

MI
291.18
93

INEC

ARGENTINA

Dirección Nacional de Estadísticas Sociodemográficas.
Serie: Documentos Técnicos.
Documento Nro.: 18



21 MAYO 1991

**"SEGUNDA REUNION DEL COMITE DE EXPERTOS EN
ESTADISTICAS SOCIODEMOGRAFICAS"**

Buenos Aires, 22 al 25 de Agosto de 1989.

SISTEMA EXPERTO DE PROPOSITO GENERAL
PARA EL AREA SOCIODEMOGRAFICA

Con la colaboración del Programa de Naciones Unidas para el
Desarrollo y del Fondo de Población de Naciones Unidas y de la
Oficina Panamericana de la Salud.

M
291.18
92

P R E S E N T A C I O N

El contenido de este documento expresa una línea de trabajo que desde el ámbito de la computación, se suma a los esfuerzos orientados a la reformulación de las estadísticas sociodemográficas.

La elaboración de este documento ha estado bajo la responsabilidad de: Susana Silberberg, Sandra Levy, Eugenio Pettigiani, Andrea Lorenzetti, Claudia Parceró, Carlos Sarachu.

C O N T E N I D O

I. INTRODUCCION.....	03
II. ANTECEDENTES.....	03
III. OBJETIVOS.....	04
III.1. Objetivo Principal.....	04
III.2. Objetivos Secundarios.....	04
IV. METODOLOGIA.....	04
IV.1. Partes que componen un Sistema Experto.....	04
IV.2. Descripción del modelo adoptado.....	05
V. DESARROLLO DE ACTIVIDADES.....	11
V.1. Primera Etapa.....	11
V.2. Segunda Etapa.....	13
V.3. Cronograma de Actividades.....	14
 Bibliografía.....	 15

I. - INTRODUCCION.

El presente documento tiene por objeto presentar un conjunto integrado de herramientas de computación, para la evaluación y análisis de información, de una manera eficaz, rápida, sin redundancias, con resultados inmediatos y de fácil utilización por el usuario NO experto, de todos los datos provenientes de encuestas.

La idea madre fue implementar un administrador de datos capaz de soportar estructuras de Redes Plex y que el usuario las pueda acceder en forma de relaciones binarias (ver mayores detalles en la obra "Organización de las Bases de Datos", de James Martin, Editorial Prentice Hall, los capítulos sobre "Esquemas y Sub-Esquemas" y "Visiones Lógicas y Físicas - Su diferenciación"). Surgió a continuación y de una manera natural el propósito de extender el soporte básico de dichas redes, haciéndolo apto para manejar proposiciones de lógica y manejo de expresiones indeterminadas (ver mayores detalles en la obra "Lógica para Computación", de Francisco S. Naishtat, Editorial EUDEBA, en particular la parte de "Cálculo proposicional").

Como resultado de lo anterior y a partir de las necesidades y herramientas disponibles, se procedió al diseño y análisis de un "Sistema Experto de Propósito General" para PC's.

El mismo estará implementado utilizando un modelo de Marcos y Reglas con una valoración probabilística de la certidumbre que responde a la teoría de Dempster-Shafer (ver mayores detalles en la obra "Ingeniería del Conocimiento", de J.L.Maté-Hernández - J.Pazos Sierra, Editorial SEPA, capítulo 4).

II. - ANTECEDENTES.

Si bien el modelo que a continuación se describe no tiene antecedentes dentro del ámbito estatal en nuestro país, está siendo ampliamente utilizado dentro del ámbito privado principalmente en aplicaciones industriales, de servicios, atención médica, etc; y en el ámbito oficial (principalmente en los EEUU), como una herramienta eficaz para reducir los tiempos y los costes de tareas que hasta el presente se consideraban "No Sistematizables".

El enfoque básico apunta a integrar un conjunto de técnicas de trabajo cuyo componente principal son las derivadas de la "Inteligencia Artificial". Dentro de estas técnicas tomaremos las aplicadas al campo de los "Sistemas Expertos", estableciendo además un punto de contacto que garantice la continuidad con lo existente dentro de los modelos de programación clásica, permitiendo capturar el conocimiento de un "experto humano" sistematizarlo y ponerlo al alcance de todos los potenciales usuarios.

Citando a los autores, Edward Feigenbaum y Enrique Castillo, podemos definir un sistema experto como:

"...un programa de computador inteligente que usa conocimiento y procedimientos de inferencia para resolver problemas que son lo suficientemente difíciles como para requerir la intervención de un experto humano para su resolución" (Feigenbaum).

"...un sistema informático que simula el proceso de aprendizaje, de memorización, de razonamiento, de comunicación y de acción de un experto humano en una determinada rama de la ciencia, suministrando, de esta forma, un consultor que puede sustituirle con unas ciertas garantías de éxito" (Castillo).

III.- OBJETIVOS

III.1.- Objetivo principal:

El objetivo principal es que las distintas subdirecciones de la Dirección Sociodemográfica y los distintos departamentos de la Encuesta Permanente de Hogares logren una comunicación fluida de los "Conocimientos" específicos que cada experto dentro de su área posee a través de una herramienta que les resulte de uso "amigable".

III.2.- Objetivos secundarios:

- Mejorar la calidad del conocimiento de los expertos humanos.
- Multiplicar el número de expertos.
- Disminuir el coste del conocimiento.
- Utilizar personal no especializado para resolver problemas que requieren especialidad.
- Inferir nuevo conocimiento en forma semi-automática.
- Inferir datos lógicos y/o atributos a partir de los existentes.
- Aumentar la velocidad de procesamiento y análisis de la información.
- Facilitar el acceso del conocimiento a poblaciones más amplias.
- Aumentar la confiabilidad del conocimiento.
- Disponer de una herramienta ágil y de fácil adaptación a nuevos requerimientos de los usuarios.

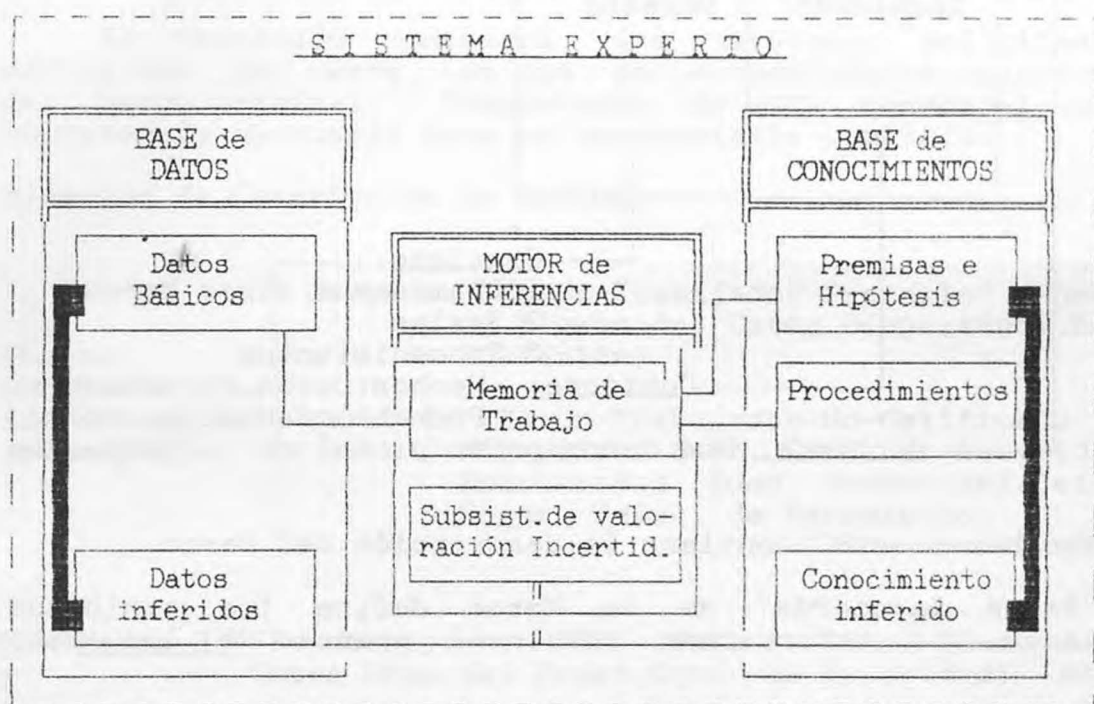
IV.- METODOLOGIA.

IV.1.- Partes que componen un Sistema Experto.

Un Sistema Experto es un sistema compuesto por una serie de Subsistemas que funcionan en forma integral de una manera similar a la del cerebro humano. Se lo puede dividir en:

- . Base de Datos (contiene la información "en bruto")
 - .. Datos y Atributos de los Datos.
 - .. Relaciones entre los datos.

- . Motor de Inferencia (es la parte "pensante" del sistema)
 - .. Memoria de Trabajo. (contiene todo lo necesario para resolver el problema que este activo).
 - .. Subsistema de Valoración de la Incertidumbre.
- . Base de Conocimientos (contiene la "experiencia" del experto humano).
 - .. Reglas
 - ... Premisas (de base o aprendidas)
 - ... Hipótesis de trabajo
 - .. Procedimientos (forman el punto de conexión con toda la programación existente y permite "salvar" aquellas situaciones en las que la "lógica" se torna insuficiente).



IV.2.- Descripción del modelo adoptado

Se decidió utilizar un "Modelo de Marcos y Reglas con una valoración probabilística de la certidumbre".

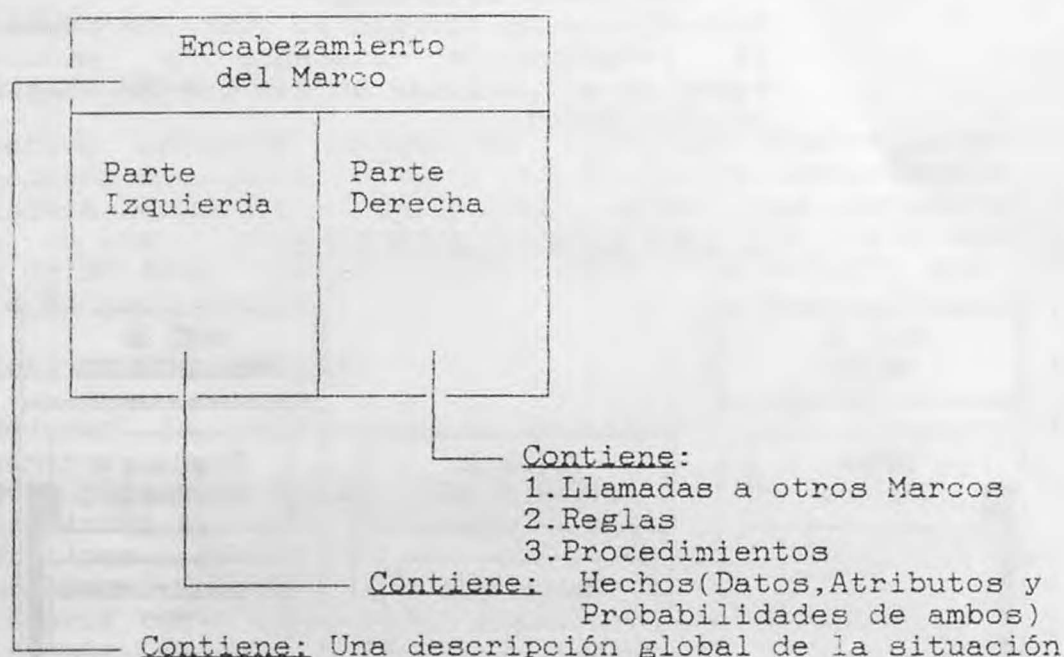
Un Marco es la descripción de una situación (obviamente dentro de algún contexto), lo mas completa posible y, en la medida en que la misma lo permita, SIN AMBIGÜIDADES (cosa que en la práctica no se da, y que nos obliga a establecer un factor de certeza o una ponderación de evidencias para poder suplir este punto). Definamos entonces que debe contener un Marco:

- a) Una representación del conocimiento a almacenar.

- b) Estructuras de control sobre el conocimiento (permítasenos decir, "El conocimiento que se tiene acerca del conocimiento") o Meta-Conocimiento.
- c) Guardar "Memoria" del conocimiento aprendido (y/o deducido) y establecer formas de inferencias.

Partes componentes de un Marco:

Un marco queda definido por tres partes principales:



El "Encabezamiento" contiene la descripción del marco

La "Parte Izquierda" de un Marco define los atributos, probabilidades, e información adicional acerca de los datos almacenados.

La "Parte Derecha" es la que nos permite deducir o inferir, nuevos hechos, realizar procedimientos algorítmicos y establecer cadenas de razonamiento inter o intra - marcos.

Descripción de un hecho: (parte izquierda de un Marco)

La descripción de un hecho debe contener no sólo los datos a almacenar sino también la forma en que se relaciona con otros hechos (de otros marcos), e información adicional útil para el usuario (atributos).

Descripción de las reglas: (parte derecha de un Marco)

Las reglas están formadas por expresiones del tipo:

- a) X1 es cierto SI X2 es cierto Y X3 es cierto
- b) X1 es cierto SI X2 es cierto
- c) X1 es cierto? / <Cte> es cierto

- d) Ejecute Procedimiento(Z)
- e) Ejecute Marco(Z)

Las reglas de tipo a) y b) son las que llamamos "Hipótesis" o "Hechos a Confirmar". Las de tipo c) nos definen hechos que son ciertos porque el experto lo dice, o nos permiten establecer interrogantes directos sobre los hechos almacenados.

Las de tipo d) nos permiten realizar cálculos y/o algoritmos que establecen el punto de conexión con la programación clásica. Y el tipo e) nos permite la realimentación y encadenamiento dentro de una "línea de razonamiento".

La equiparación de un hecho dentro de una regla se realiza "instanciando" los atributos solicitados en la regla con los existentes en el hecho, y además, por medio de la probabilidad asignada y la requerida en la regla, seleccionando aquellos hechos que verifiquen el factor de certeza pedido.

El resultado contendrá los atributos solicitados en la definición del marco, los que no necesariamente coinciden con los del hecho original. Completando de esta manera el circuito de recurrencia necesario para un razonamiento complejo.

Ejemplos de Descripción de MARCOS:

(marco con parte izquierda solamente)	[	Marco 1: "Es-Un-Trabajador-Precario" (Nombre, Cond. de Migración, Categ.Ocupacional, Residencia, Educación)
		Marco 2: "Es-Un-Trabajador-No-Calificado" (Cond.de Migración, Sexo, Cond. de Actividad, Escolaridad, Edad, Nombre del Jefe del Hogar, Relac. de Parentesco)

[	Marco 3: "Es-un-hijo-con-NBI" (Nombre, Sexo, Edad, Categ.Ocup.del Padre, Cond. de Actividad)
	REGLA: X1 Es-un-hijo-con-NBI (Nombre, Edad, Sexo, ---, ---)
	SI:
	X1 pertenece_a MARCO(2)
	CON: "Relac. de Parentesco = hijo"
	CON: "Nombre del jefe del hogar = X2"
Y:	
	X2 pertenece_a MARCO(1)

Ejemplos de descripción de hechos:

Hecho-Marco 1:

Nombre del hecho: "Juan" Factor de Certeza result.: 75 Marco al que pertenece: "Es-Un-Trabajador-Precario"		
Nombre del Atributo	Valor del atributo	Probabilidad asig.
Cond. Migración	Migr. de País Lim.	80
Categ. Ocupacional	Asalariado	30
Residencia	Bar. Marg. Urbano	75
Educación	Primar. Incompleto	90

Hecho-Marco 2:

Nombre del hecho: "Marcela" Factor de Certeza result.: 75 Marco al que pertenece: "Es-Un-Trabajador-No-Calificado"		
Nombre del Atributo	Valor del atributo	Probabilidad asig.
Cond. Migración	Migr. de País Lim.	75
Sexo	Femenino	30
Cond. Actividad	Inactivo	75
Escolaridad	Primar. Incompleto	90
Edad	+ de 40	40
Nombre del jefe del Hogar	Juan	
Relac. de Parentesco	Cónyuge	47

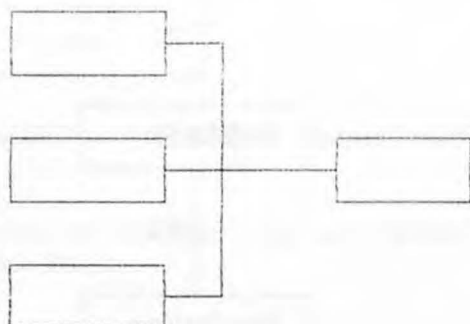
Los marcos 1 y 2 sólo contienen parte izquierda (hechos que corresponden a las bases de datos primarias). El marco 3 puede almacenar los hechos a los que da origen ordenándolos por "Nombre".

Obsérvese que el marco 3 genera hechos que difieren en los atributos respecto de los hechos que le dieron origen.

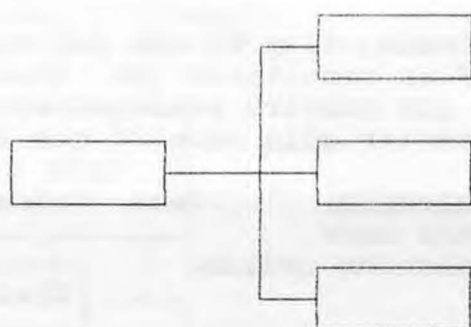
Descripción de las estructuras de datos utilizadas en la BASE de DATOS

Existen 4 tipos de básicos de relaciones entre estructuras de datos que a nosotros nos interesan, y que de acuerdo al tipo de relación que se establece entre los mismos se les asigna los nombres siguientes:

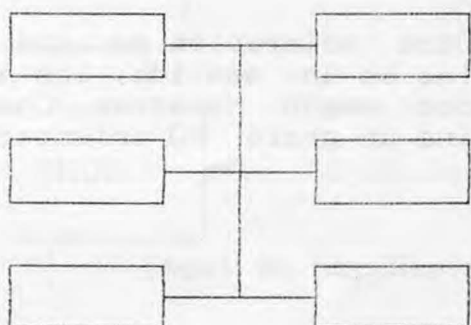
a) Relación: Muchos a Uno



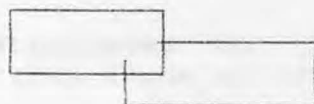
b) Relación: Uno a Muchos



c) Relación: Muchos a Muchos
(red de datos)



d) Relación: de Lazo o Recursiva

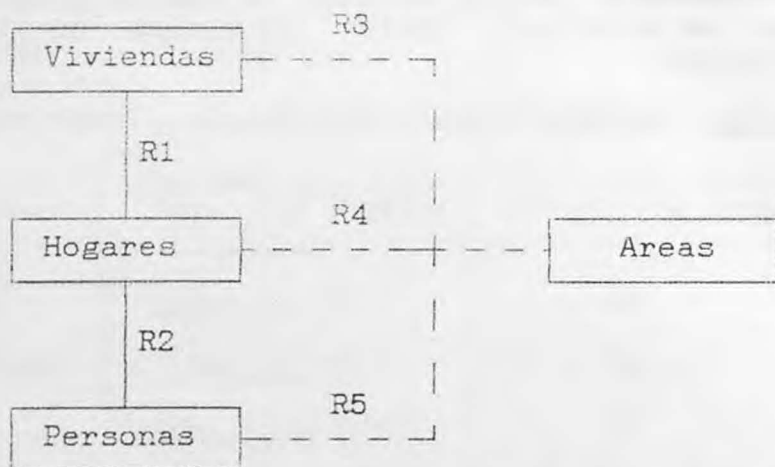


Entonces nuestras necesidades quedan basadas en un administrador de datos capaz de monitorear estos tipos de relaciones.

Convengamos primero que para hacer uso de las mismas, éstas se deben implementar a partir de relaciones binarias, siguiendo los 3 pasos básicos de "Normalización de Datos" propuestos por "CODD" (para mas detalles ver la obra: "Diseño e Implementación de Bases de Datos", de Gio Wiederhold, Ed. Mc.Graw-Hill).

Otro aspecto importante es el hecho que el usuario sólo verá la relación binaria que solicita, siéndole transparente el esquema general implementado.

Por ejemplo:



Donde R1 y R2 son del tipo b) y R3,R4,R5 son del tipo a). La Red Plex resultante (de tipo 'c' en este caso), es el producto final que nuestro administrador de datos debe manejar internamente y presentar sólo aquello que se le pida.

Por ejemplo: "...dame todas las personas que habitan en una vivienda dada."

Relación pedida:

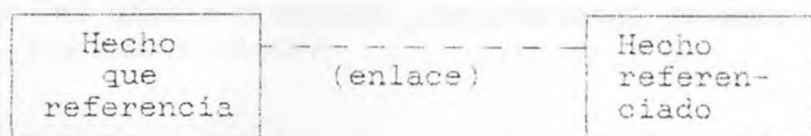


(el paso intermedio, o sea hogares, en este caso no se ve).

Otra aclaración es que si bien se describen las relaciones binarias en un sentido, los accesos se pueden realizar en ambos sentidos según nuestras necesidades (conformando los que se denomina un grafo "NO orientado" o "RED PLEX COMPLEJA").

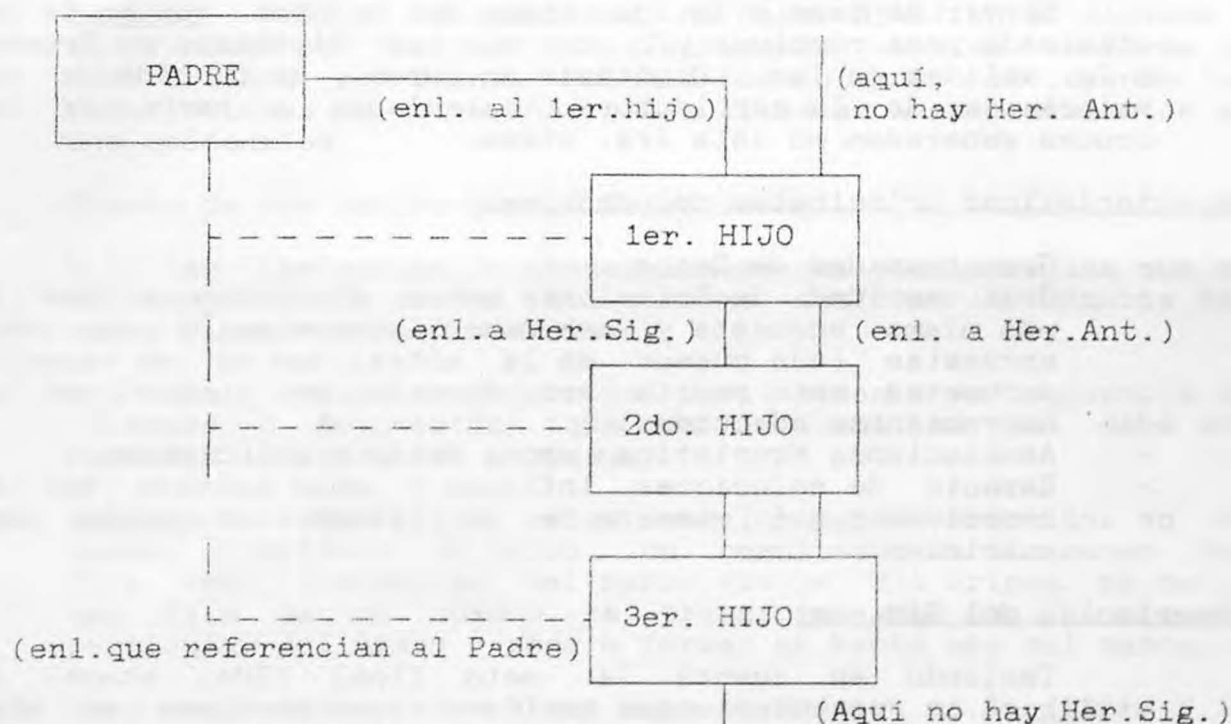
Descripción de los tipos de relaciones implementados: (para soportar la red de datos)

- a) Relación de REFERENCIA (equivale a una relación "MUCHOS a UNO").



- b) Relación Jerárquica o de "PADRE a HIJOS" (equivale a una relación de "UNO a MUCHOS").

En este caso la relación se arma a través de un campo en el hecho padre y tres campos en el hecho hijo.



V.- DESARROLLO DE ACTIVIDADES.

Las actividades han sido desarrolladas en base a las necesidades prioritarias que se presentaron desde el comienzo de la elaboración del sistema aquí propuesto.

V.1.- Primera etapa: (Esta etapa ya está en parte implementada y se encuentra en período de prueba y ajuste)

Objetivos:

Sistematizar la generación de Tabulados, basándonos en "reglas de agrupamiento de variables" de manera de obtener una coherencia en los distintos cruces realizados para las mismas y que éstas no sean arbitrarias, sino condicionadas por el

equipo de trabajo que tiene a su cargo el procesamiento y análisis.

Así definidos los grupos de variables, se permite a los potenciales "usuarios" (otras áreas), el acceso directo a los "cruces necesarios" en forma directa y sin intervención del equipo original (salvo para tratar inconsistencias de datos y/o de cruces).

Problema Principal:

Permitir la generación de cruces de variables de una manera formal y semi-automática, haciendo transparente al usuario la mayor cantidad posible de pasos para lograr su objetivo.

Sub-Problemas:

Servir de base a la 2da. etapa del sistema, que ha de ser utilizada para resolver y/o decidir las "Hipótesis en Prueba" y la validez de las "Hipótesis en curso" en función de una valoración de la certidumbre, calculadas a partir de los cruces generados en esta 1ra. etapa.

Características principales del problema:

- Gran cantidad de Datos.
- Gran cantidad de Relaciones entre distintos módulos de una misma encuesta y una menor interrelación con otras encuestas (aún cuando en la actualidad no se vinculan encuestas esto podría ser deseable de contar con las herramientas adecuadas).
- Asociaciones Heurísticas entre datos y soluciones.
- Espacio de soluciones infinito y solo acotado por las necesidades del momento y el tiempo disponible para cubrirlas.

Descripción del Sistema:

Teniendo en cuenta la meta final (2da. etapa) se utilizará un vocabulario que conlleva a estos fines, si bien queda aclarado que no es el mas adecuado en esta etapa del sistema.

El sistema global está basado en un "Modelo de Marcos" o "Frames" como más comumente se los denomina. La idea es que un marco es la descripción de una situación, dentro de un contexto dado, es decir representa "un punto de vista" de la realidad observada.

En otros términos, expresamos en los marcos nuestros "puntos de vista" sobre un mismo "cuerpo de conocimientos almacenado", según el momento y necesidades existentes.

Hecha la definición global. En esta 1ra. etapa, el objetivo principal es lograr que un "Plan de Tabulados" se realice de una manera;

- 1) Rápida,
- 2) Homogénea (sobre todo en el transcurso del tiempo),
- 3) Eliminando inconsistencias entre cruces, definiciones conceptuales, etc.

Estas premisas nos llevaron a definir el problema de la siguiente manera:

- a) Reglas: Son el conjunto de "restricciones" a utilizar en un marco, y pueden actuar sobre ninguna, una o más variables, (sería el equivalente a una estratificación de datos). Nos permite realizar restricciones de universo, de las variables dependientes o independientes, y la interrelación con procedimientos de cálculo (algoritmos externos) en caso necesario.
- b) Marcos: Es el agrupamiento de Mayor Nivel en el sistema y su función es "Definir" explícitamente como se conforma un cuadro línea por línea, columna por columna, incluyendo las restricciones de universos, y las llamadas a procedimientos si son necesarios.

Consta de dos partes bien diferenciadas:

- b.1) Las "instancias" o conocimientos que contiene, los que en el conjunto de marcos del sistema conforma la memoria del mismo o conocimiento acumulado.
- b.2) Las "reglas de deducción" utilizadas y su secuencia de "Disparo o Activación" (que es la forma en que debe ser resuelta la situación invocada).
- c) Hechos: En esta etapa, estos son los datos obtenidos en el cuadro al activar un marco, los resultados se almacenan por fila como "instancias" del marco que le dió origen, es decir una fila de un cuadro se transforma en una instancia particular del mismo y pasa a formar un hecho más del marco.

Podríamos esquematizar el sistema como se muestra en la figura V.1.

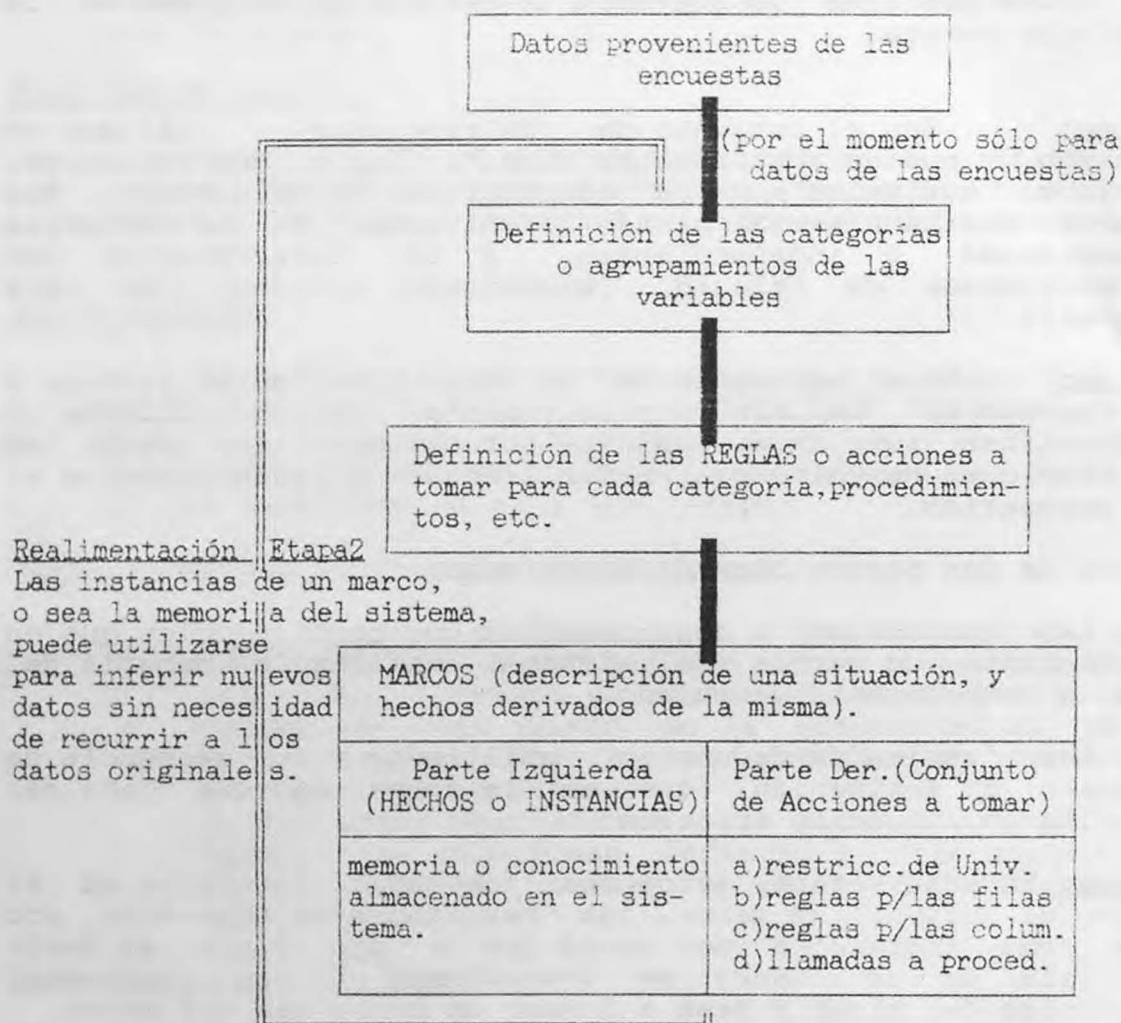
V.2.- Segunda Etapa:

Queda como tarea de investigación y de desarrollo para el período 1991/1992, la implementación final y completa del sistema propuesto. Cabe aclarar que la 2da. etapa se verá favorecida por la retro-alimentación de la 1ra., que ya está funcionando y que nos ha permitido "bajar" los tiempos de proceso, programación y análisis en una relación de 3 a 1, es decir ha permitido obtener resultados en la 3ra. parte de los tiempos que anteriormente se realizaban.

Cabe recordar como acotación que el objetivo final del sistema propuesto es que puedan coexistir datos provenientes de varias encuestas, varias ondas, y varios módulos y lograr en el futuro su interrelación y paulatina integración como un sólido bloque de

conocimientos a los cuales el usuario pueda preguntar y obtener respuestas en forma dinámica, conformando una suerte de Pin-Pong de preguntas y respuestas, entre el Usuario y el Sistema.

FIGURA V.1: Esquema en bloques del Sistema propuesto.



V.3.- Cronograma de Actividades:

Primera Etapa: 1989 - 1990

Segunda Etapa: 1991 - 1992

BIBLIOGRAFIA

1. ENRIQUE CASTILLO Y ELENA ALVAREZ, "Sistemas Expertos - Aprendizaje e Incertidumbre". Ed. Paraninfo, 1989. - Madrid - España.
2. Jose L. MATE HERNANDEZ Y JUAN PAZOS SIERRA, "Ingeniería del Conocimiento - Diseño y Construcción de Sistemas Expertos". Ed. SEPA, 1988 - Córdoba - Argentina.
3. HERMAN E. DOLDER, "Diseño de Bases de Datos utilizando conceptos y técnicas de Inteligencia Artificial". Ed. EUDEBA, 1987 - Bs. As. - Argentina.
4. MARVIN MINSKY, "La Sociedad de la Mente - La Inteligencia Humana a la luz de la Inteligencia Artificial". Ed. Galápagos, 1986 - Bs. As. - Argentina.
5. RAUL J. CARNOTA Y ALBERTO TESZKIEWICZ, "Sistemas Expertos y Representación del Conocimiento". EBAI/ESLAI, 1987 - Tandil - Argentina.
6. FRANCISCO S. NAISHTAT, "Lógica para computación". EUDEBA, 1986. Bs. As. - Argentina.
7. PHILIP R. ROBINSON, "Turbo - Prolog" (marca Registrada de Borland International). Ed. Osborne/Mc.Graw-Hill, 1987 - México.
8. JAMES MARTIN, "Organización de las Bases de Datos". Ed. Prentice Hall, 1977 - México.
9. GIO WIEDERHOLD, "Diseño e Implementación de Bases de Datos". Ed. Mc.Graw-Hill, 1985 - México.
10. BORGE LANGEFORS, "Teoría de los Sistemas de Información". Ed. El Ateneo, 1976/1982 - Bs. As. - Argentina.
11. J. P. AUBERT Y R. SCHOMBERG, "Inteligencia Artificial". Ed. Paraninfo, 1986 - Madrid - España.
12. RICHARD GAUTHIER Y STHEPEN D. PONTO, "Diseño de Programas para Sistemas". Ed. Paraninfo, 1984 - Madrid - España.