relacionadas entre si, de tal manera que una capa sirve para acceder o localizar conocimientos de otras. como se ve en la figura 1.

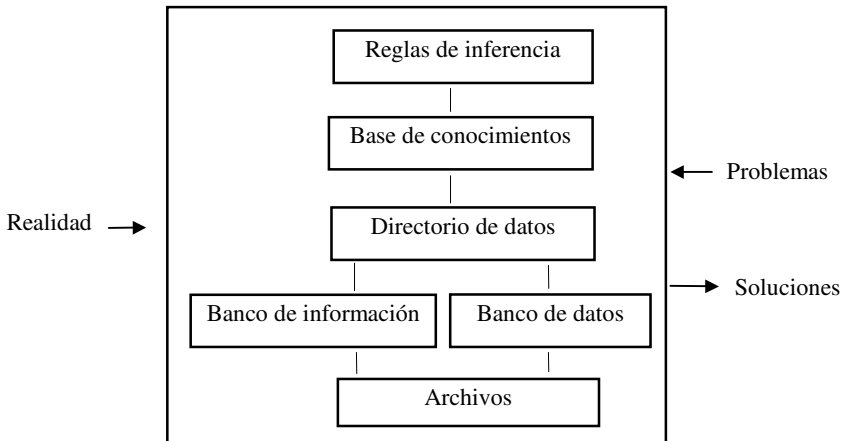


Figura 1. Arquitectura de los Sistemas Basados en Conocimiento

Como se observa, el sistema esta formado por un conjunto de mecanismos de representación de conocimiento relacionados entre si y un conjunto de mecanismos de percepción y actuación que permiten interactuar con el sistema y mantener actualizada su información.

Sin embargo la arquitectura no es tan simple ya que un problema muy grave de este tipo de sistemas se presenta en la adquisición, mantenimiento y explotación del conocimiento, ya que llega un momento en que el volumen de datos a manejar hace extremadamente complejo su tratamiento manual.

Por lo que. en este trabajo se plantea que el proceso de Adquisición del conocimiento se debe realizar en forma automatizada, se propone un conjunto de herramientas evolutivas orientadas a la adquisición y mantenimiento del sistema basado en conocimiento y en los siguientes puntos se describirá como se puede construir un Sistema Evolutivo Basado en Conocimiento y bajo una arquitectura por capas.

II GENERACIÓN AUTOMÁTICA DEL DIRECTORIO DE DATOS

Por ejemplo la creación del directorio de datos en forma manual puede terminar siendo un proceso muy tedioso, ya que. cada palabra clave en el directorio se necesita ubicar en los bancos de información para detectar en que párrafos se presenta, de tal forma que cuando un usuario pregunte por esa palabra el sistema le indique en que párrafos se encuentra.

Realizar lo anterior a mano puede ser un proceso aburrido o impráctico. por lo que desde hace muchos años se han desarrollado herramientas que construyen los directorios en forma automatizada.

Por ejemplo durante los años 70's en El Colegio de México se inicio el proyecto del Diccionario del Español en México donde se propuso hacer un diccionario con todas las palabras que utilizan los hispanohablantes en México, para realizar este proyecto se analizaron varios millones de documentos buscando las diferentes palabras y sus significados.

Definitivamente este proyecto no se hubiera podido realizar manualmente. ya que la cantidad de información a tratar era enorme, por lo que se construyó. bajo la dirección del Ing. José

Cen Zubieta un sistema de apoyo automatizado, al cual se le daban como entrada los documentos a analizar y el sistema generaba para cada palabra la lista de los párrafos donde aparecía esa palabra. de tal forma que los lingüistas podían estudiar el uso de cada palabra en diferentes contextos y de ahí le asociaban su significado.

Basados en la idea anterior se puede construir una herramienta que analice los documentos almacenados en el banco de información y a partir de ahí la primera capa del sistema evolutivo toma los datos del banco y construye el directorio de datos, indicando para cada palabra clave la localización de los párrafos donde aparece.

III GENERACIÓN AUTOMÁTICA DE LA BASE DE CONOCIMIENTOS

No sólo la creación del directorio de datos puede ser un problema oneroso, también la creación y mantenimiento de los otros componentes del sistema puede llegar a ser muy pesado para realizarse a mano, por lo que como siguiente capa del sistema se integrará un mecanismo que permita la construcción automática de la Base de Conocimientos.

La Base de Conocimientos es uno de los componentes mas importantes de un sistema basado en conocimiento. ya que permite representar las interrelaciones entre las diferentes palabra clave y su significado.

Existen varios mecanismos para representar el conocimiento, pero uno de los mas usados son las redes semánticas, donde una red semántica esta formada por hechos o reglas de la forma objeto relación objeto (o r o), por ejemplo a partir de la oración:

Juan es hermano de Pedro

se detecta que Juan es un objeto, es hermano de es una relación y Pedro es otro objeto, por lo que de esta oración se obtienen directamente los componentes de una red semántica.

Por lo común los documentos no vienen divididos en párrafos de la forma o r o, por ejemplo si observa este artículo se dará

cuenta que existen pocas oraciones donde directamente aparezca un objeto seguido de una relación y de otro objeto. por lo que si se quiere construir la red semántica a partir de las oraciones presentes en un documento se requiere de un análisis mas profundo, esto fue lo que se realizó en la tesis de licenciatura de Jesús Olivares titulada “Sistema Evolutivo para Representación de Conocimiento” desarrollada en la UPIICSA del IPN [3], ahí se planteó un mecanismo evolutivo para construir un base de conocimientos basada en una Red Semántica Aumentada (que almacena datos, imágenes, procesos, sonido y otros elementos).

La idea de Jesús consistió básicamente en aplicar una serie de técnicas de Inferencia Gramatical basadas en la distribución lingüística [3][4], para transformar una oración compuesta de múltiples elementos en un conjunto de oraciones semánticamente equivalentes a la original, donde cada oración final es de la forma **o r o**. por lo que, son fácilmente representadas con una red semántica.

Por ejemplo en la oración:

Juan estudia en el Poli, trabaja en el INEGI y vive en Lindavista
se tienen varios hechos pero sin embargo no son directamente representables mediante una red semántica, por lo que se le aplicará la distribución lingüística, por facilidad representaremos los elementos de la oración mediante letras y números, de tal manera que se note directamente cuales son los objetos, cuales las relaciones y cuales los separadores:

Juan	es un objeto	o1
estudia en el	es una relación	r1
Poli	es un objeto	o2
,	es un separador	+
trabaja en el	es una relación	r2
INEGI	es un objeto	o3
y	es un separador	+
vive en	es una relación	r3
Lindavista	es un objeto	o4

de donde la oración quede representada como:

$$o1(r1o2 + r2o3+r3o4)$$

El paréntesis que esta después. de o1 indica que tiene relación con los elementos restantes de la oración (en la tesis de Elsa Berruecos[6] se presenta un método para encontrar en forma automática la dependencia entre los diferentes elementos de una oración).

Si se toma la oración y se le aplica la operación de distribución del álgebra, queda:

$$o1 (r1o2 + r2o3 + r3o4) = o1r1o2 + o1r2o3 + o1r3o4$$

y si sustituimos el significado de los diferentes elementos tenemos que:

o1r1o2	Juan estudia en el Poli
+	
o1r2o3	Juan trabaja en el INEGI
+	
o1r3o4	Juan vive en Lindavista

o sea que a partir de una oración compleja se obtuvieron tres oraciones simples del tipo

o r o

que mantienen el significado de la oración original y que son directamente representables en una red semántica

IV GENERACIÓN AUTOMÁTICA DE LAS REGLAS DE INFERENCIA

El conjunto de reglas de inferencia es otro de los componentes básicos de un Sistema Basado en Conocimiento nuevamente su generación manual puede ser muy tediosa por lo que nos basaremos en un trabajo desarrollado junto con Javier Ortiz en el CENIDET de Cuernavaca [5] para la generación automatizada de reglas de inferencia. Para lo cual partiremos que en general una regla de inferencia es un regla de la forma:

Si A entonces B

esto es, si se cumple A entonces B es cierto donde A en general es un conjunto de condiciones o síntomas sobre un problema y B en

general se puede ver como una acción o conjunto de acciones que se ejecutan si se cumple A, en particular B es el diagnóstico y tratamiento asociado con los síntomas de A.

Para encontrar las reglas de inferencia a partir de un documento se toma del documento un párrafo donde se presente un conjunto de síntomas asociados con un diagnóstico y con un conjunto de tratamientos, se sustituye cada elemento por una representación interna donde se indica si un elemento es síntoma, diagnóstico o tratamiento e ignorando los elementos que no tengan significado por lo que al final queda una representación de la forma:

$$S_1 S_2 S_3 \dots S_n D_1 T_1 T_2 T_3 \dots T_m$$

esto se realiza para todos los párrafos del documento y al final se cuenta con un conjunto de reglas como:

$$S_{1_1} S_{2_1} S_{3_1} \dots S_{n_1} D_1 T_{1_1} T_{2_1} \dots T_{m_1}$$

$$S_{1_2} S_{2_2} S_{3_2} \dots S_{n_2} D_2 T_{1_2} T_{2_2} \dots T_{m_2}$$

...

$$S_{1_k} S_{2_k} S_{3_k} \dots S_{n_k} D_p T_{1_k} T_{2_k} \dots T_{m_k}$$

Ahora bien asociada con la distribución lingüística se encuentra otra operación algebraica conocida como factorización lingüística y que es inversa de la distribución o sea que si factorizamos o_1 de la oración:

$$o_1 r_1 o_2 + o_1 r_2 o_3 + o_1 r_3 o_4$$

nos queda:

$$o_1 (r_1 o_2 + r_2 o_3 + r_3 o_4)$$

Lo interesante es que si mediante la distribución se pueden encontrar los componentes de la base de conocimientos, mediante la Factorización Lingüística se pueden encontrar las reglas de inferencia del sistema basado en conocimiento. ya que sólo se necesitan tomar las reglas encontradas anteriormente verlas como una expresión algebraica y factorizar los elementos comunes, de tal manera que los síntomas comunes a varios diagnósticos quedan juntos y los síntomas particulares de un diagnóstico quedan asociados con él.

CONCLUSIÓN

Como se puede ver en este documento partimos de que se cuenta con un banco de información al que se le aplican técnicas evolutivas para construir el Directorio de Datos, la Base de Conocimiento y el conjunto de Regias de Inferencia.

Lo maravilloso de este enfoque es que a partir de la misma fuente de información se pueden encontrar diferentes niveles de representación de conocimiento y la que se esta realizando es un diálogo con el interior del Banco de Información para encontrar reglas y conocimiento que se encuentra inmerso en las fuentes de información y que por lo común se pierden.

Con este trabajo se quiere mostrar que en los bancos de información se tiene una gran cantidad de información oculta, que sólo es necesario construir herramientas capaces de realizar procesos de diálogo con los almacenes de información y estos mecanismos son capaces de encontrar una gran cantidad de información inmersa y que hasta el momento no se aprovecha por no contar con las herramientas adecuadas, pero principalmente por desconocimiento de que estamos sentados sobre una mina de información y no sabemos aprovecharla o siquiera que existe.

BIBLIOGRAFÍA

- [1] Galindo Soria, Fernando, *Sistemas Evolutivos: Nuevo Paradigma de la Informática*, Memorias de la XVII conferencia Latinoamericana de Informática Caracas, Ven. Julio. 1991
- [2] Galindo Soria, Fernando, *Sistemas Evolutivos en Boletín de Política Informática*, INEGI-SPP, México. Sept 1986
- [3] Olivares Ceja, Jesús Manuel, *Sistema Evolutivo para Representación del Conocimiento*, UPIICSA-IPN, México, D.F., 1991
- [4] Galindo Soria, Fernando, *Algunas Propiedades Matemáticas de los Sistemas Lingüísticos*, Memorias del Simposium Nacional de Computación, México, nov. de 1992
- [5] Ortiz, Javier, Galindo Soria, Fernando, *Sistemas Evolutivos Constructores de Sistemas Expertos*, CENIDET-DGIT. UPIICSA-IPN, Cuernavaca. Mor. 1988
- [6] Berruecos Rodríguez, Elsa, *Sistema Evolutivo Generador de Esquemas Lógicos de Bases de Datos*, UPIICSA-IPN, México. 1990