



**SEP**

**SEIT**

**Jgit**

---

INSTITUTO TECNOLÓGICO DE MINATITLÁN

---

INGENIERÍA EN SISTEMAS COMPUTACIONALES

T E S I S

**" IMPLEMENTACIÓN DEL MÉTODO DE  
BÚSQUEDA UTILIZADO EN EL DESARROLLO  
DE UN SISTEMA EXPERTO AUXILIAR EN EL  
DIAGNÓSTICO DE CÁNCER EN NIÑOS."**

**PARA OBTENER EL TÍTULO DE  
INGENIERO EN SISTEMAS COMPUTACIONALES**

**PRESENTA**

José David Montaña Gómez

**ASESOR**

M. C. Ana Estela Ruiz Linares

MINATITLÁN, VERACRUZ

2003

# *Dedicatoria A Mis Padres*

*Porque gracias a su cariño, guía y apoyo he llegado a realizar uno de los anhelos más grandes de mi vida, fruto de inmenso apoyo, amor y confianza que en mi se depositó y con los cuales he logrado terminar mis estudios profesionales que constituyen el legado más grande que pudiera recibir y por el cual les viviré eternamente agradecido.*

# *Agradecimientos*

## **A DIOS**

Por permitirme vivir cada día, estar siempre conmigo y brindarme todo lo que soy y todo lo que tengo.

## **A MIS PADRES**

Que con su infinito amor, apoyo y comprensión he podido seguir adelante y conseguir lo hasta ahora logrado.

## **A MI ASESOR M.C. ANA ESTELA RUIZ LINARES**

Por su paciencia y dedicación.

## **A MIS AMIGOS**

Por ser comprensivos y ayudarme a salir adelante en los momentos difíciles.

## **AL DR. ARTURO FAJARDO GUTIÉRREZ**

Por los consejos, tiempo dedicado y empuje que nos motivo a seguir adelante.

## **A CONACYT-SIGOLFO**

Por el apoyo y facilidades que en todo momento nos proporcionaron.

## **AL DR. ROBERTO RODRÍGUEZ GARCÍA**

Por sus conocimientos y experiencias transmitidas.

## **A RODRIGO CARRASQUEDO**

**VELÁZQUEZ**

Por toda la confianza, enseñanzas y experiencias que me permitieron reflexionar y tomar decisiones sobre mi futuro.

# INDICE GENERAL

<b>Introducción</b>	<b>1</b>
Objetivos	4
Objetivo General	4
Objetivos Particulares	4
Justificación	5
Hipótesis	6
Alcances y limitaciones	7
Generalidades	9
<b>Capítulo I Antecedentes.</b>	<b>12</b>
1.1 Historia de la inteligencia artificial.	13
1.2 ¿Qué es la Inteligencia Artificial?	14
1.2.1 Base de datos, retroseguimiento y procesamiento de listas.	16
1.3 Sistemas expertos.	17
1.3.1 ¿Qué es un sistema experto?	17
1.3.2 ¿Cuáles son las ventajas de un sistema experto?	18
1.3.3 Algunos ejemplos de sistemas expertos comerciales.	19
1.3.4 Arquitectura de los sistemas expertos.	20
1.3.5 Cómo funciona un sistema experto.	20
1.3.5.1 La base de conocimiento.	21
1.3.5.2 El motor de inferencia.	21
1.3.5.2.1 El método de encadenamiento hacia adelante.	22
1.3.5.2.2 El método de encadenamiento hacia atrás.	22
1.3.5.2.3 El método de reglas de producción.	23
1.4 Técnicas de búsquedas	23
1.4.1 La técnica de búsqueda del primero en profundidad.	24
1.4.1.1 Evaluación de la búsqueda del primero en profundidad.	24
1.4.2 La técnica de búsqueda de primero en anchura.	25
1.4.2.1 Evaluación de la búsqueda del primero en anchura.	25
1.4.3 Adición de la información heurística.	26
1.4.3.1 La técnica de búsqueda de la escalada de la colina.	27
1.4.3.1.1 Evaluación de la de búsqueda de la escalada de la colina.	27
1.4.3.2 La técnica de búsqueda del menor coste.	28
1.4.3.2.1 Evaluación de la búsqueda del menor coste.	28
1.5 Lógica e incertidumbre.	28
1.5.1 Lógica proposicional.	29
1.5.2 Cálculo de predicados.	29
1.5.3 Incertidumbre.	30
1.5.4 Lógica incierta o difusa.	31
1.5.5 Sistemas probabilísticos.	31

1.5.6	Teoría probabilística clásica.	32
1.5.7	El método del eslabón más débil.	33
1.5.8	El método del eslabón más fuerte.	34
<b>Capítulo II Diseño del Sistema Experto.</b>		<b>35</b>
2.1	Diseño de la solución.	36
2.1.1	Desarrollo de los sistemas expertos.	36
2.1.1.1	El equipo de desarrollo.	36
2.1.1.2	Construcción de prototipos (Rapid Prototyping).	39
2.1.2	Diseño del prototipo (Rapid Prototyping).	40
2.1.2.1	Diseño del procesamiento de lenguaje natural y la base de conocimientos.	46
2.1.2.1.1	Elección del método para el sistema probabilístico.	48
2.1.2.2	Diseño del mecanismo de búsqueda.	49
2.1.2.2.1	Elección del método para el mecanismo de búsqueda	50
2.1.2.2.2	Elección de método para el mecanismo de inferencia	50
<b>Capítulo III Desarrollo del Sistema Experto.</b>		<b>54</b>
3.1	Propuesta de solución.	55
3.1.1	¿Por qué usar C/C++ para la programación en inteligencia artificial?	55
3.2	Desarrollo del mecanismo de inferencia.	58
<b>Capítulo IV Resultados.</b>		<b>64</b>
<b>Conclusiones</b>		<b>73</b>
<b>Anexo A</b> Proyecto: “EPIDEMIOLOGÍA DE LAS NEOPLASIAS MALIGNAS EN NIÑOS DERECHOHABIENTES DEL INSTITUTO MEXICANO DEL SEGURO SOCIAL”.		<b>76</b>
<b>Anexo B</b> Resultados Proyecto: “EPIDEMIOLOGÍA DE LAS NEOPLASIAS MALIGNAS EN NIÑOS DERECHOHABIENTES DEL INSTITUTO MEXICANO DEL SEGURO SOCIAL”.		<b>84</b>
<b>Anexo C</b> Cartas compromiso de usuarios y reconocimientos		<b>89</b>
<b>Glosario</b>		<b>100</b>
<b>Bibliografía</b>		<b>103</b>

# *Introducción*

El problema del cáncer está tomando tintes alarmantes. Estudios estadísticos revelan una tendencia al incremento del cáncer en niños provenientes de los países desarrollados correspondiente al 1% anual. Estos mismos indicadores muestran que, en contra parte, la tasa de mortalidad debida a esta enfermedad observa una tendencia a la baja considerable. Ejemplo de esto se presenta en la población infantil de los Estados Unidos. La mortalidad por cáncer registra una disminución en promedio del 58%. Países como Canadá, países de Europa y Asia, muestran resultados similares. En cambio en países africanos, India y países latinoamericanos no se presenta el mismo impacto.<sup>4</sup>

Lo anterior se ha logrado por la creación de grupos cooperativos, los cuales registran a la mayoría de los niños con cáncer. Por mencionar algunos, los dos grandes grupos cooperativos en Estados Unidos, *Children Cancer Study Group* y el *Pediatric Oncology Group* registran aproximadamente el 95% de los niños con cáncer.

Estos grupos proporcionan un tratamiento integral, además de efectuar un diagnóstico lo más temprano posible. Por lo tanto se puede decir que la disminución de la mortalidad en los países donde se han establecido grupos cooperativos descansa en dos aspectos importantes:

- a) Tratamiento integral del niño con cáncer.
- b) Diagnóstico temprano.

---

<sup>4</sup> Fajardo Arturo. Epidemiología descriptiva del cáncer en el niño.



Debido a esto, un diagnóstico temprano en un niño con cáncer resulta de importancia crucial para tener mayores probabilidades de un tratamiento integral efectivo y erradicar la enfermedad del paciente.

En los países subdesarrollados existen muchos problemas para la atención integral del niño con cáncer. En primer lugar no se cuenta con registros integrales y confiables que abarquen el ámbito nacional, por lo cual se desconoce el impacto originado por el cáncer en la población pediátrica. Asimismo y posiblemente debido al desconocimiento de los resultados actuales del tratamiento del niño con cáncer, en los cuales se ha incrementado la sobrevida e incluso para algunos cánceres se habla de curación, frecuentemente los médicos, quienes no los tratan directamente (dígase médicos generales, médicos familiares, pediatras, etc.) tengan una actitud fatalista ante los niños con este tipo de enfermedades.<sup>5</sup>

---

<sup>5</sup> Fajardo Arturo. Epidemiología descriptiva del cáncer en el niño.

## **OBJETIVOS**

### **OBJETIVO GENERAL**

- Implementar en C++ un método de búsqueda utilizado en un sistema experto de diagnóstico médico.

### **OBJETIVOS ESPECÍFICOS**

- Contribuir en el desarrollo de un sistema experto orientado en la detección temprana de cáncer en niños.
- Coadyuvar en la vinculación del Instituto Tecnológico de Minatitlán con el sector salud para el desarrollo de proyectos de investigación tecnológica.

## JUSTIFICACIÓN

Poco se sabe acerca de las neoplasias malignas en la niñez, sin embargo ha habido una tendencia al incremento en los últimos años, siendo en nuestro país la segunda causa de muerte en los niños de 5 a 14 años.<sup>6</sup> El impacto del cáncer en esta edad es mucho mayor que en el adulto dado que se tiene una mayor pérdida de años de vida en los casos fatales, el efecto emocional en los padres y hermanos y la posibilidad de ruptura de la vida familiar en la mayoría de los casos. En los sobrevivientes, la ausencia escolar puede dañar las perspectivas futuras y el tratamiento conducir a la pérdida de la fertilidad y efectos adversos al crecimiento. En adición y aunque la etiología del cáncer en la niñez es desconocida, hay evidencia de que la exposición a ciertos factores ambientales como los derivados del benceno y los pesticidas, así como la exposición a campos electromagnéticos, en unión a factores como la ocupación de los padres puede ocasionar cáncer en la población pediátrica. Esto implica que el cáncer en la niñez puede ser prevenible.

El correcto y eficiente diseño del método de búsqueda resulta de importancia crucial ya que por medio de él se construye y da seguimiento al conjunto de preguntas para la investigación de un posible caso de cáncer. Además proporciona el análisis probabilístico de la orientación hipotética del tipo de cáncer que se puede presentar. Esto es lo que permite al sistema experto emitir al diagnóstico al usuario.

---

<sup>6</sup> Fajardo Arturo. Epidemiología descriptiva del cáncer en el niño.

## Hipótesis

El diseño e implementación del método de búsqueda a través de la técnica de las reglas de producción construido a partir del método de encadenamiento hacia atrás permitirá al sistema experto la eliminación de la mayor incertidumbre posible para la construcción del árbol de búsqueda y el camino solución a través del conjunto de preguntas de la sintomatología que presente el paciente, orientadas al probable caso de cáncer. Además apoyado en los sistemas probabilísticos para el tratamiento de la incertidumbre y basado en el método del eslabón más fuerte permitirá la adición de heurística a través de un análisis porcentual de los síntomas en relación con el tipo de cáncer, lo cual permitirá al sistema emitir un diagnóstico veraz y oportuno.

## Alcances y limitaciones

El desarrollo de este sistema realiza un aporte en el campo de la medicina al proporcionar una herramienta para el diagnóstico temprano del cáncer en niños. Una restricción importante es que el diagnóstico emitido por el sistema está basado en la probabilidad arrojada a partir del estudio de los síntomas que el paciente presenta en relación con los existentes en la base de conocimientos de otros casos de cáncer, es decir, está marcado por la similitud de síntomas presentados por casos anteriores de pacientes con cáncer comprobados y analizados por médicos especialistas en el área. Esto nos indica que los factores con incidencia en el desarrollo del cáncer, tales como los aspectos propios del área geográfica en donde viva el paciente, por ejemplo el origen del agua que toma puede ser de un pozo o río que pudiera estar contaminado, si vive cerca de industrias y en consecuencia pueda tener contacto con alguna sustancia o agente cancerígeno; todavía no se encuentra definido su nivel influencia en la aparición de cáncer en el paciente, además estas condiciones pueden variar a partir de diferencias en el estatus social, lugar de residencia, lugar de nacimiento, malformaciones congénitas, etc. A través de esta herramienta se pretende estudiar el grado de incidencia de estos y otros factores, que posteriormente podrían ser tomados en cuenta para un diagnóstico oportuno de esta enfermedad. De esta forma los alcances del sistema se podrían expandir, aunque de la misma manera estaría regido por la capacidad del médico para observar las tendencias a las que los factores de influencia se dirigen y al campo de acción en el que este se desenvuelva.

Debido a la naturaleza del problema y al entorno en que éste se desarrolla, el sistema sólo podrá ser utilizado por personas con conocimientos en medicina ya que en su mayoría el lenguaje utilizado en el mismo es de corte clínico, por lo cual algunos términos serían desconocidos para personas con pocos conocimientos en esta área, y provocaría posibles errores en la información proporcionada al sistema, con lo cual el índice de veracidad ofrecido en el diagnóstico se vería seriamente afectado.

## Generalidades

La “Implementación del método de búsqueda utilizado en el desarrollo de un sistema experto auxiliar en el diagnóstico de cáncer en niños”, forma parte del proyecto de investigación “Desarrollo de un sistema experto auxiliar en el diagnóstico de cáncer en niños” clave 99-05-001-V que fue apoyado y financiado por CONACYT – SIGOLFO en el periodo comprendido entre 1999 y 2001.

Este trabajo tiene como finalidad presentar el empleo de un sistema experto como un método de apoyo para el área médica en la detección temprana del cáncer en la población infantil. Este documento muestra la forma en que es estructurado y desarrollado el mecanismo de búsqueda, que es parte del sistema experto. A manera de complemento se analiza la estructura de la base de conocimientos y el procesamiento de lenguaje natural debido a que el mecanismo de búsqueda se apoya en ellos para lograr su óptimo desempeño.

En el primer capítulo, se introducirán algunos conceptos y teorías empleadas en el desarrollo de aplicaciones de inteligencia artificial tal es el caso de los métodos y mecanismos de búsqueda. Analizaremos las técnicas de búsqueda con la inclusión de un método heurístico que es esta característica la cual, en definitiva hace marcada la diferencia entre una aplicación de software tradicional y una aplicación de software que emplea técnicas de inteligencia artificial. Estudiaremos las técnicas de búsqueda heurística de escalada de la colina y la del menor coste. Se definirá que es un sistema experto, sus características y elementos que lo componen, además de la influencia de la lógica e incertidumbre sobre la implementación de

problemas reales en mecanismos y algoritmos informáticos que permitan resolver estos por medio de una computadora.

En el segundo capítulo describiremos el diseño conceptual proporcionado al sistema experto para la detección del cáncer en sus etapas tempranas. Analizaremos cómo se diseña un sistema experto y los mecanismos que ayuden a la detección del cáncer. Se explicará que técnicas y métodos fueron empleados para el diseño del sistema experto, por qué se eligieron dichas técnicas y qué criterios de evaluación se emplearon para seleccionarlas. Se explicará el proceso de desarrollo de sistemas expertos así como los módulos que intervienen, desde el equipo de expertos en el campo de estudio e ingenieros del conocimiento y el rol que cada uno de ellos juega en el proceso de desarrollo, hasta herramientas empleadas para la implementación de los mismos

En el tercer capítulo se desglosará la parte medular del código comprendida por los objetos, atributos, métodos y funciones, que hacen posible la búsqueda de la solución. Se estudiará la herramienta y el lenguaje de programación utilizados para el desarrollo del Sistema Experto y cuáles fueron los factores que motivaron la elección del lenguaje.

El cuarto capítulo muestra el estado actual que observa el sistema experto, su funcionamiento a fondo y el resultado que se obtiene a partir del procesamiento de información por el mecanismo de inferencia que deriva en el análisis de probabilidad que se tiene de que el paciente presente algún tipo de cáncer, es decir el diagnóstico final arrojado por el experto.



En la sección de anexos se presenta parte de la investigación realizada en el Centro Médico Siglo XXI con el proyecto: “EPIDEMIOLOGÍA DE LAS NEOPLASIAS MALIGNAS EN NIÑOS DERECHOHABIENTES DEL INSTITUTO MEXICANO DEL SEGURO SOCIAL”.

En el anexo A se observa el cuestionario llevado a cabo para recabar la información del paciente de entre los cuales se destaca el historial de la sintomatología presentada por el paciente a lo largo del posible desarrollo del cáncer, el anexo B muestra fragmentos esenciales de los resultados estadísticos obtenidos, los que derivaron en la construcción de la base de conocimientos del sistema experto así como del mecanismo para la búsqueda de soluciones.

En el anexo C se muestran las cartas compromisos de los usuarios potenciales que participaron en el proyecto. Así también se incluyen las copias de las constancias de entrega del prototipo a los usuarios comprometidos. Se incluye una copia de la portada de la Revista Médica de Pediatría en donde este trabajo se menciona en la carta al editor; y finalmente el reconocimiento al primer lugar fase regional del XVI Concurso Nacional de Creatividad de los Institutos Tecnológicos.

# Capítulo I

## *Antecedentes*

En la actualidad la inteligencia artificial abarca una enorme cantidad de subcampos, desde áreas de propósito general, como es el caso de la percepción y del razonamiento lógico, hasta tareas específicas, como el ajedrez, la demostración de teoremas matemáticos, la poesía y el diagnóstico de enfermedades. La inteligencia artificial es un campo genuinamente universal.

### 3.1 Historia de la inteligencia artificial.

La inteligencia artificial tuvo sus inicios hacia 1960, cuando en el MIT, John McCarthy creó el LISP, el primer lenguaje de investigación dentro de la inteligencia artificial. El término *inteligencia artificial* suele atribuírsele a Marvin Minsky, también del MIT, quien en 1961 escribió un artículo titulado “Hacia la inteligencia artificial”. En los sesenta surgió la primera computadora que jugaba al ajedrez, las primeras pruebas matemáticas informatizadas y el conocido programa ELIZA escrito en 1964 por Joseph Weizenbaum del MIT, este programa actuaba como un psicoanalizador rogeriano.

Hacia finales de los setenta, muchos éxitos como el procesador de lenguaje natural, la representación del conocimiento y la resolución de problemas; habían sido alcanzadas en áreas específicas de la inteligencia artificial. Estos éxitos formaron la base para la introducción de uno de los primeros productos comerciales de la inteligencia artificial: el *sistema experto*. Un sistema experto es un programa que contiene información sobre un cierto campo y que, cuando se le pregunta, responde igual que un humano experto. Uno de los primeros sistemas

de este tipo fue MYCIN, desarrollado en la Universidad de Stanford, para ayudar a los médicos en el diagnóstico de enfermedades.<sup>7</sup>

En 1972 Alain Colmerauer creó el PROLOG en Marsella, Francia. Al igual que LISP, Prolog era un lenguaje diseñado para ayudar a resolver problemas relativos a la inteligencia artificial; posee una base de datos incorporada y una sintaxis bastante simple. En 1981, los japoneses anunciaron que usarían Prolog como base de sus computadoras de la “quinta generación”.

### 3.2 ¿Qué es la Inteligencia Artificial?

Un gran número de autores definen a la inteligencia artificial en torno a dos dimensiones principales, algunas se refieren a **procesos mentales** y el **razonamiento**, en tanto que otros lo hacen aludiendo a la **conducta**. Por otra parte algunas definiciones miden la condición deseable en función de **eficiencia humana**, mientras otras lo hacen en conformidad con un concepto de inteligencia artificial denominado **racionalidad**.<sup>8</sup>

Se considera que un sistema es racional si hace lo correcto. De acuerdo con lo anterior en la inteligencia artificial existen cuatro posibles objetivos por alcanzar, como se comenta en la *figura 1*.

Sistemas que piensan como humanos.	Sistemas que piensan racionalmente.
Sistemas que actúan como humanos.	Sistemas que actúan racionalmente.

Figura 1. Algunas definiciones de Inteligencia Artificial. Se agrupan en cuatro categorías.

<sup>7</sup> Hebert Schildt. Utilización de C en la inteligencia artificial.

<sup>8</sup> Russell Stuart. Norving Peter. Inteligencia Artificial. Un enfoque moderno.

A lo largo de la historia se han adoptado los cuatro enfoques anteriormente mencionados. El enfoque centrado en el comportamiento humano constituye una ciencia empírica, que entraña el empleo de hipótesis y de la confirmación mediante experimentos. El enfoque racionalista combina matemáticas e ingeniería.

Aunque el campo de la inteligencia artificial es amplio, los temas fundamentales son:

- Búsqueda (de soluciones).
- Sistemas expertos.
- Incertidumbre y “lógica difusa”.
- Procesamiento de lenguaje natural.
- Robótica.
- Lógica.
- Aprendizaje de las máquinas.
- Visión y reconocimiento de patrones.
- Reconocimiento de modelos.

Algunas de las áreas representan aplicaciones finales, tales como los sistemas expertos; otras como el procesamiento del lenguaje natural y búsqueda de soluciones, son bloques de la inteligencia artificial que se añaden a otros programas para llevar a cabo su realización.

Cuando se hace referencia a la inteligencia artificial, el término *búsqueda* se refiere a la búsqueda de soluciones a un problema, no significa encontrar una información específica dentro de una base de datos.

Un sistema experto tiene dos atributos principales. Primero, le permite introducir información sobre un tema en una computadora. A esta información se le suele llamar *base de conocimiento*. En segundo lugar, le permite interrogar a esta base de conocimiento y luego actúa como si fuese un experto en la materia, que es en definitiva la razón de su nombre.<sup>9</sup>

El procesamiento del lenguaje natural permite a la computadora la entrada del lenguaje humano de forma directa. El mayor obstáculo para alcanzar esta meta es el tamaño y la complejidad de los lenguajes humanos. Además, tenemos el problema de que la computadora sea consciente de la información contextual, la cual puede aparecer en cualquier situación que no sea de las más simples.

### 3.2.1 Base de datos, retroseguimiento y procesamiento de listas.

La mayoría de los algoritmos basados en inteligencia artificial necesitan fundamentalmente de una o más bases de datos, un método de retroseguimiento y un método de procesamiento de listas. Puede que no se vea la importancia de estas tres tareas; sin embargo, estas tareas son tan importantes para la mayoría de los algoritmos de inteligencia artificial como lo son por ejemplo las sentencias de control de bucle para la mayoría de las rutinas convencionales.

---

<sup>9</sup> Hebert Schildt. Utilización de C en la inteligencia artificial.

Un procedimiento basado en inteligencia artificial utiliza bases de datos para mantener las distintas informaciones que recibe u obtiene. El retroseguimiento es por definición más complejo. Aporta una rutina para encontrar la solución de un problema. La capacidad de retroseguimiento es importante para muchas de las rutinas de inteligencia artificial porque permite que un algoritmo busque una solución a un problema siguiendo varios caminos o líneas de razonamiento. Si la rutina encuentra un punto muerto, simplemente vuelve atrás hacia el punto más cercano del procesamiento e intenta otra vía. El retroseguimiento es básicamente una operación de almacenamiento. Los primeros investigadores en inteligencia artificial valoraron la posibilidad del procesamiento de listas y creyeron que era fundamental para crear inteligencia. Como muestra Prolog las listas, aun siendo importantes, han sido reemplazadas por bases de datos dinámicas en muchas aplicaciones. Sin embargo, las listas son importantes todavía para diversos tipos de rutina.<sup>10</sup>

### 3.3 Sistemas expertos.

#### 3.3.1 ¿Qué es un sistema experto?

Los sistemas expertos son programas que imitan el comportamiento de un humano. Utilizan la información que el usuario le proporciona para darle una opinión sobre cierta materia. Por tanto, el sistema experto le hace preguntas hasta que pueda identificar un objeto que relacione con sus respuestas. Un sistema experto intenta dar consejo a su usuario sobre el tema que conoce.

---

<sup>10</sup> Hebert Schildt. Utilización de C en la inteligencia artificial.

“Un sistema experto es una aplicación informática que soluciona problemas complicados que de otra manera exigirían ampliamente la pericia humana. Para lograr esto, se simula el proceso de razonamiento humano mediante la aplicación específica de *conocimientos* y de *inferencias*.”<sup>11</sup>

Internamente, un sistema experto ideal se puede caracterizar como un sistema que comprende:

- Amplio conocimiento específico a partir del campo de interés.
- Aplicación de técnicas de búsqueda.
- Soporte para Análisis Heurístico.
- Habilidad para inferir nuevos conocimientos a partir de conocimientos ya existentes.
- Procesamiento de símbolos.
- Capacidad para explicar su propio razonamiento.

### 3.3.2 ¿Cuáles son las ventajas de un sistema experto?

El atractivo de un sistema experto es fundamentalmente su disponibilidad y conveniencia. Tal vez la característica más fascinante y poderosa de los sistemas expertos, que los distingue de la mayoría de las aplicaciones tradicionales de la computación, es su capacidad para enfrentar problemas que constituyen un reto del mundo real, por medio de la aplicación de procesos que reflejan el discernimiento y la intuición humanas.

---

<sup>11</sup> Rolston D. W. Inteligencia artificial y sistemas expertos.



### 3.3.3 Algunos ejemplos de sistemas expertos comerciales.

MYCIN fue el primer sistema experto de éxito mundial. Desarrollado en la Universidad Stanford en la mitad de los setenta, fue diseñado para ayudar a los médicos a diagnosticar algunas enfermedades bacterianas. El diagnóstico de enfermedades es en esencia una tarea basada en la comparación de los síntomas mostrados por un paciente con los síntomas propios de una enfermedad hasta que se encuentra la relación. El problema es que resulta difícil para un médico diagnosticar rápida y confiadamente todas las innumerables enfermedades que existen.

El PROSPECTOR, que fue creado en 1978 por Richard Duda, Peter Hard y Rene Reboh. PROSPECTOR es un experto en geología: predice la posibilidad de que ciertos depósitos de minerales se encuentren en una región en particular. Ha habido distintas variaciones en este programa, incluyendo programas que pueden predecir el descubrimiento de petróleo, gas natural y helio.

El comienzo de la década de los ochenta vio la introducción de sistemas expertos que podían proporcionar consultas de impuestos, consejos sobre seguros y ayuda legal.

### 3.3.4 Arquitectura de los sistemas expertos.

Los sistemas expertos emplean una amplia variedad de arquitecturas específicas en sus sistemas, principalmente porque una arquitectura es más aplicable que otra cuando se considera una aplicación dada. Actualmente se lleva a cabo una vasta investigación para estudiar diferentes aspectos de las arquitecturas de sistemas expertos y subsiste un considerable debate al respecto.<sup>12</sup> A pesar de las diferencias significativas, la mayoría de las arquitecturas tienen muchos componentes en común. La *figura 2* muestra una arquitectura general con los componentes típicos.

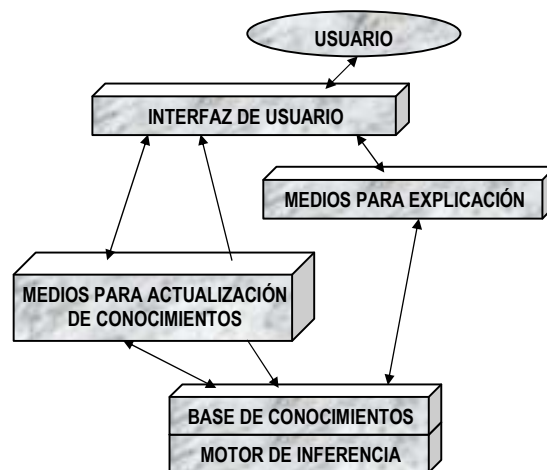


Figura 2. Arquitectura Típica de un Sistema Experto.

### 3.3.5 Cómo funciona un sistema experto.

Todos los sistemas expertos básicamente tienen dos partes: la *base del conocimiento* y el *motor de inferencia*.

<sup>12</sup> Rolston D. W. Inteligencia artificial y sistemas expertos.

### 3.3.5.1 La base de conocimiento.

La base de conocimiento es una base de datos que posee una información y unas reglas específicas sobre una materia determinada. Existen dos conceptos relacionados con una base de conocimientos.

- *Objeto*: La conclusión que es definida por sus reglas asociadas.
- *Atributo*: Una cualidad específica que, con su regla, ayuda a definir el objeto.

Por tanto, considere la base de conocimiento como una lista de objetos con sus reglas y atributos asociados. En el sentido más simple y para muchas aplicaciones, la regla que se aplica a un atributo determina si un objeto tiene o no tiene dicho atributo. Así pues se puede definir un objeto usando una lista de atributos que el objeto posea o no.

### 3.3.5.2 El motor de inferencia.

El motor de inferencia es la parte del sistema experto que intenta utilizar la información que le es suministrada para encontrar un objeto que le encaje. Hay dos categorías muy amplias sobre los motores de inferencia: la determinista y la probabilista. La mayoría de las disciplinas no son deterministas, sino que pueden considerarse probabilistas hasta cierto punto. No obstante, para muchas de éstas, el factor de incertidumbre no es estadísticamente importante, de manera que se pueden tratar como situaciones deterministas.

Hay tres formas básicas de construir el motor de inferencia: encadenamiento hacia adelante, encadenamiento hacia atrás y mediante reglas de producción. Las diferencias entre estos métodos se relacionan con la forma en que su motor de inferencia intenta alcanzar su meta.

#### **3.3.5.2.1 El método de encadenamiento hacia adelante.**

Al encadenamiento hacia adelante se llama algunas veces *conducido por datos* porque el motor de inferencia utiliza la información que el usuario le proporciona para moverse a través de una red de Ies griegas (ANDs) y Oes (ORs) lógicos hasta que encuentra un punto terminal, que es el objeto. Si el motor de inferencia no puede encontrar un objeto usando la información existente, entonces le pide más. Los atributos que definen al objeto crean un camino que conduce al mismo objeto: la única forma de alcanzar dicho objeto es la de satisfacer todas sus reglas. Por tanto, un motor de inferencia de encadenamiento hacia adelante comienza con alguna información y luego intenta encontrar un objeto que encaje con dicha información.

#### **3.3.5.2.2 El método de encadenamiento hacia atrás.**

Un motor de inferencia de encadenamiento hacia atrás comienza con una hipótesis (un objeto) y pide más información para confirmarlo o negarlo. Al encadenamiento hacia atrás se le llama algunas veces *conducido por objetos* porque el sistema experto empieza con un objeto e intenta verificarlo. El encadenamiento hacia atrás poda un árbol. Este es el proceso opuesto al del encadenamiento hacia adelante, que construye un árbol.

### 3.3.5.2.3 El método de reglas de producción.

Un motor de inferencia de las reglas de producción es teóricamente superior tanto a un sistema de encadenamiento hacia adelante como hacia atrás porque pide la información que tenga una mayor importancia de acuerdo con el estado actual del sistema. Un motor de inferencia de reglas de producción es en realidad, un motor de encadenamiento hacia atrás mejorado. La teoría operativa general es que el sistema pide como nueva información aquella que *elimine la mayor incertidumbre posible del sistema*. Para lograr esto el método requieren que la base de conocimiento contenga no sólo la información objeto-atributo estándar, sino también un valor cuantificador.

## 3.4 Técnicas de búsqueda.

Hay varias técnicas para la búsqueda de una posible solución. Estas son las más comunes e importantes:

- Primero en profundidad.
- Primero en anchura.
- Escalada de la colina.
- Menor coste.

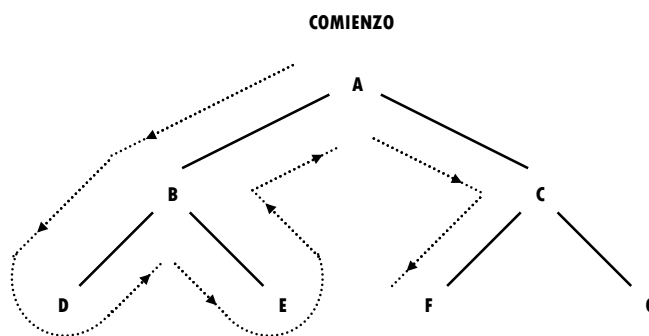


Figura 3. Diagrama de la búsqueda del primero en profundidad.

### 3.4.2 La técnica de búsqueda del primero en anchura

La técnica de búsqueda del primero en anchura es la opuesta a la de profundidad. La búsqueda del primero en anchura garantiza el encontrar una solución si existe, porque puede que en algunos casos degenera en una búsqueda exhaustiva. La *figura 4* muestra el funcionamiento de la búsqueda del primero en anchura.

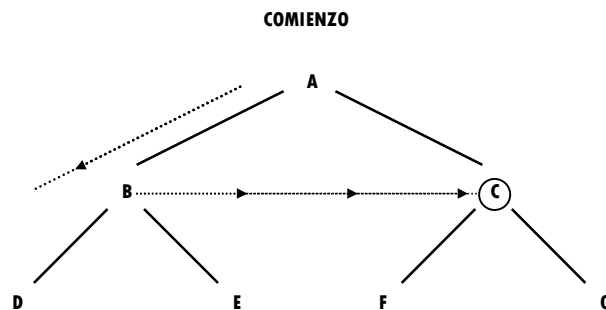


Figura 4. Diagrama de la búsqueda del primero en anchura

#### 3.4.2.1 Evaluación de la búsqueda del primero en anchura.

Encontrará una dificultad en la búsqueda del primero en anchura cuando intente encontrar una meta que esté situada varios niveles más abajo. En este caso, la búsqueda del primero en anchura derrocha un esfuerzo considerable para encontrarla. Generalmente, el programador elige entre la búsqueda del primero en profundidad y en anchura en base a la suposición de la situación más probable a la meta.

### 3.4.3 Adición de la información heurística.

Hasta este punto puede que haya supuesto que las búsquedas en profundidad y en anchura sean rutinas ciegas. Buscan una solución basándose solamente en el movimiento desde una meta a otra sin uso de una suposición inteligente. Mientras que este proceso puede ser correcto para ciertas situaciones controladas por un programador, posee la información que le aconseja el uso de un método sobre otro, lo que un programa generalizado de inteligencia artificial necesita es un proceso de búsqueda que es, normalmente, superior a cualquiera de estas dos técnicas. La única manera de alcanzar tal tipo de búsqueda es la adición de capacidades heurísticas.

La heurística son reglas simples que habilitan la posibilidad de que una búsqueda proceda en una dirección correcta. La información heurística, aunque en ocasiones imprecisa y no garantizada, incrementa las posibilidades de que un método de búsqueda encuentre una meta rápidamente de manera óptima, o ambas a un mismo tiempo. En definitiva, incrementan las posibilidades de un rápido éxito. El uso de heurísticas contribuye grandemente a la *potencia* y *flexibilidad* de los sistemas expertos y tiende a distinguirlos aún más del software tradicional.

Puede pensar que podría insertar fácilmente heurística en programas que estuvieran diseñados para unas aplicaciones específicas, pero sería imposible crear búsquedas heurísticas generalizadas.



Muy frecuentemente, los métodos de búsqueda heurísticos se basan en exagerar o minimizar algunos aspectos del problema. Las dos aproximaciones heurísticas que se examinarán, utilizan un procedimiento opuesto y encuentran resultados diferentes. Estos dos sistemas están basados esencialmente en rutinas de búsqueda del primero en profundidad.

#### 3.4.3.1 La técnica de búsqueda de la escalada de la colina.

La ruta más corta no implica necesariamente un menor número de conexiones. El algoritmo de escalada de la colina escoge como próximo paso el nodo que aparezca en el lugar más cercano a la meta.

##### **3.4.3.1.1 Evaluación de la búsqueda de la escalada de la colina.**

En muchos casos, este tipo de búsqueda es bastante bueno porque tiende a reducir el número de nodos a visitar antes de encontrar una solución. No obstante, tiene tres posibles inconvenientes. Primero, está el problema de la “colina falsa”, en este caso la búsqueda se retrosigue de manera extensiva para encontrar la solución. En segundo lugar, está el problema de las “mesetas”, en el que todos los pasos siguientes parecen iguales de buenos o malos. En este caso la escalada de la colina no es mejor que una búsqueda primero en profundidad. En tercer lugar, está el problema de las “crestas”. En este caso, la escalada de la colina no es efectiva porque el algoritmo provocará que crucemos la cresta varias veces durante el retroseguimiento.

A pesar de estos problemas potenciales, la escalada de la colina generalmente le lleva a una solución óptima, más rápidamente que cualquier otro método no heurístico.

#### 3.4.3.2 La técnica de búsqueda del menor coste.

La opuesta a la técnica de la escalada de la colina es la del *menor coste*. El menor coste toma el camino del menor esfuerzo.

##### 3.4.3.2.1 Evaluación de la búsqueda del menor coste.

La búsqueda del menor coste tiene las mismas ventajas y desventajas que la búsqueda de escalada de la colina, excepto en el orden inverso. Puede haber falsos valles, tierras bajas y barrancos; pero, sobre todo, la búsqueda de menor coste tiende a funcionar bastante bien. Solamente se puede decir que, por término medio, sobrepasa a una búsqueda a ciegas.

### 3.5 Lógica e incertidumbre.

Muchos conocimientos, especialmente si se adquieren por la observación, son probabilísticos en cierta medida: es decir, uno puede creer que algo es verdad aún admitiendo la posibilidad de un error. Para que una computadora piense y razone como un humano, se exige y se requiere la integración de esta incertidumbre con la lógica.

### 3.5.1 Lógica proposicional.

La lógica proposicional trata de la determinación de la verdad o falsedad de distintas proposiciones. Una proposición es una afirmación formada correctamente que es, o verdadera o falsa. La lógica proposicional utiliza operadores para conectar las proposiciones. Aquí tiene los operadores más comunes:

Y  
O  
NO  
IMPLICA  
EQUIVALENTE

Aunque la lógica proposicional constituye la base para los lenguajes, tanto inteligentes como informáticos, usted no puede usarla por sí misma para representar el conocimiento humano del mundo. Debido a que la lógica proposicional carece de la habilidad de representar relaciones entre los objetos, está limitada solamente a determinar la verdad o falsedad de una oración dada y, por tanto, no puede usarse en clasificaciones. Esta carencia nos lleva al cálculo de predicados.

### 3.5.2 Cálculo de predicados.

El cálculo de predicados, llamado frecuentemente lógica de predicados, es simplemente una extensión de la lógica proposicional que permite el examen de los valores de verdad en afirmaciones que se hagan sobre objetos determinados.

La base del cálculo de predicados es el predicado que esencialmente es una función que devuelve un valor verdadero o falso dependiendo de su argumento.

La diferencia fundamental entre la lógica de predicados y la lógica proposicional es la separación del atributo del objeto que posiblemente lo posee; es decir, en el cálculo de predicados, usted puede crear una función para determinar la dureza de un objeto. Sin embargo, en la lógica proposicional, usted debe generar un nuevo enunciado para cada caso.

### 3.5.3 Incertidumbre.

Todas las cosas representadas por la lógica son solamente verdadero o falso. Sin embargo en el mundo real no es así, la validez o verdad de la mayoría de las observaciones está gobernada por un factor de probabilidad asociado a ella.

Al intentar resolver situaciones donde alguna información es desconocida, llegamos a dos importantes áreas de investigación en inteligencia artificial. La primera área es la de la *lógica difusa*, que trata de la evaluación de expresiones lógicas que contiene valores desconocidos. La segunda área cubre los *sistemas probabilísticos*, que utilizan la probabilidad de aparición de distintos hechos para llegar a una respuesta.

### 3.5.4 Lógica incierta o difusa.

La *lógica difusa* se refiere a la evaluación de expresiones lógicas que pueden contener valores desconocidos. Debido a que la lógica es determinista, puede pensar que la presencia de algo desconocido invalidaría una expresión lógica: pero, en algunos casos, algo desconocido no implica que la expresión no pueda ser evaluada.

En la lógica difusa, un valor lógico puede ser uno de los siguientes:

- Verdad.
- Falso.
- Desconocido (llamado algunas veces “puede ser”).

Un valor desconocido significa que, mientras puede ser verdad o mentira su valor se desconoce.

### 3.5.5 Sistemas probabilísticos.

Una de las áreas más importantes que requiere el procesamiento de la incertidumbre es la del sistema experto. Si el sistema experto va a dar un buen consejo, debe también informar de la probabilidad de que dicho consejo sea correcto. A continuación tiene una revisión de los

fundamentos de las teorías de probabilidad clásicas. Estas reglas se deben modificar para aplicarlas a la toma de decisiones en inteligencia artificial.

### 3.5.6 Teoría probabilística clásica.

El número de posibles acontecimientos que pueden ocurrir determina la probabilidad de que dicho acontecimiento ocurra. Al conjunto de todos los sucesos posibles relacionados con una cierta situación se le denomina espacio de sucesos. Si la posibilidad de cada acontecimiento en el espacio de los sucesos ocurridos es igual, y si no se duplican los acontecimientos, entonces la probabilidad ( $P$ ) de un acontecimiento específico  $x$  es igual a:

$$P_x = \frac{1}{N}$$

Donde  $N$  es el número de posibles eventos. Por ejemplo, la posibilidad de que una moneda lanzada al aire salga cara es de  $\frac{1}{2}$ , ya que hay 2 posibles eventos – caras y cruces – y ambos tienen la misma posibilidad de ocurrir. La posibilidad de sacar un 3 en una sola tirada de dado es de  $\frac{1}{6}$  porque hay 6 posibles resultados y todos los resultados puede ocurrir igualmente.

La parte de la probabilidad en la que más se interesan los programadores de inteligencia artificial es la de la computarización de la posibilidad de que una secuencia o combinación de eventos ocurra. Dentro de la teoría probabilística formal, la probabilidad de una combinación de eventos es igual al producto de las probabilidades de cada uno de dichos eventos. Por

ejemplo, la posibilidad de que una moneda salga cara tres veces seguidas cuando se la lanza es:

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

Los eventos que forman una secuencia o combinación no necesitan estar relacionados. Por ejemplo, la posibilidad de que salga cara en una moneda y un 4 en una tirada de dado son:

$$\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$$

Todas las posibilidades están entre 0 y 1, ambos inclusive. Un evento con 1 de probabilidad es una certeza; por ejemplo, la muerte es una certeza. Un evento con 0 de probabilidad es una imposibilidad: por ejemplo, vivir para siempre sería considerado como una imposibilidad. No es posible tener una probabilidad menor de 0 o mayor de 1.

El método de probabilidad clásico no funciona bien cuando lo aplica a la resolución de la incertidumbre de la información. Veremos dos métodos más satisfactorios para resolver este problema.

### 3.5.7 El método del eslabón más débil.

Cuando un factor está asociado con un atributo específico, significa que hay una pequeña posibilidad de que el atributo pertenezca a un ejemplar del objeto en cuestión. Por tanto, un método sugiere que la certeza de una solución debería ser el coeficiente de certeza más bajo

que esté asociado con el objeto. Este concepto es análogo al de una cadena que fuera sólo tan fuerte como su eslabón más débil.

La mayor ventaja del método del eslabón más débil es que da la oportunidad de éxito como si todos los atributos estuvieran asociados con el objeto independientemente: es decir, como si la probabilidad de una solución estuviera basada en una combinación de factores. Por otra parte, hay situaciones en las que el argumento más convincente es más importante que el eslabón más débil. Estas situaciones necesitan el método del eslabón más fuerte.

### 3.5.8 El método del eslabón más fuerte.

En ciertas aplicaciones, la probabilidad de que una solución sea la correcta no se basa en la evidencia más débil, sino más bien, en la más fuerte. Esto es similar a un debate en el que sólo un argumento fuera suficiente para decidir el resultado. Por ejemplo ¿por qué no es bueno dejar armas cargadas donde los niños puedan alcanzarlas? Hay varias razones, tales como que el niño podría disparar el arma y hacer un ruido ensordecedor, o que el niño podría disparar y romper algo. Pero el argumento más convincente es que el niño podría disparar el arma y herir o matar a alguien. Claramente, es el último argumento el más importante.

Cuando se aplica a un sistema experto, el método del eslabón más fuerte simplemente implica que la probabilidad de que una solución sea correcta es la misma que la probabilidad de que un atributo sea el que más probabilidades tiene de aparecer.



# Capítulo II

## *Diseño del Sistema Experto*

### 3.1 Diseño de la solución

Para poder diseñar un Sistema Experto como propuesta de solución de la problemática del diagnóstico temprano de cáncer, se debe conocer cuál es el proceso de desarrollo de un sistema experto.

#### 3.1.1 Desarrollo de los sistemas expertos

Para desarrollar el software de un sistema experto, primero debemos conocer el equipo de gente necesario para la elaboración del mismo, después los métodos que utiliza ese equipo de gente y por último cómo prueban y construyen prototipos de software para concluir en el Sistema Experto final.

- Equipo de desarrollo.
- Métodos auxiliares en el desarrollo.
- Construcción de prototipos.

##### 3.1.1.1 El equipo de desarrollo

Las personas que componen un grupo o un equipo, como en todos los ámbitos deben cumplir ciertas características y cada uno de ellos dentro del equipo desarrolla un papel distinto. A

continuación detallaremos cada componente del equipo dentro del desarrollo y cuál es la función de cada uno:

- El experto. La función del experto es la de poner sus conocimientos especializados a disposición del sistema experto.
- El ingeniero del conocimiento. El ingeniero que plantea las preguntas al experto, estructura sus conocimientos y los implementa en la base de conocimientos.
- El usuario. El usuario aporta sus deseos y sus ideas, determinando especialmente el escenario en el que debe aplicarse el Sistema Experto.

En el desarrollo del sistema experto, el ingeniero del conocimiento y el experto trabajan muy unidos. El primer paso consiste en elaborar los problemas que deben ser resueltos por el sistema. Precisamente en la primera fase de un proyecto es de vital importancia determinar correctamente el ámbito estrechamente delimitado de trabajo. Aquí se incluye ya el usuario posterior, o un representante del grupo de usuarios. Para la aceptación, y en consecuencia el éxito, es de vital y suma importancia tener en cuenta los deseos y las ideas del usuario.

En este sentido tanto el grupo de expertos mismo como el grupo de colaboradores son usuarios potenciales del Sistema. Así que en este rubro de usuario los participantes emiten sus opiniones sobre la manera en que deberá estar estructurado el sistema y los resultados a emitir.

Una vez delimitado el dominio, nos disponemos a "engrosar" nuestro sistema con los conocimientos del experto. El experto debe comprobar constantemente si su conocimiento ha sido transmitido de la forma más conveniente. El ingeniero del conocimiento es responsable de una implementación correcta, pero no de la exactitud del conocimiento. La responsabilidad de esta exactitud recae en el experto.

La *figura 5* muestra el proceso de desarrollo de un sistema experto tomando en cuenta que forma parte del proceso de representación del conocimiento.



Figura 5. Proceso de la representación de conocimientos.

A ser posible, el experto deberá tener comprensión para los problemas que depara el procesamiento de datos. Ello facilitará mucho el trabajo. Además, no debe ignorarse nunca al

usuario durante el desarrollo, para que al final se disponga de un sistema que le sea de máxima utilidad.

La estricta separación entre usuario, experto e ingeniero del conocimiento no deberá estar siempre presente. Pueden surgir situaciones en las que el experto puede ser también el usuario. Este es el caso, cuando exista un tema muy complejo cuyas relaciones e interacciones deben ser determinadas una y otra vez con un gran consumo de tiempo. De esta forma el experto puede ahorrarse trabajos repetitivos. Esta situación se presenta en nuestro caso en particular.

La separación entre experto e ingeniero del conocimiento permanece, por regla general inalterada.

### 3.1.1.2 Construcción de prototipos (Rapid Prototyping)

En el desarrollo de sistemas expertos se nos plantean dos importantes riesgos:

- No existen implementaciones similares que puedan servir de orientación al encargado del desarrollo en la casi totalidad de los casos.
- En muchos puntos, los requisitos necesarios están esbozados con muy poca precisión.

El diseño y la especificación requieren una temprana determinación de la interface del software y de la funcionalidad de los componentes. En el desarrollo de sistemas expertos deben alterarse a menudo durante y también después de su implementación, ya que los

requisitos se han ido configurando y han obtenido mayor precisión, o porque se ha descubierto que deben iniciarse otras vías de solución. Durante el desarrollo, resulta más apropiado empezar con implementaciones tipo test para encontrar el camino hacia una solución definitiva y para hacerlas coincidir con las necesidades del usuario.

Un método efectivo es la implementación de un prototipo de sistema experto que permita llevar a cabo las funciones más importantes de éste, aunque con un esfuerzo de desarrollo considerablemente inferior al de una implementación convencional. Este proceder se define bajo el nombre de "Rapid Prototyping". Las máquinas de inteligencia artificial especialmente desarrolladas, los lenguajes de programación de inteligencia artificial y en determinados casos los Shells, ofrecen una considerable ayuda para el "Rapid Prototyping".

Para sistemas expertos, el "Rapid Prototyping" es el procedimiento más adecuado, pues posibilita una rápida reacción a los deseos en constante cambio tanto por parte de los expertos como parte del usuario.

### 3.1.2 Diseño del prototipo (Rapid Prototyping)

Ya que nuestro prototipo está orientado a un diagnóstico médico, nuestro "experto" está conformado por un equipo de médicos pediatras y epidemiólogos, investigadores del Centro Médico Siglo XXI del Instituto Mexicano del Seguro Social. Los usuarios potenciales de este prototipo son los médicos de primer nivel de atención representados por los médicos generales. Finalmente, nuestro trabajo fue fungir como ingenieros del conocimiento.

Delimitado el campo de acción de cada participante y la forma en que se dará el desarrollo del sistema el siguiente paso es diseñar el prototipo. En primer lugar ¿Qué tipos de cáncer detectará el sistema experto? Se adoptaron los diez primeros grupos de neoplasias de la clasificación utilizada internacionalmente para denotar el cáncer en niños, ya que son los de mayor incidencia en la población infantil de México.<sup>13</sup> Por otro lado fundada en esta clasificación se realizan de igual forma la clasificación de los casos de cáncer objeto de estudio de las investigaciones efectuadas por nuestro equipo de expertos humanos que es la base en la cual se sustenta el diseño del sistema experto. En la *figura 6* se observa la gráfica de los índices de ocurrencia de cáncer en la población infantil de México denotados de estas investigaciones.

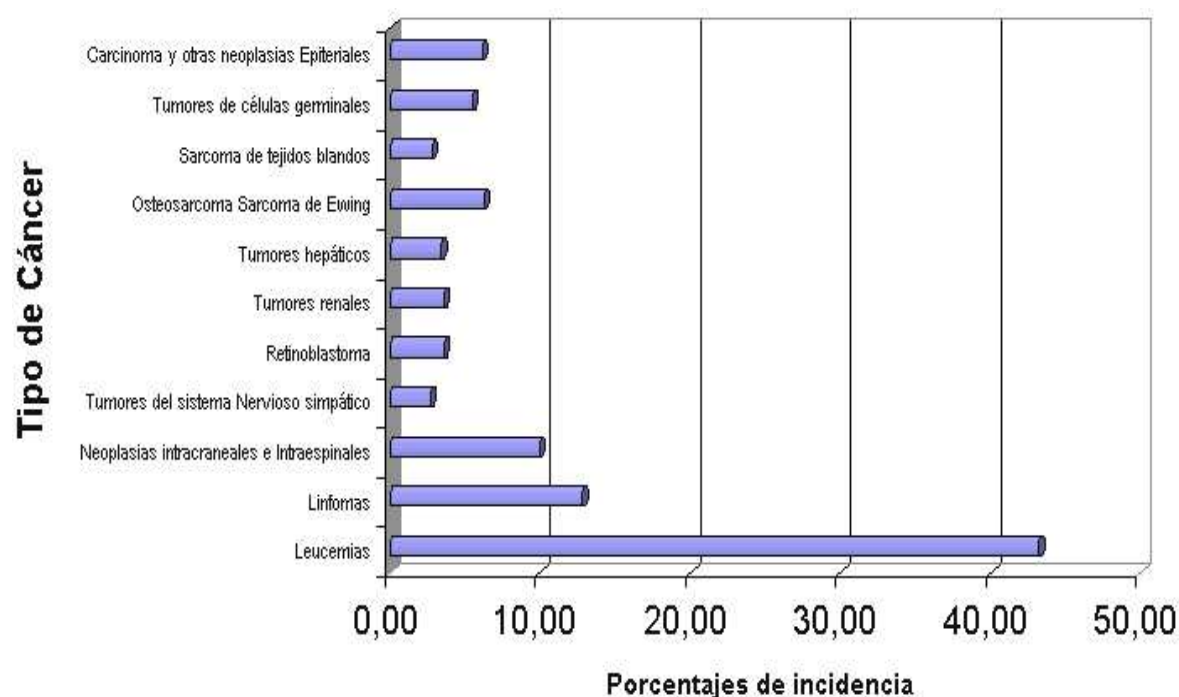


Figura 6. Análisis univariado de los porcentajes de incidencia de cáncer en niños.

<sup>13</sup> Fajardo Arturo. Epidemiología descriptiva del cáncer en el niño.

La información obtenida del grupo de expertos médicos fue proporcionada en una base de datos cuyo parte del esquema se muestra en *figura 7*. Aquí se muestra el segmento de la información utilizada por el sistema.

CLASIBM	Clasificación de Birch y Marsden
CLASIMORFO	Clasificación morfológica
CLASITPOPO	Clasificación topográfica
SEXO	Sexo del paciente
LUGARDERE	Estado donde reside el paciente
SIGNO1	Cual fue el primer síntoma presentado
SIGNO2	Cual fue el segundo síntoma presentado
SIGNO3	Cual fue el tercer síntoma presentado
SIGNO4	Cual fue el cuarto síntoma presentado
SIGNO5	Cual fue el quinto síntoma presentado
SIGNO6	Cual fue el sexto síntoma presentado
SIGNO7	Cual fue el séptimo síntoma presentado

Figura 7. Representación del esquema de la Base de Datos proporcionada por el IMSS.

En la *figura 8* se muestra la información que aparece en la Base de datos donde se distingue en la clasificación de Birch y Marsden que los dos primeros dígitos determinan el tipo de cáncer que presenta el paciente.

Contents of SQLBASE6.DBF							
Definition		Data		Enter SQL			
	CLASIMORFO	CLASITPOPO	CLASIBM	ESTUDIOH	REGISTRO	ESTADIO	LUGARDERE
►	9821/3	C91.0	01A9821	1		07	15
	9821/3	C91.0	01A9821	1		06	15
	9821/3	C91.0	01A9821	1		06	15
	9180/3	C40.2	08A9180	1		04	15
	9473/3	C72.0	03C9473	1		03	15
	9821/3	C91.0	01A9821	1		06	15

Figura 8. Información de la Base de Datos proporcionada por el IMSS.

Debido a la especificidad de la información, en particular de la clasificación de Birch y Marsden en donde es claro observar que parte del término para clasificar el tipo de cáncer es



tomado de su clasificación morfológica y ya que el objetivo del sistema es el diagnóstico del cáncer de manera general es necesario eliminar del término la clasificación morfológica. Para llevar a cabo esto se determinó utilizar solo los dos primeros dígitos de la clasificación de Birch y Marsden. En la *figura 9* se observa la nomenclatura correspondiente a cada tipo de cáncer.

01A9821	Leucemia
02B9595	Linfoma
03C9473	Tumor de Sistema Nervioso Central
04A9500	Tumor de Sistema Nervioso Simpático
059510	Retinoblastoma
06A8960	Tumores Renales
07A8970	Tumores Hepáticos
08A9180	Tumores Óseos
09D9040	Tumores Tejidos Blandos
10C9060	Tumores Células Gerninales

Figura 9. Nomenclatura entre el termino de la Clasificación de Birch y Marsden y el cáncer especificado.

Ahora bien ya aclaramos la disyuntiva de los tipos de cáncer que el sistema detectará y como obtenerlos, pero ¿Cómo lograr que el sistema detecte el tipo de cáncer? En primer lugar hay que determinar la manera en que se realiza un diagnóstico por parte de un experto humano. En términos generales un médico realiza el diagnóstico de una enfermedad en base a los síntomas que el paciente presenta que es el fenómeno que aparece como consecuencia de una alteración funcional u orgánica en cualquier parte del organismo. En este caso en particular este tipo de “mecánica” del diagnóstico no es posible llevarla a cabo debido a que en su mayoría la sintomatología conocida de esta enfermedad, conocido en la jerga médica como “sintomatología florida”, corresponde a etapas finales de la enfermedad en donde en su mayoría el cáncer se ha desarrollado casi por completo.

El conocimiento de la sintomatología del cáncer en etapas tan avanzadas de la enfermedad es de poca ayuda ya que la finalidad del sistema es la detección temprana.

Nuestro grupo de expertos cuenta, como resultado de varios trabajos de investigación, con una base de datos de síntomas y signos presentados por el niño con diagnóstico positivo de cáncer a lo largo del posible desarrollo de la enfermedad, desde el primer signo que apareció, hasta el padecimiento que presentó en el momento del diagnóstico. Es por esto que el grupo de expertos médicos esta tratando de validar el esquema mostrado en la figura 10 en donde se observan los primeros diez grupos de neoplasias empleadas internacionalmente para clasificar la sintomatología característica de cada uno de los tipos de cáncer.

<u>Grupo de Neoplasias</u>	<u>Signo o síntoma</u>
<b>I. Leucemias</b>	Petequias
<b>II. Linfomas</b>	Anemia que no se corrige
<b>III. Tumores del Sistema Nervioso Central</b>	Tumoración en cuello
<b>IV. Tumores del Sistema Nervioso simpático</b>	Distensión abdominal
<b>V. Retinoblastoma</b>	Ataxia, vómito y convulsiones que no se controlan
<b>VI. Tumores renales</b>	Tumoración?
<b>VII. Tumores hepáticos</b>	Estrabismo después de 6 meses
<b>VIII. Tumores óseos</b>	Leucocoria
<b>IX. Tumores de los tejidos blandos</b>	Distensión abdominal
<b>X. Tumores de células Germinales</b>	Distensión abdominal
	Dolor por traumatismo con duración de mas de dos semanas
	Tumoración
	Hidrocele después de seis meses
Problema: Sintomatología poco específica	

Figura 10. Algunos signos y síntomas tempranos en el niño con cáncer (Hospital de Pediatría CMN Siglo XXI)

En consecuencia, se tomó como alternativa utilizar estos síntomas para crear la base de conocimientos del sistema experto. En este sentido los signos se encontraron divididos en siete campos (signo1, ... , signo7), primero fue necesario reunir el contenido de estos campos para

poder ingresarlo como primitivas a la base de conocimientos. Pero surge otra dificultad, la terminología utilizada resulta en ocasiones ambigua además de hacer uso de expresiones similares para enunciar síntomas y la dificultad atribuible a la gramática del lenguaje empleada en donde en ocasiones se emplean abreviaturas incorrectamente. Es necesario automatizar hasta donde sea posible la creación de la base de conocimientos para de esta manera poder actualizarla cuando surjan nuevos estudios. Los síntomas que se determinen pertenecerán a la base de conocimientos, serán organizados en base a la frecuencia y porcentaje de aparición orientando el diseño del sistema experto a un enfoque probabilístico basándonos en el tratamiento de incertidumbre a través de un **sistema probabilístico**, lo cual sustenta la teoría de automatizar la creación de la base de conocimientos.

Para esto es necesario la creación de una aplicación de apoyo en la creación de la base de conocimientos. Un tipo de aplicación idónea para esta labor resulta ser un programa de **procesamiento de lenguaje natural** (PLN) utilizando la **técnica de la eliminación de ruidos**.<sup>14</sup> En este sentido es necesario indicarle al PLN que términos debe identificar como validos.

Hasta este momento se ha determinado la estructura necesaria para poder formar la base de conocimientos, es decir, los conocimientos necesarios para saber “los síntomas que indican la presencia del cáncer”.

---

<sup>14</sup> Ignacio Pérez Aguirre. Utilización de procesamiento de lenguaje natural en el diseño de la interfaz de usuario en un sistema experto auxiliar en diagnóstico temprano de cáncer en niños.

Necesitamos un mecanismo que de el seguimiento de los síntomas que el paciente presenta y al igual que un médico genere una hipótesis del tipo de cáncer que este investigando, y en base a esto, determine un posible camino solución que elimine la incertidumbre de asegurar el tipo de cáncer propuesto, además deberá identificar otros factores que incidan en la comprobación de la hipótesis, así como también otros síntomas característicos del cáncer propuesto. Para esto necesitamos del mecanismo de inferencia que apoyado en la base de conocimientos y el PLN desempeñará estas funciones. En términos generales estos son las partes medulares que componen al sistema experto necesarias para desempeñar eficientemente las funciones del experto médico.

El mecanismo de inferencia representa la forma en que razona el experto médico, la base de conocimientos, la teoría del experto adquirida a lo largo de años de estudio y el PLN representa la interacción sistema – usuario , sistema – experto humano.

#### 3.1.2.1 Diseño del procesamiento del lenguaje natural y la base de conocimientos.

El subsistema de PLN interactúa entre el sistema representado por el mecanismo de inferencia y el usuario ya sea bien para expresarle el resultado del diagnóstico del mecanismo de inferencia o para interpretar las preguntas del mecanismo de inferencia e investigar acerca de otros síntomas que podría presentar el paciente. De aquí se deriva la importancia del PLN para el mecanismo de inferencia.

Se comentó en la sección anterior que era necesario para la implementación del PLN indicar que términos debería reconocer como válidos. Esto lo hacemos a través de una relación dentro de la base de conocimientos.

El diseño de la estructura de almacenamiento de la base de conocimientos se fundamentó, en una de las áreas de investigación en inteligencia artificial en el campo de la lógica, los sistemas probabilísticos, en específico apoyados en conceptos básicos de la teoría probabilística clásica para la realización del método del eslabón más fuerte.

En la *figura 11* se muestra el diseño conceptual de la base de conocimientos. En el diseño conceptual se muestra la relación TOKENS y COMPOUND, que muestran los términos que el PLN es capaz de interpretar. Para aquellos síntomas definidos en un solo término (tal es el caso de vómito, diarrea, etc.) se utiliza la relación TOKENS. La relación COMPOUND permite interpretar aquellos síntomas definidos en dos o más términos (como podría ser otorrea crónica, dolor miembros inferiores, etc.).

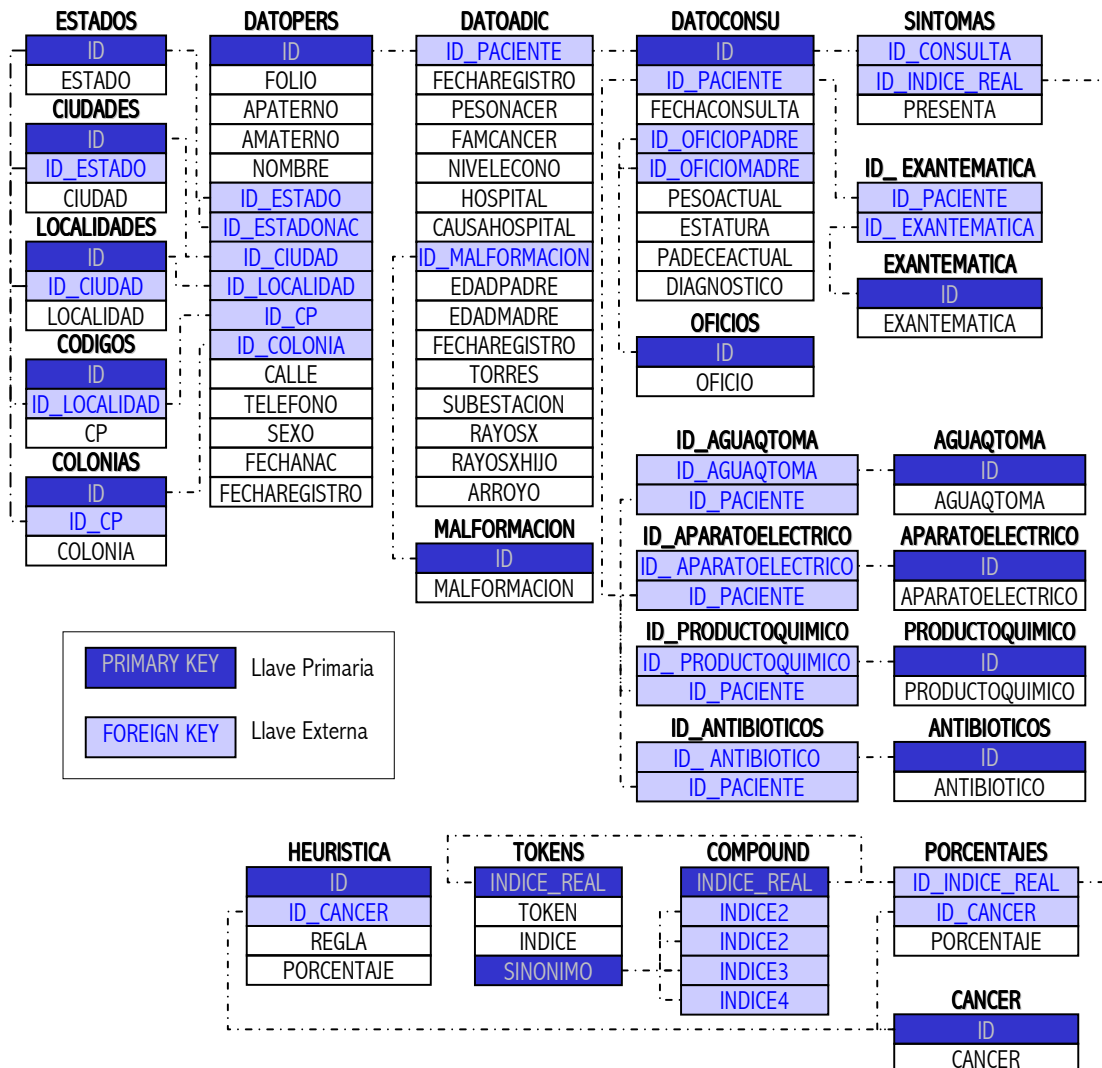


Figura 11. Representación del esquema de la Base de Conocimientos

### 3.1.2.1.1 Elección del método para el sistema probabilístico.

La elección del método del eslabón más fuerte para la construcción del sistema probabilístico se fundamentó en que se debe usar este método cuando se pueda usar cualquier atributo para identificar un objeto. En este caso, el factor de certeza refleja la probabilidad del argumento más convincente. En el sistema experto para la detección temprana del cáncer en niños el objeto a identificar es el tipo de cáncer y los atributos que lo caracterizan son los síntomas.

En nuestro problema en particular, no es posible implementar la versión del eslabón más débil (donde se considera una combinación de todos los atributos), ya que una combinación de síntomas para definir el tipo de cáncer es prácticamente imposible. Esto está respaldado por una situación expresada por los expertos médicos donde argumentan que se han presentado casos de Leucemia en los cuales el único síntoma que se ha registrado es anemia que no cede al tratamiento en un periodo de tiempo prolongado.

### 3.1.2.2 Diseño del mecanismo de búsqueda

Una vez diseñado y estructurado el método de almacenamiento de la información en la base de conocimientos, estamos preparados para abordar el tema del motor de inferencia. Este motor de inferencia es el corazón de un sistema experto. Aunque pueda parecer simple en sentido abstracto, la implementación de un motor de inferencia puede ser bastante difícil. Primero que nada se realizó la elección del mecanismo de búsqueda de solución. En primera instancia la mayoría de las técnicas de búsqueda (nos referimos a las técnicas de búsqueda a ciegas) realizan recorridos por espacios de búsqueda determinados es decir, se basan en recorridos de árboles; característica que en definitiva es imposible aplicar al problema actual debido a que no se tiene definido ningún árbol de búsqueda, esto porque no se ha definido con certeza la sintomatología característica de cada una de las neoplasias malignas.

#### **3.1.2.2.1 Elección del método para el mecanismo de búsqueda.**

La decisión del método de búsqueda se inclinó a favor de las técnicas de búsqueda heurística, ya que la información heurística con que se cuenta está relacionada con un valor cuantificador, obtenido a través de un cálculo estadístico de la probabilidad de que el paciente presente determinada sintomatología dependiendo del tipo de cáncer que se analice. En específico la técnica de búsqueda heurística elegida fue la de escalada de la colina, ya que se busca reducir el número de nodos visitados basándose en escoger el nodo que presente la mayor probabilidad de aparición.

#### **3.1.2.2.2 Elección del método para el mecanismo de inferencia.**

El motor de inferencia fue construido utilizando el método de las reglas de producción ya que se requiere que vaya solicitando la información que tenga una mayor importancia de acuerdo con el estado actual del sistema, además de que elimine la mayor incertidumbre posible del sistema; características propias de este método. Como se cito en el capítulo anteriores, los sistemas expertos de reglas de producción empezaron siendo, bien sistemas de encaminamiento hacia adelante o hacia atrás. Se determinó utilizar el método de reglas de producción basado en el sistema de encaminamiento hacia atrás ya que se inicia con una hipótesis del tipo de cáncer que tiene el paciente, y pide información para confirmarlo o negarlo. En la *figura 12* se muestra el algoritmo del proceso de seguimiento de cada uno de los tipos de cáncer.



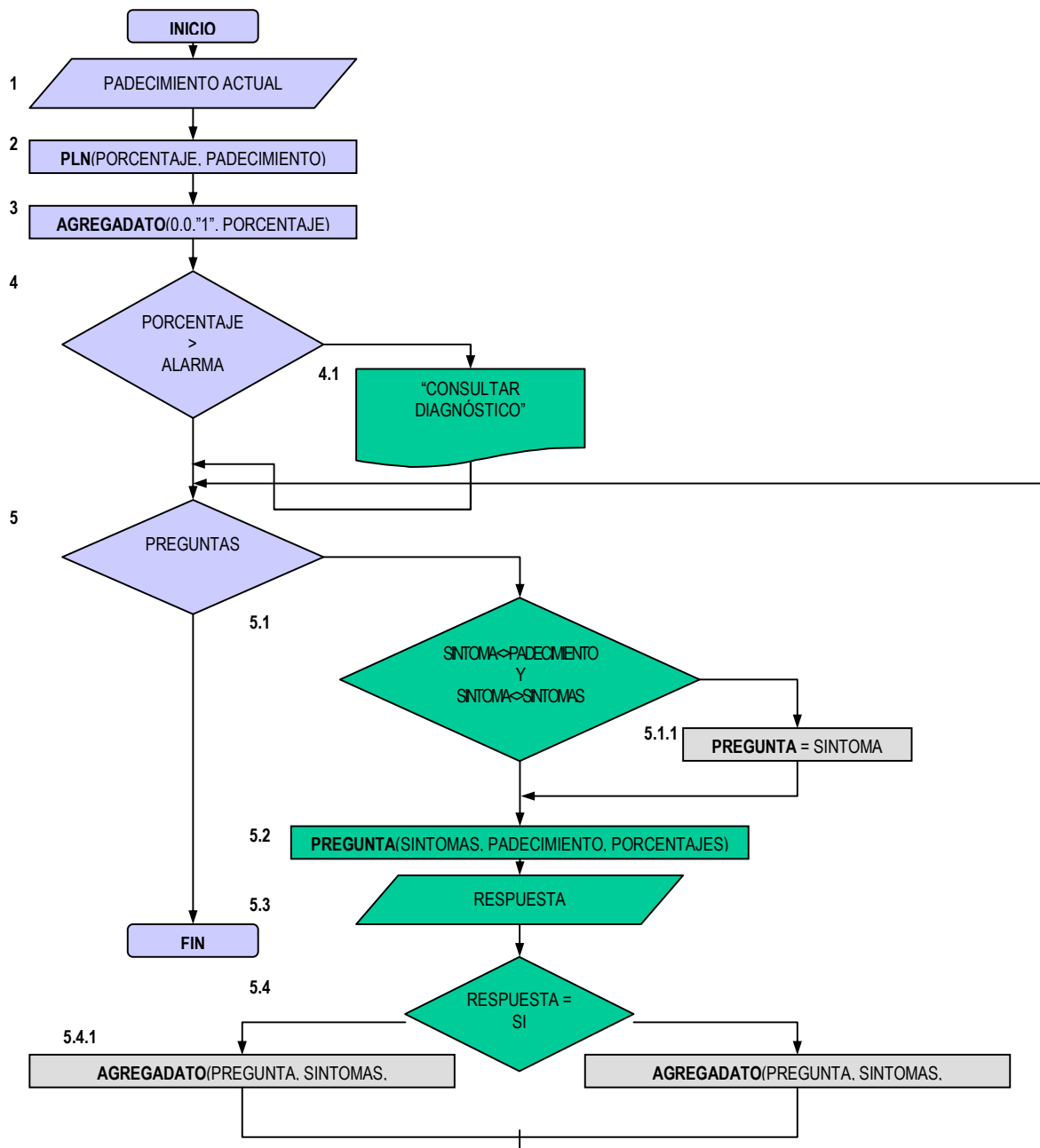


Figura 12. Diagrama de flujo de la sesión de preguntas realizada por el mecanismo de inferencia.

A continuación se describe el algoritmo de la sesión de preguntas realizado por el mecanismo de inferencia:

1. La sesión de preguntas inicia con la inserción del motivo de asistencia a consulta médica por parte del paciente, es decir el padecimiento actual.
2. El padecimiento actual es analizado por el PLN para obtener la sintomatología del paciente, sintomatología que es almacenada en una estructura dinámica conocida como lista enlazada empleada para el seguimiento de la consulta.
3. Debido a que la mayor incidencia de cáncer en niños la tiene los casos de leucemia, la primera línea de búsqueda que intenta generar está orientada a tratar de determinar si el paciente puede presentar leucemia. Esto se logra agregando el dato de hipótesis de leucemia a la lista enlazada de porcentajes con un índice 0%. En la lista de porcentajes se da seguimiento a los índices de probabilidad del tipo de cáncer que se puede presentar en el paciente.
4. Si el índice porcentual de la mayor probabilidad del tipo de neoplasia maligna es superior al máximo establecido como alarma, se realiza lo siguiente:
  - 4.1. Se debe indicar al usuario que consulte el diagnóstico del mecanismo de inferencia para revisar los índices porcentuales.
5. Si existen preguntas que aún no se han realizado, basado en esta teoría de los índices porcentuales, el sistema comienza la sesión de preguntas con el signo que tenga la mayor incidencia y elimine la mayor incertidumbre hacia qué tipo de cáncer debe seguir orientándose el diagnóstico.
  - 5.1. Se realiza una consulta a la base de conocimientos para obtener un síntoma que no se encuentre registrado dentro del padecimiento actual o síntomas ya cuestionados y además tenga el mayor índice de probabilidad de aparecer.
    - 5.1.1. Se obtiene el síntoma como la siguiente pregunta a realizar.
  - 5.2. Se lanza la pregunta al usuario.

5.3. Se obtiene la respuesta del usuario.

5.4. Se almacena la respuesta a la pregunta en la lista de síntomas.

5.4.1. Se agrega el nuevo síntoma a la lista de síntomas dependiendo de la respuesta, si la respuesta es afirmativa, se ingresa con valor verdadero, en caso contrario, se ingresa con un valor falso.

Si en algún momento de la consulta el índice porcentual de algún tipo de cáncer supera a los demás, a partir de ese momento la sesión de preguntas se orienta a realizar cuestionamientos acerca de los signos característicos de los pacientes que desarrollaron ese cáncer.

En la *figura 13* se presenta un esquema en donde se muestra la forma en que interactúa el sistema para identificar un posible caso de cáncer.

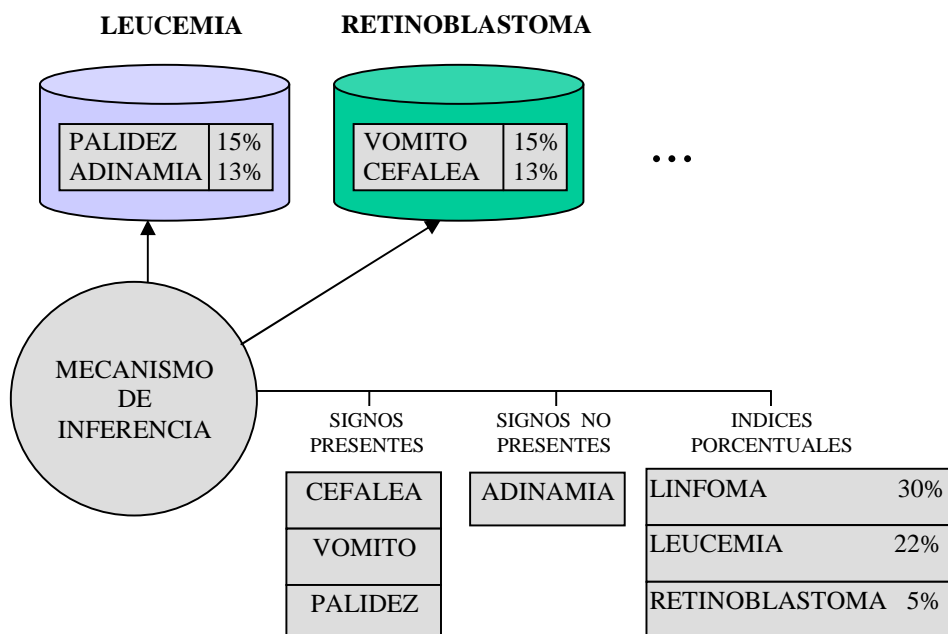


Figura 13. Diseño del mecanismo de inferencia.

# Capítulo III

## *Desarrollo del Sistema Experto*

### 3.1 Propuesta de solución

Para alcanzar los objetivos de este trabajo, se propone el desarrollo de un sistema experto que funcione como un medio de apoyo para el estudio de los signos y síntomas que ocasionan el cáncer en los niños menores de 15 años, registrarlos y estadificarlos, esto para tener una idea más clara de la epidemiología básica que da origen a esta enfermedad y otros aspectos, los cuales podrían ser tomados en cuenta para un diagnóstico oportuno. Además el sistema podrá ser consultado por el médico familiar o pediatra para determinar si, basado en la sintomatología presentada por el paciente, existe riesgo de que pueda ser algún tipo de las neoplasias malignas que el sistema es capaz de detectar.

Para el desarrollo del sistema experto se utilizó el lenguaje de programación de alto nivel C/C++ y en particular el entorno de programación visual de Borland, C++ Builder versión 5.0, una herramienta RAD (*Rapid Application Developed*) para el desarrollo de software. En el siguiente apartado se justifica la utilización de un lenguaje de propósito general como lo es C/C++ en lugar de algún lenguaje de programación de inteligencia artificial como lo son Prolog o LISP.

#### 3.1.1 ¿Por qué usar C/C++ para la programación en inteligencia artificial?

C/C++ es uno de los lenguajes de programación más populares en uso. Proporciona un esqueleto estructurado sin límite a la creatividad del programador, además los compiladores C

producen programas muy rápidos y eficientes al ejecutarlos. Por estas y otras razones, muchos paquetes de software están escritos en C/C++.

Ya que hay más programadores que conocen C/C++ que los que conocen algunos otros lenguajes normalmente usados en investigación en inteligencia artificial, parece claro que si las técnicas de inteligencia artificial van a usarse en el mundo real necesitan ser codificadas en C/C++.

Mientras que los lenguajes usados para la investigación en inteligencia artificial son excelentes para la exploración, son difíciles de adaptar a una programación general. Por ejemplo, Prolog carece de la capacidad de una programación general. Deficiencia ésta que hace que ciertas tareas, como es la suma de números de una lista, sea más difícil de lo que debería ser en realidad. Además, si la aplicación a desarrollar no requiere el retroseguimiento del Prolog ni de los recursos de una base de datos, éstos se convierte en un peso extra que debe soportar la aplicación.

La mayoría de los lenguajes de inteligencia artificial (incluyendo LISP y Prolog) son lenguajes *declarativos*, mientras que los lenguajes comunes son *de procedimiento*. En un lenguaje declarativo, el programador le dice a la computadora *lo que* debe hacer; en un lenguaje de procedimiento, el programador le dice a la computadora *cómo* ha de hacerlo. Por tanto, los lenguajes de inteligencia artificial y los restantes tienden, a convertirse en polos contrapuestos dentro del ámbito del lenguaje de programación.

No hay ni siquiera una sola técnica de IA que no pueda ser desarrollada usando un lenguaje de procedimiento como lo es el C/C++. De hecho, el desarrollo de ciertas rutinas es, en verdad, más claro en C/C++ que en un lenguaje de inteligencia artificial. Hay tan sólo vagas razones históricas para explicar por qué algunos lenguajes especiales de inteligencia artificial fueron inventados en primer lugar.<sup>15</sup>

Aprender a desarrollar varios elementos de inteligencia artificial a través del C/C++ es importante porque estos elementos pueden adaptarse a un gran número de aplicaciones existentes. Muchas aplicaciones han sido, y continuarán siéndolo, escritas en C/C++. Muchas de estas aplicaciones pueden beneficiarse de la suma de los distintos avances basados en la inteligencia artificial. Por ejemplo, un gestor de base de datos o un sistema operativo podría usar un procesador de lenguaje natural “front-end” para crear un mejor procesador de órdenes, mientras que un paquete de contabilidad podría incluir un módulo de sistema experto que podría ayudar al usuario en la planificación de impuestos o en el inventario de la empresa.

Aunque puede en ciertos momentos unir un módulo escrito en C/C++ con otro escrito en lenguaje inteligencia artificial (como el Turbo Prolog), no es ésta la solución más conveniente.

En primer lugar, usar dos lenguajes separados supone unos mayores problemas de dirección y coordinación porque se debe en realidad mantener dos proyectos separados. En segundo lugar, usar un lenguaje separado de inteligencia artificial supone, o bien que los programadores originales del sistema tengan que aprender el lenguaje, o bien que se tenga que contratar a

---

<sup>15</sup> Hebert Schildt. Utilización de C en la inteligencia artificial.

otros programadores. La mejor forma es implementar cualquiera de las técnicas basadas en inteligencia artificial requeridas en el lenguaje original del paquete de aplicaciones.

### 3.2 Desarrollo del mecanismo de inferencia

El mecanismo de inferencia se apoya fuertemente en una base estadística comprendida en la base de conocimientos. El mecanismo de inferencia interactúa de manera directa con la interfaz de usuario, representada por el procesamiento de lenguaje natural, y con la base de conocimientos.

El código de la *figura 14* muestra la definición del objeto lista enlazada en donde se lleva el seguimiento de los signos y síntomas presentes y ausentes en el paciente, así como el porcentaje para cada uno de los diferentes tipos de cáncer determinado por la sintomatología presentada por el paciente.

```
class ClaseLista{
private:
    ClaseLista *Siguiente;
public:
    String Index;
    String Heuristica;
    int Ocurrencia;
    float Porcentaje;
    bool Presenta;
    ClaseLista* ObtenMemoria(String, ClaseLista* *);
    int AgregaDato(String, ClaseLista* *);
    int AgregaDato(String data,ClaseLista* *Ptr,bool );
    int AgregaDato(float, String, ClaseLista* *);
    int AgregaDato(String, ClaseLista* *, int);
    int ObtenDato(ClaseLista* *);
    void Elimina(ClaseLista* *);
    void Ordena(ClaseLista* *,int );
    int ObtenTotales(ClaseLista* *);
    int GenerarGrupos(ClaseLista* *,short int,int);
```



```

int ObtenUltimo(Claselista *);
int EliminaUltimo(Claselista* *);
int BuscarIndex(AnsiString Indice, Claselista *Temp);
void EliminaSimilar(Claselista *Primer, Claselista* *Segundo)
void Next(Claselista* *);
};

```

Figura 14. Definición del objeto que manipulará tanto los signos y síntomas presenta y ausentes en el paciente como la probabilidad de que cada uno de los diferentes tipos de cáncer este presente.

Una *lista enlazada* es una colección lineal de estructuras autorreferenciadas llamadas *nodos* conectado por *enlaces* de apuntador –de ahí el termino lista enlazada<sup>16</sup>. Se tiene acceso a una lista enlazada vía apuntador al primer nodo de la lista (Padecimiento y Síntomas). En una lista enlazada los datos se almacenan dinámicamente –cada nodo se crea conforme es necesario. En la figura 15 se observa la representación gráfica de la lista enlazada.

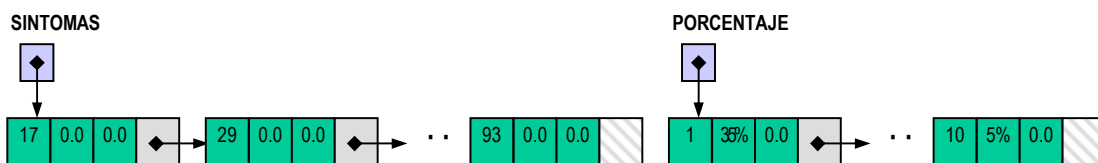


Figura 15. Representación gráfica de la lista enlazada

En el caso de la lista enlazada de síntomas y padecimientos, se tienen rutinas básicas de recorrido de listas enlazadas y rutinas para obtención de memoria para almacenar un nuevo nodo que representa al síntoma presente o no en el paciente. En el caso de la lista enlazada de porcentajes, se guardan los índices porcentuales de cada uno de los tipos de cáncer que el sistema reconoce. Otras rutinas básicas implementadas en esta última lista son para agregar un dato a la lista enlazada, obtener totales para el cálculo de índices porcentuales, ordenar la lista enlazada en base a los índices porcentuales, eliminar los datos almacenados en la lista así como también los espacios de memoria que estos utilizan, eliminar la duplicidad de la

<sup>16</sup> H. M. Deitel, P. J. Deitel. Como programar en C/C++

información contenida en la lista, buscar un dato específico de la lista, eliminar el último nodo insertado y recorrer el apuntador del nodo actual para que apunte al siguiente nodo de la lista.

En la *figura 16* se muestra la rutina *Pregunta* que representa el mecanismo de inferencia. Esta función es crucial para el correcto desempeño del sistema experto ya que se encarga de ejecutar el cálculo de porcentajes en base a las respuestas que vaya proporcionando el usuario. Además esta rutina determina las preguntas que el usuario deberá responder durante la consulta.

```
bool TFormConsulta::Pregunta(AnsiString *Texto, bool recorrer)
{
    TQuery *Consulta = new TQuery(this);
    bool Found;
    iPregunta = 0;

    Consulta->DatabaseName = "SIDICAN";
    *Texto = "";

    SumarPorcentajes(recorrer);

    if(Porcentaje)
    {
        if(Porcentaje->Porciento >= MAX_ALARMA)
        {
            Application->MessageBox("Probabilidad de Cáncer elevada.\nVerifique su Diagnostico",
                "Posible caso de Cáncer", MB_ICONEXCLAMATION);
            Diagnostico = StrToInt(Porcentaje->Index);
        }
        else Diagnostico = 0;
    }else Diagnostico = 0;

    Porcentaje->AgregaDato(0.0,"1",&Porcentaje);

    Found = false;
    for(ClaseLista *Aux = Porcentaje;!Found && Aux;)
    {
        Consulta->Close();
        Consulta->SQL->Clear();
        Consulta->SQL->Add("SELECT * FROM PORCENTAJES WHERE ID_CANCER = :x");

        for( ClaseLista *Sintom = Padecimiento; Sintom; Sintom->Next(&Sintom))
            Consulta->SQL->Add("AND ID_INDICE_REAL <> " + Sintom->Index);

        for( ClaseLista *Sintom = Sintomas; Sintom; Sintom->Next(&Sintom))
```

```

Consulta->SQL->Add("AND ID_INDICE_REAL <> " + Sintom->Index);

Consulta->SQL->Add("ORDER BY PORCENTAJE DESC");
Consulta->ParamByName("x")->AsInteger = StrToInt(Aux->Index);
Consulta->Open();
if(!Consulta->IsEmpty())
{
    Found = true;
    iPregunta = Consulta->FieldByName("ID_INDICE_REAL")->AsInteger;
}
else
    Aux->Next(&Aux);
}
if(Found)
{
    FormExpertSys->Termino(Texto, iPregunta);
    delete Consulta;
    return true;
}
*Texto = "Preguntas agotadas. Consulte su Diagnostico";
delete Consulta;
return false;
}

```

Figura 16. Código necesario para formular la pregunta a realizar y dar seguimiento de las probabilidades de cáncer para cada uno de los tipos objeto de estudio.

A continuación se explica brevemente su funcionamiento:

- 1 La variable *Consulta* de tipo *TQuery* es una variable de tipo *puntero* que se emplea para establecer consultas a la base de conocimientos. La variable *Found* de tipo *bool* es una bandera para conocer cuándo se ha encontrado el siguiente síntoma a cuestionar y la variable *iPregunta* de tipo *int* representa la pregunta a efectuar por el mecanismo de inferencia.
- 2 *SumarPorcentajes(recorrer)*; es una rutina que recalcula y ordena los índices porcentuales para la lista de porcentajes de acuerdo a cada uno de los tipos de cáncer que el sistema detecta.

- 3 `if(Porcentaje->Porciento >= MAX_ALARMA)` si el índice porcentual de la mayor probabilidad de la lista de porcentajes de tipos de cáncer es superior al máximo establecido como alarma, se emite el mensaje de consulta de diagnóstico.
- 4 `Porcentaje->AgregaDato( 0.0, "1", &Porcentaje);` se establece a **Leucemia** como hipótesis del tipo de cáncer. Cuando no existe ningún síntoma registrado o se está iniciando la sesión de preguntas el establecimiento de esta hipótesis resulta de gran ayuda ya que de no ser así el mecanismo de inferencia no tendrá punto de partida para iniciar la búsqueda. Cuando ya existen síntomas que establezcan como hipótesis algún otro tipo de cáncer, esta instrucción no afectará en nada dicha suposición debido a que el porcentaje que se agrega es 0.0.
- 5 `for(ClaseLista *Aux = Porcentaje; ;Found && Aux;)` inicia un ciclo para buscar el siguiente síntoma a preguntar, este ciclo se repite hasta que se encuentre dicho síntoma, se agoten las preguntas, o se hayan descartado todas las hipótesis (Aux).
  - 5.1 En el bloque de código dentro del anterior ciclo `for` se establece la consulta en la base de conocimientos para buscar un síntoma que no existe tanto en el padecimiento actual del paciente (`Padecimiento`) como en los síntomas que ha presentado (`Síntomas`) para la hipótesis del tipo de cáncer (`Aux->Index`).
  - 5.2 `if(!Consulta->IsEmpty())` si la consulta no es vacía.
    - 5.2.1 `Found = True;` se establece a verdadero la bandera para indicar que encontró el síntoma a preguntar.
    - 5.2.2 `iPregunta = Consulta->FieldByName("ID_INDICE_REAL")->AsInteger;` asigna el síntoma a preguntar.

5.2.3 Si la consulta es vacía `Aux->Next(&Aux);` intenta buscar un síntoma para la siguiente hipótesis.

6 `if(found)` si se encontró el síntoma.

6.1 `FormExpertSys->Termino(Texto,iPregunta);` lanza la siguiente pregunta al usuario.

6.2 `return true;` regresa un valor verdadero que indica que se pudo establecer satisfactoriamente la pregunta al usuario.

7 `return false;` regresa un valor falso que indica que no se pudo establecer satisfactoriamente la pregunta al usuario.

La implementación del PLN y del mecanismo de inferencia en C++ Builder versión 5.0 se puede analizar en el capítulo IV de resultados, donde se observa la implementación adicional de una herramienta estadística para el análisis de la base de conocimientos y de la base de datos de consulta.

# Capítulo IV

## *Resultados*

A continuación se desglosa el funcionamiento que presenta el sistema durante el desarrollo de una consulta.

El sistema al ser consultado por el usuario (entiéndase por usuario médico familiar, pediatra, residente, prestador de servicio social en alguna institución de salud en el área de medicina, estudiante de medicina, etc.), solicita en primera instancia datos personales del paciente concernientes al nombre, edad, domicilio, estado en que radica actualmente, estado de nacimiento, sexo y fecha de nacimiento. En la *figura 17* se observa la pantalla para la inserción de datos personales.

Datos Personales

ID:  Fecha: 14/04/02

Nombre

Apellido Paterno: MONTAÑO Apellido Materno: GOMEZ Nombre(s): JOSE DAVID

Domicilio

Estado: VERACRUZ Ciudad: COATZACOALCOS Localidad: COATZACOALCOS

C.P.: 96536 Colonia: SAN SILVERIO Calle y No.: ARROYO # 210

Telefono: 019212682456

Adicionales

Sexo: MASCULINO Fecha de Nac.: 21/09/89 Estado de Nacimiento: TLAXCALA

DD/MM/AA

**Figura 17. Pantalla de ingreso de los datos personales del paciente.**

Posteriormente solicita al usuario datos adicionales del paciente referentes a su peso al nacer, edad de los padres, si el paciente presenta alguna malformación congénita, si presentó alguna enfermedad exantemática (rubéola, varicela, etc.), si durante el primer año de vida tuvieron necesidad de hospitalizarlo por más de 24 horas y por qué causa, el nivel socioeconómico de la familia del paciente y si se ha presentado algún tipo de cáncer en sus familiares. En la *figura 18* se muestra la pantalla de inserción de los datos adicionales.

**Datos Personales Adicionales**

Peso al Nacer: 8500 gr

Edad del Padre: 40 Años      Edad de la Madre: 41 Años

Internamientos Hospitalarios en el 1er. año de vida. Tiempo mayor de 24 hrs.: SI      Causa del internamiento: ATAQUES ASMATICOS

Malformaciones: NO      Familiares con Cáncer: NO

Enfermedades Exantemáticas:

- ☐ SARAMPION
- ☒ VARICELA
- ☒ RUBEOLA
- ☐ ESCARLATINA
- ☐ OTRAS

Nivel Socioeconómico: MEDIO

Continuar      Cancelar      Regresar

**Figura 18. Pantalla de ingreso de los datos adicionales del paciente.**

Continuando con la consulta se pide al usuario que ingrese los factores de riesgo, es decir, factores que representan posibles agentes cancerígenos que propician el desarrollo del cáncer, aunque no se tenga aún la certeza del nivel de influencia de éstos. Pregunta al usuario de donde proviene el agua que toma el paciente, si su lugar de residencia se encuentra cerca de arroyos de desagüe, si cerca de su casa existe alguna subestación eléctrica, si se le practicaron estudios de rayos X a la madre del paciente durante el embarazo, si el paciente ha recibido tratamiento de rayos X en un periodo de un año, si cerca de la casa del paciente existen torres de alta tensión y si ha estado en contacto con algún aparato electrodoméstico, producto químico o se le ha suministrado algún antibiótico.

En la *figura 19* se observa la pantalla de ingreso de los factores de riesgo.



Figura 19. Pantalla de ingreso de los factores de riesgo.

Por último solicita información concerniente a la consulta actual del paciente. Esta información consiste de la fecha de la consulta, peso actual, estatura actual, oficio de los padres y la sintomatología actual presentada por el paciente. En la *figura 20* se presenta la pantalla de inserción de los datos de la consulta.

Figura 20. Pantalla de inserción de los datos de la consulta actual.

Una vez conocida esta información, el sistema experto plantea una hipótesis; esto es, a través de la sintomatología actual presentada por el paciente e introducida por el usuario y después del análisis de ésta, el sistema trata de determinar el tipo de cáncer que es más probable que

presente el paciente, ya sea leucemia, linfoma, tumor en sistema nervioso central, etc.; es decir, selecciona un objeto y trata de aseverar o negar el mismo; formula o toma el camino de búsqueda y el conjunto de preguntas que le ayudarán a corroborar o descartar la suposición del tipo de cáncer del paciente y formular un nuevo camino de búsqueda si así se requiere. Para poder verificarlas comprueba primero la exactitud de las premisas de la regla. Esto se realiza mediante una búsqueda de enunciados correspondientes en la base de conocimientos.

Como primera etapa el Mecanismo de Inferencia toma como entrada el motivo de consulta del paciente y algún dato adicional opcional que el usuario del sistema pueda observar relevante para que sea utilizado como parte del diagnóstico para el sistema.

Con estos datos iniciales el sistema trata de trazar el área de búsqueda para llevar a cabo su diagnóstico o la forma y el camino en que llegará a él. Esto lo realiza tomando en cuenta que en la base de conocimientos existe una relación de signos y síntomas, una por cada uno de los diferentes tipos de cáncer analizados por el sistema, que contiene la influencia o incidencia de determinado signo y/o síntoma en relación con el cáncer en que se presenta, y le va dando un peso específico para cada elemento introducido por el usuario.

Con esta información el Sistema hace un análisis de cuál es la mayor probabilidad de que el paciente presente cáncer y de qué tipo (de los diez tipos de cáncer que abarcamos) podría ser. Una vez obtenida la mayor probabilidad, el sistema comienza con la elaboración de preguntas en base al tipo de cáncer que tiene contemplado que podría ser el resultado del diagnóstico. Tomando como base los primeros síntomas introducidos por el usuario, el sistema empieza a

hacer preguntas sobre los signos y/o síntomas que pudiera presentar el paciente, véase la *figura 21*.



Figura 21. Fase de realización de preguntas del sistema al usuario.

Aunque esto no quiere decir que no considere ir tomando en cuenta a los otros tipos de cáncer implicados en el sistema; en todo momento y aunque las preguntas se elaboran tomando en consideración a un sólo tipo de cáncer, en caso de que se tenga un registro en la base de conocimientos de que ese signo y/o síntoma guarda alguna relación con cualquiera de los otros tipos de cáncer, el sistema se encarga de incrementar el índice de probabilidad de cáncer para cada uno de los tipos analizados en el diagnóstico final.

Si en algún momento el mecanismo de inferencia, en base a las preguntas que fueron realizadas, detecta que la probabilidad de que el diagnóstico final se inclina hacia cualquiera de los otros tipos de cáncer, el área de búsqueda y en este caso la elaboración de preguntas se realiza en base a los conocimientos que se tengan registrados para el cáncer que en ese momento tenga la mayor probabilidad.

Si el sistema detecta que el tipo de cáncer rebasa un máximo establecido para asegurar que el paciente tiene cáncer se informa al usuario a través de una alarma para que revise el diagnóstico arrojado por el mecanismo de inferencia. En la *figura 22* se observa el mensaje de alarma.



Figura 22. Mensaje de alarma de posible caso de cáncer.

La *figura 23* muestra el diagnóstico que proporciona el mecanismo de inferencia después de una sesión de preguntas.

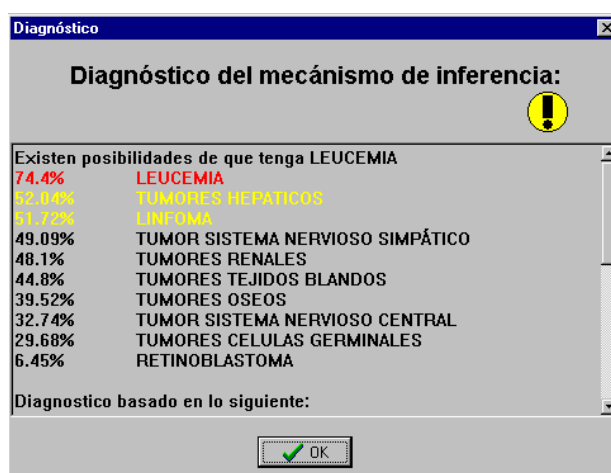
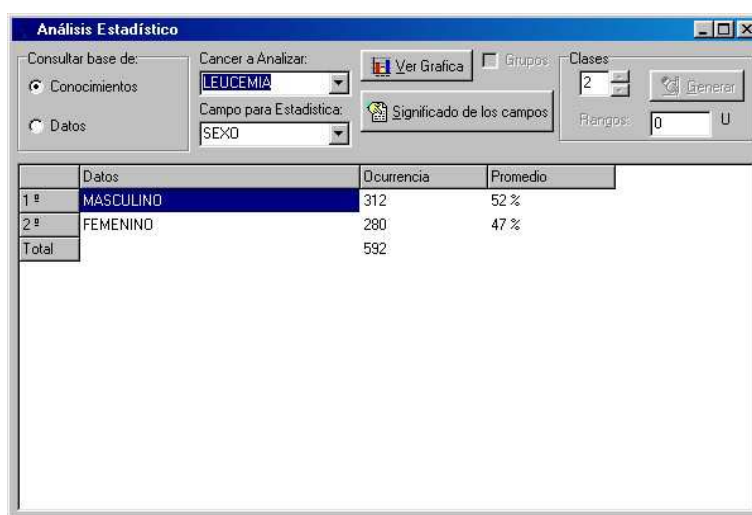


Figura 23. Diagnóstico del mecanismo de inferencia.

La sesión de preguntas continuará hasta agotar todas las preguntas registradas en la base de conocimientos aunque, como es de suponer, el índice de probabilidad que de estos síntomas se obtenga cada vez se irá aproximando al mínimo con lo cual los porcentajes obtenidos no variarán significativamente, es decir, los síntomas que eliminen la mayor incertidumbre serán los primeros por los que el sistema pregunte al usuario si los presenta el paciente y, conforme transcurra el número de preguntas, el índice porcentual de síntomas subsecuentes cada vez

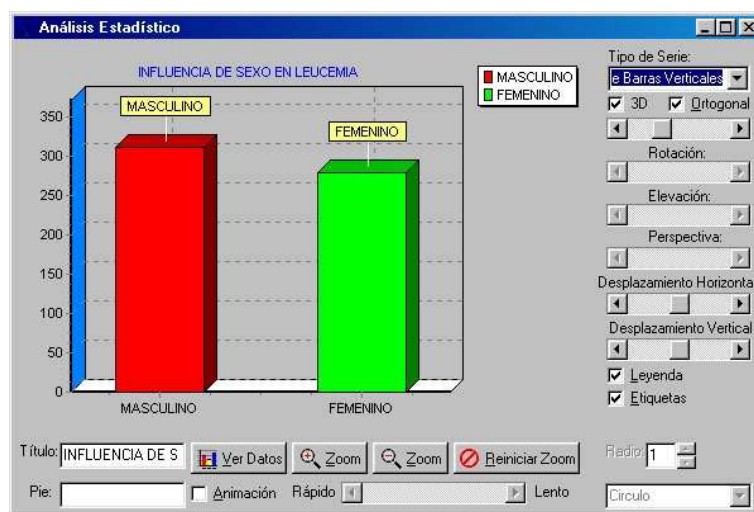
será mucho menor con tendencia a un índice del 0.00% para el tipo de cáncer que se considera como hipótesis, aunque para otras neoplásicas malignas es posible que tenga un índice porcentual mucho mayor.

El sistema experto cuenta con una herramienta de análisis porcentual. Esta herramienta permite representar el análisis porcentual de los elementos que componen la base de conocimientos y la base de datos de las consultas, además permite generar una gráfica del resultado de análisis porcentual. En la *figura 24* se presenta la pantalla de los datos del análisis porcentual.



**Figura 24. Pantalla de los datos del análisis porcentual.**

En la pantalla de datos del análisis porcentual se puede seleccionar la base a consultar, si es la base de conocimientos o la base de datos de las consultas. Se elige el cáncer a analizar y el campo para la estadística. La gráfica del análisis porcentual se presenta en la pantalla de la gráfica del análisis del porcentual a través del botón “Ver Gráfica”. En la figura 25 se observa la pantalla de la gráfica del análisis porcentual.



**Figura 25. Pantalla de la gráfica del análisis porcentual.**

En la pantalla de la gráfica del análisis porcentual se puede seleccionar el tipo de serie con el que se quiere representar la gráfica ( gráfica de barras horizontales o verticales, gráfica de pastel, etc), si se quiere representar la serie con una gráfica en tercera dimensión. Se pueden configurar parámetros que aparezcan en la gráfica como el título, el pie de la gráfica, la leyenda, la etiqueta y los parámetros para la graficación en tercera dimensión.

# Conclusiones

Al final del proyecto se contó con la implementación de un mecanismo de búsqueda con el cual se integró el sistema experto auxiliar en el diagnóstico temprano de cáncer en niños.

Durante las pruebas preliminares el mecanismo de búsqueda implementado permitió que el sistema experto presentara un porcentaje de certidumbre del 75%. Este porcentaje se obtuvo probando el sistema con 8 niños con diagnóstico positivo de cáncer del Hospital N° 36 de la Cd. de Coatzacoalcos, Ver., de los cuales el sistema experto arrojó un diagnóstico correcto del tipo exacto de cáncer que presentaban los niños en 6 de los casos. En los otros 2 casos el sistema detectó que sí existía un diagnóstico positivo de cáncer, aunque el tipo no coincidió. Esto es, el tipo de cáncer que el sistema diagnosticó no era el que en realidad presentaban los pacientes.

Sin embargo, debido a que el prototipo tiene como objeto ser aplicado al sector salud, se le realizará una validación para conocer su grado de sensibilidad y especificidad como herramienta de diagnóstico. Para realizar esta validación actualmente están desarrollando, de forma conjunta, la M. C. Ana Estela Ruiz Linares, Directora de Proyectos Biomédicos del Departamento de Ingeniería en Electrónica del Instituto Tecnológico de Minatitlán y el Dr. Leoncio Miguel Guzmán García, Jefe de Proyectos de Investigación del Hospital General de Zona N° 36 del IMSS localizado en la Cd. de Coatzacoalcos, Ver., el protocolo de validación del sistema experto.

Es necesario resaltar la importancia de este tipo de proyectos ya que contribuye a la vinculación del Instituto Tecnológico de Minatitlán con otras instituciones preocupadas por el desarrollo tecnológico de nuestra región. Para la carrera de Ingeniería en Sistemas, el



desarrollo de este trabajo permitió al departamento iniciarse en el campo de investigación tecnológica con financiamiento de una institución externa como lo es el CONACYT, y obtener el primer lugar en el área de Ciencias Computacionales del XVI Concurso Nacional de Creatividad de los Institutos Tecnológicos fase regional.

Finalmente, en el aspecto personal, participar en un proyecto de investigación tecnológica con carácter multidisciplinario, me permitió aplicar los conocimientos adquiridos en el aula, en el desarrollo de un prototipo encaminado a encontrar una solución a un problema de salud pública como lo es el cáncer en niños.

# *Anexo A*

Cuestionario aplicado en el proyecto:

“EPIDEMIOLOGÍA DE LAS NEOPLASIAS MALIGNAS EN NIÑOS DERECHOHABIENTES DEL  
INSTITUTO MEXICANO DEL SEGURO SOCIAL”

1. Núm. de Afiliación \_\_\_\_\_ FOLIO \_\_\_\_\_
  2. Nombre del paciente .....  
(apellido paterno) (apellido materno) (nombres)
  3. Sexo \_\_\_\_ 1)Masculino 0)Femenino
  4. Fecha de nacimiento ..... mes de nac \_\_\_\_ año de nac \_\_\_\_
  5. Fecha del diagnóstico ..... mesdg \_\_\_\_ añodg \_\_\_\_
  6. Edad al diagnóstico \_\_\_\_ (meses)
  7. Tipo de cáncer .....  
Clasificación morfológica \_\_\_\_\_  
Clasificación topográfica \_\_\_\_\_  
Clasificación Birch y Marsden \_\_\_\_\_
  8. Estudio histopatológico \_\_\_\_ 1)Si 0)No Num. de registro \_\_\_\_\_
  9. Diagnóstico histopatológico .....  
.....
  10. Estadio del tumor o tipo de Leucemia (FAB) \_\_\_\_ .....
  11. Ciudad de residencia lugardere \_\_\_\_ .....
  12. Domicilio .....
  13. Teléfono .....
  14. Clínica IMSS \_\_\_\_\_
  15. Delegación IMSS \_\_\_\_\_
  16. Delegación política \_\_\_\_\_
  17. Fue atendido previamente en otro hospital \_\_\_\_ 1)Si 0)No 2)No sabe
- ANTECEDENTES HEREDOFAMILIARES
18. Familiares con cáncer \_\_\_\_ 1)Si 0)No 2)No sabe
  19. Parentesco parefamca \_\_\_\_\_
  20. Tipo de cáncer familiar \_\_\_\_\_

## ANTECEDENTES PERSONALES NO PATOLÓGICOS

21. Lugar de nacimiento \_\_\_\_\_
22. Mes de concepción \_\_\_\_\_
23. Orden al nacimiento \_\_\_\_\_ Embarazo de término \_\_\_\_\_ 1)Si 2)No
24. Peso al nacer \_\_\_\_\_ (gramos)
25. Ocupación del padre \_\_\_\_\_  
(especificar)
26. Ocupación de la madre \_\_\_\_\_  
(especificar)
27. Escolaridad del padre \_\_\_\_\_  
(escribir años cursados, posteriormente se recodificará)
28. Escolaridad de la madre \_\_\_\_\_  
(escribir años cursados, posteriormente se recodificará)
- 01)Analfabeta                      05)Secundaria incompleta                      09)Licenciatura  
02)Sabe leer y escribir                      06)Secundaria completa                      10)Otros  
03)Primaria incompleta                      07)Bachillerato  
04)Primaria completa                      08)Técnico
29. Salario mensual del padre \$ \_\_\_\_\_
30. Salario mensual de la madre \$ \_\_\_\_\_
31. Salario mensual conjunto \$ \_\_\_\_\_
32. Núm. de cuartos en la habitación \_\_\_\_\_
33. Núm. de cuartos para dormir Cdormir \_\_\_\_\_
34. Núm. de personas que viven en la casa \_\_\_\_\_
35. Agua intradomiciliaria \_\_\_\_\_ 1)Si 0)No
36. Eliminación de excretas \_\_\_\_\_ 1)T. Inglés 2)Letrina 3)Fecalismo
37. Nivel socioeconómico \_\_\_\_\_ 1)Alto 2)Medio 0)Bajo
- 37a. El piso de la habitación es de \_\_\_\_\_  
1)Mosaico 2)Cemento 3)Madera 4)Tierra 5)Otra
38. Edad del padre a la concepción del hijo \_\_\_\_\_ (años)
39. Edad de la madre a la concepción del hijo \_\_\_\_\_ (años)
40. Edad del padre al diagnóstico del hijo \_\_\_\_\_ (años)

41. Edad de la madre al diagnóstico del hijo \_\_\_\_ (años)

#### ANTECEDENTES PERSONALES PATOLÓGICOS

42. Enfermedades exantemáticas \_\_\_\_ (tachar las que haya padecido)

00)Ninguna 01)Sarampión 02)Varicela 04)Rubeola 16)Escarlatina  
32)Otras .....

43. Internamientos hospitalarios durante el primer año de vida \_\_\_\_ 1)Si 2)No 3)No se sabe

44. Causa Causahp \_\_\_\_ .....

45. Malformaciones \_\_\_\_ 1)Si 0)No 3)No sabe

46. Tipo \_\_\_\_ .....

#### ANTECEDENTES RELACIONADOS CON EL PADECIMIENTO

##### PADECIMIENTO ACTUAL

(Haga una síntesis del padecimiento del niño con fechas de atención y número de consultas recibidas)

.....  
.....  
.....  
.....  
.....  
.....

47. Fecha de inicio de signos y síntomas : .....

Días iniciodías \_\_\_\_ mes iniciomes \_\_\_\_ año inicioaño \_\_\_\_

48. ¿Cuales fueron los signos y síntomas que presentó?

.....  
.....  
.....

49. ¿Por qué llevó a su hijo a consulta?

.....  
.....

50. ¿Cuál fue el primer signo y síntoma que tuvo su hijo?

Signo1 .....

51. Además del primer signo o síntoma, ¿Qué otros signos y síntomas presentó al inicio del padecimiento?

Signo2 \_\_\_\_\_ Signo5 \_\_\_\_\_

Signo3 \_\_\_\_\_ Signo6 \_\_\_\_\_

Signo4 \_\_\_\_\_ Signo7 \_\_\_\_\_

52. Localización del tumor \_\_\_\_ (tachar el o los signos donde se localizó el tumor)

001)Cabeza      002)Cuello      004)Tórax      008)Abdomen      016)Pelvis

032)Genitales      064)Extremidades sup.      128)Extremidades inf.

256)Otras .....

53. Tuvo metástasis al diagnóstico \_\_\_\_ (tachar si hubo metástasis)

001)SNC    002)Pulmón    004)Hígado    008)Riñón    016)Huesos

032)No hubo metástasis    064)Otra .....

54. Sitio del tumor \_\_\_\_

1)Unilateral derecho      2)Bilateral      3)Generalizado      4)Unilateral izquierdo

#### TIEMPO DE DIAGNÓSTICO

55. Antes de ser atendido en el IMSS fue atendido en otra institución antesimss \_\_\_\_ 1)Si      0)No

56. Fecha de atención .....

57. Cuál institución \_\_\_\_

001)Pediatra privado      002)M-General privado      004)ISSSTE      008)INP

016)SSA      032)DDF      064)Ejército      128)Naval      256)HIM

512)Otra.....      999)No se aplica

58. Cuál fue el diagnóstico dgprivada \_\_\_\_ .....

59. Cuál fue el tratamiento Txprivado1 \_\_\_\_\_

Txprivado2 \_\_\_\_\_

Txprivado3 \_\_\_\_\_

60. Número de consultas a nivel privado o en otra institución IMSS \_\_\_\_

61. Fue atendido en 1er nivel de atención del IMSS \_\_\_\_      1)Si      0)No

62. Cual fue el diagnóstico dgIMSS1 \_\_\_\_ .....

63. Cuál fue el tratamiento TxIMSS1 \_\_\_\_\_  
TxIMSS2 \_\_\_\_\_  
TxIMSS3 \_\_\_\_\_
64. Número de consultas en primer nivel IMSS \_\_\_\_
65. Fecha 1 IMSS .....
66. Fue atendido en 2º nivel de atención del IMSS \_\_\_\_ 1)Si 0)No
- 66a. Antes de ser atendido en el 2º nivel del IMSS fue atendido en otra institución antes2imss \_\_\_\_ 1)Si  
0)No
- 66b. Cuál institución antes2i \_\_\_\_  
001)Pediatra privado 002)M-General privado 004)ISSSTE 008)INP  
016)SSA 032)DDF 064)Ejército 128)Naval 256)HIM  
512)Otra..... 999)No se aplica
- 66c. Cuál fue el diagnóstico antesdg2p \_\_\_\_
- 66d. Cuál fue el tratamiento antes2tx1p \_\_\_\_\_  
antes2tx2p \_\_\_\_\_  
antes2tx3p \_\_\_\_\_
- 66e. Número de consultas a nivel privado o en otra institución antes2conp
67. Cuál fue el diagnóstico dgIMSS2 \_\_\_\_ .....  
.....
68. Cuál fue el diagnóstico txIMSS21 \_\_\_\_\_  
txIMSS22 \_\_\_\_\_  
txIMSS23 \_\_\_\_\_
69. Número de consultas en segundo nivel IMSS \_\_\_\_
70. Fecha 2 IMSS.....
71. Fecha de atención de 3er. Nivel del IMSS.....
- 71a. Antes de ser atendido en el 3er. Nivel del IMSS fue atendido en otra institución en otra institución antes3imss  
\_\_\_\_ 1)Si 0)No
- 71b. Cuál institución antes3i \_\_\_\_  
001)Pediatra privado 002)M-General privado 004)ISSSTE 008)INP  
016)SSA 032)DDF 064)Ejército 128)Naval 256)HIM  
512)Otra..... 999)No se aplica

71c. Cuál fue el diagnóstico antesdg3p \_\_\_\_\_.  
.....

71d. Cuál fue el tratamiento        antes3tx1p \_\_\_\_\_  
  antes3tx2p \_\_\_\_\_  
  antes3tx3p \_\_\_\_\_

71e. Número de consultas a nivel privado o en otra institución antes3conp

72. Número de consultas en tercer nivel IMSS \_\_\_\_

73. Tiempo transcurrido entre el inicio de los síntomas relacionados con el padecimiento y primera consulta en cualquier institución t1consulta \_\_\_\_ (días).

(Se obtiene con la fecha de inicio de los síntomas y la 1ª vez que solicitó consulta con cualquier sitio).

74. Tiempo transcurrido entre el inicio de los síntomas y primera atención en el IMSS t1IMSS (días).

(Se obtiene con la fecha de inicio de los síntomas y la fecha de atención en el IMSS).

75. Tiempo transcurrido entre primera consulta en 1er. Nivel en el IMSS y primera consulta 2º nivel en el IMSS t2IMMS \_\_\_\_ (días).

(Se obtiene con la fecha de atención en 1er nivel en el IMSS y la fecha de atención de 2º nivel en el IMSS siempre y cuando haya sido atendido en los dos niveles, pero no haber sido visto en 1º ó 2º nivel).

76. Tiempo transcurrido entre primera consulta en 2º nivel en el IMSS y primera atención en 3er. Nivel t3IMSS \_\_\_\_ (días).

(Se obtiene con la fecha de atención en 2º nivel en el IMSS y la fecha de atención de 3er nivel en el IMSS siempre y cuando haya sido atendido en los dos niveles, pero no haber sido visto en 2º nivel).

76a. ¿Se hizo el diagnóstico en el segundo nivel? \_\_\_\_        1)Si        0)No

76b. Tiempo transcurrido entre el inicio de los síntomas y su diagnóstico en el segundo nivel IMSS tiempo t1a2IMSS \_\_\_\_ (días).

76c. Tiempo transcurrido entre la 1ª atención del IMSS y de la del diagnóstico en el 2º nivel IMSS t1a2IMSS \_\_\_\_ (días).

76d. Tiempo transcurrido entre el inicio de los síntomas y su diagnóstico clínico en 2º nivel IMSS timepodx2 \_\_\_\_ (días).

77. Tiempo transcurrido entre el inicio de los síntomas y su diagnóstico en el 3er nivel de atención en el IMSS tini3IMSS \_\_\_\_ (días).

78. Tiempo transcurrido entre la 1ª atención en el IMSS y la del diagnóstico en el 3er nivel IMSS t1a3IMSS \_\_\_\_ (días).



79. Tiempo transcurrido para realizar el diagnóstico clínico en 3er nivel del IMSS tdx3 \_\_\_\_ (días).

80. Tiempo transcurrido entre el diagnóstico clínico en el 3er nivel del IMSS tdx3 \_\_\_\_ (días).

81. Número unidad médica \_\_\_\_

81a. Tiempo transcurrido entre el ingreso a 3er nivel e inicio de QT tpo3nivel \_\_\_\_

El cuestionario fue contestado por \_

1)PADRE

2)MADRE

3)OTROS

4)AMBOS PADRES

El cuestionario fue aplicado en \_\_\_\_ (escribir la ciudad en donde fue aplicado) .....

.....

OBSERVACIONES.....

.....

.....

# *Anexo B*

Resultados utilizados para la creación de la Base de Conocimientos.

Se registraron 562 neoplasias y de acuerdo al esquema propuesto por Birch y Marsden se distribuyeron de siguiente manera:

TIPO DE NEOPLASIA	NUMERO	%
<b>I Leucemias</b>	<b>242</b>	<b>43.0</b>
Leucemia linfoblástica aguda	196	
Leucemia mieloblástica aguda	46	
<b>II Linfomas y otras enfermedades Reticuloendoterilaes</b>	<b>72</b>	<b>12.8</b>
Enfermedad Hodgkin	20	
Linfoma no Hodgkin	31	
Linfoma Burkitt	7	
Histiocitosis	14	
<b>III Neoplasias intracraneales e Intraespinales</b>	<b>56</b>	<b>9.9</b>
<b>IV Tumores del sistema Nervioso simpático</b>	<b>15</b>	<b>2.7</b>
<b>V Retinoblastoma</b>	<b>20</b>	<b>3.6</b>
<b>VI Tumores renales</b>	<b>20</b>	<b>3.6</b>
<b>VII Tumores hepáticos</b>	<b>19</b>	<b>3.4</b>
<b>VIII Osteosarcoma Sarcoma de Ewing</b>	<b>35</b>	<b>6.2</b>
<b>IX Sarcoma de tejidos blandos</b>	<b>16</b>	<b>2.8</b>
<b>X Tumores de células germinales</b>	<b>31</b>	<b>5.5</b>
<b>XI Carcinoma y otras neoplasias Epiteriales</b>	<b>34</b>	<b>6.0</b>
<b>TOTAL</b>	<b>562</b>	<b>100.0</b>

SEXO		
	NUMERO	%
Masculino	260	46.3 %
Femenino	302	53.7%
Total	562	100.0

EDAD AL DIAGNOSTICO		
EDAD	NUMERO	%
< 1	78	13.9
1 – 4	124	22.0
5 – 9	166	29.5
10 y mas	194	34.5
Total	562	100.0

ESTADIFICACION DE NEOPLASIAS EN NIÑOS MENORES 15 AÑOS		
ESTADIO	NUMERO	%
01 (I)	12 (8.9)	3.2
02 (II)	26 (19.3)	7.0
03 (III)	60 (44.4)	16.2
04 (IV)	37 (27.4)	10.0
Total	135	100

LUGAR DE RESIDENCIA DE NIÑOS DERECHO HABIENTES DEL IMSS 1997 – 1998		
RESIDENCIA	NUMERO	%
Aguascalientes	3	0.5
Baja California Norte	3	0.5
Baja California Sur	1	0.2
Campeche	4	0.7
Coahuila	14	2.5
Colima	8	1.4
Chiapas	21	3.7
Chihuahua	14	2.5
Distrito Federal	98	17.4
Durango	10	1.8
Guanajuato	19	3.4
Guerrero	14	2.5
Hidalgo	12	2.1
Jalisco	54	9.6
Estado de México	128	22.8
Michoacán	12	2.1
Morelos	18	3.2
Nayarit	6	1.1
Nuevo León	2	0.4
Oaxaca	1	0.2
Puebla	1	0.2
Querétaro	10	1.8
Quintana Roo	8	1.4
San Luis Potosí	1	0.2
Sinaloa	14	2.5
Sonora	22	3.9
Tabasco	1	0.2
Veracruz	37	6.6
Yucatán	14	2.5
Zacatecas	12	2.1
Total	562	100.0

ANTECEDENTES FAMILIARES DE CANCER		
AHF CANCER	NUMERO	%
Si	388	69.0
No	174	31.0
Total	562	100.0

INTERNAMIENTOS HOSPITALARIOS DURANTE PRIMER AÑO DE VIDA EN NIÑOS CON CANCER DERECHOHABIENTES DEL IMSS 1997 – 1998		
Hospitalización	FRECUENCIA	%
No	484	86.1
Si	78	13.9
Total	562	100.0

PRINCIPLAES MALFORMACIONES CONGENITAS EN NIÑOS CON CANCER DERECHOHABIENTES DEL IMSS 19978 – 1998		
Malformación	FRECUENCIA	%
Síndrome Down	5	22.7
Corazón	3	13.6
Endocrino	2	9.1
Pie equinovaro	2	9.1
Digestivo	2	9.1
Otras	8	36.3
Total	22	100.0

ANTECEDENTES DE ENFERMEDAD EXANTEMATICA EN NIÑOS CON CANCER DERECHOHABIENTES DEL IMSS 1997 – 1998		
ENFERMEDAD	FRECUENCIA	%
Ninguna	294	52.4
Sarampión	19	3.4
Varicela	110	19.6
Sarampión y varicela	13	2.3
Rubéola	14	2.5
Sarampión y rubéola	1	0.2
Varicela y rubéola	10	1.8
Sarampión, varicela y rubéola	6	1.1
Escarlatina	4	0.7
Escarlatina y sarampión	1	0.2
Varicela y escarlatina	4	0.7
Sarampión, rubéola y escarlatina	1	0.2
Sarampión, varicela, rubéola y escarlatina	3	0.5
Otras	81	14.4
Total	561	100.0

PRINCIPALES CAUSAS DE HOSPITALIZACION DURANTE EL PRIMER AÑO DE VIDA EN NIÑOS CON CANCER DERECHOHABIENTES DEL IMSS		
CAUSA	FRECUENCIA	%
GEPI	8	10.5
Prematurez	8	10.5
Deshidratación	8	10.5
Neumonía	5	6.6
Bronconeumonía	5	6.6
Bronquitis	4	5.3
Cardiopatía	3	4.0
No sabe	2	2.6
Infección respiratoria	2	2.6
Pie equinovaro	2	2.6
Ictericia	2	2.6
Hipoxia	2	2.6
Otras	27	35.5
Total	76	100.0

MALFORMACIONES CONGENITAS EN NIÑOS CON CANCER DERECHOHABIENTES DEL IMSS 1997 - 1998		
MALFORMACION	FRECUENCIA	%
No	537	96.0
Si	22	4.0
Total	559	100.0

PRINCIPALES LUGARES DE METASTASIS EN NIÑOS CON CANCER DERECHOHABIENTES DEL IMSS 1997 – 1998		
METASTASIS	FRECUENCIA	%
SNC	15	2.7
Pulmón	11	2.0
Hígado	3	0.5
Pulmón e hígado	2	0.4
Riñón	1	0.2
Riñón e hígado	1	0.2
Huesos	4	0.7
Huesos y SNC	1	0.2
Sin metástasis	495	88.1
Otras	29	5.1
Total	562	100.0

PRINCIPALES MOTIVOS DE CONSULTA		
MOTIVO DE CONSULTA	FRECUENCIA	%
A nivel cuello Bolita en cuello Inflamación región cervical	30	5.4
Palidez	22	3.9
Dolor abdominal	20	3.6
Dolor miembros inferiores	17	3.0
Fiebre	17	3.0
Cefalea	14	2.5
Vómito	11	2.0
Infección vías respiratorias	10	1.8
Epistaxis	9	1.6
Dolor miembros inferiores y superiores	8	1.4
Leucocoria	8	1.4
Distensión abdominal	8	1.4
Fiebre y palidez	8	1.4
Aumento volumen testículo	7	1.2
Aumento volumen abdomen	6	1.1
Aumento volumen miembros superiores	6	1.1
Otras	361	64.2
Total**	562	100.0

LOCALIZACION ANATOMICA DEL TUMOR EN NIÑOS DERECHOHABIENTES DEL IMSS 1997 - 1998		
LOCALIZACION	FRECUENCIA	%
Cabeza	97	17.3
Cuello	30	5.3
Hígado	12	2.1
Abdomen	92	16.4
Genitales	12	2.1
Extremidades	55	9.8
Otras	264	47.0
Total	562	100.0

SITIO ANATOMICO DE LA NEOPLASIA EN NIÑOS DERECHOHABIENTES DEL IMSS 1997 - 1998		
SITIO	FRECUENCIA	%
Unilateral derecho	85	18.1
Bilateral	13	2.8
General	283	60.2
Unilateral izquierdo	89	18.9
Total	470	100.0

# *Anexo C*

# GLOSARIO

**ÁRBOL.** Estructura de datos construida con un nodo llamado padre, que a su vez puede tener ramas, llamados hijos y a su vez pueden tener más ramificaciones, por su forma se le conoce como el nombre de árbol.

**BASE DE CONOCIMIENTOS.** Representa un depósito de las primitivas del conocimiento (por ejemplo, hechos fundamentales, reglas de procedimientos y heurísticas) disponibles para el sistema.

**CAMINO SOLUCIÓN.** Un grafo dirigido de los nodos visitados que nos llevan a la solución.

**CÁNCER.** Designación colectiva de todos los tumores malignos. El cáncer es la enfermedad ocasionada por células del organismo que pierden sus mecanismos normales de control y presentan un crecimiento desordenado.

**CARCINÓGENO.** Agente que ocasiona el cáncer.

**ESPACIO DE BÚSQUEDA.** El conjunto de todos los nodos.

**ETIOLOGÍA.** Estudio sobre las causas de las cosas. Parte de la medicina que estudia las causas de las enfermedades.

**FACTOR DE RIESGO.** Son aquellas situaciones, en donde la posibilidad de desarrollar aumenta debido a ciertos factores que influyen directa o indirectamente, ejemplo: residuos peligrosos, aguas contaminadas.

**HEURÍSTICA.** Información acerca de la posibilidad de que un nodo específico sea mejor para intentar la próxima elección que cualquier otro nodo.



**INTELIGENCIA ARTIFICIAL.** Área de la computación enfocada a la simulación del proceso de razonamiento humano mediante la aplicación de técnicas computacionales.

**INVESTIGACIÓN EPIDEMIOLÓGICA.** Investigación que se realiza para determinar los factores que influyen en que una enfermedad se desarrolle más fácilmente, también abarca el estudio de signos y síntomas que se presentan en una enfermedad con mayor incidencia.

**LENGUAJE NATURAL.** El lenguaje natural es el medio que utilizamos de manera cotidiana para establecer nuestra comunicación con las demás personas.

**MALFORMACIÓN CONGÉNITA.** Alteración de la morfología de un órgano o estructura, presente ya en el momento del nacimiento. Las causas pueden dividirse, a grandes rasgos, en los siguientes grupos: 1)Factores ambientales (exógenos), 2)Factores hereditarios (genéticos), 3)Factores genéticos de origen cromosómico y, 4)Factores desconocidos.

**MECANISMO DE INFERENCIA.** Implementación de técnicas de búsqueda utilizadas en la IA para encontrar una solución a determinado problema.

**META.** El nodo objeto de la búsqueda.

**METÁSTASIS.** A medida que las células cancerosas crecen y se multiplican, forman una masa de tejido que puede invadir órganos adyacentes, esta diseminación se conoce como metástasis.

**MIT.** Massachusetts Institute Technologic.

**MORFOLOGÍA.** Rama de la biología. Ciencia que tiene por objeto el estudio y la descripción de los caracteres somáticos de las especies vegetales y animales.

**NEOPLASIA.** Formación de un tejido con carácter tumoral, generalmente maligno.

**NODO.** Un punto concreto y posible fin.

**NODO TERMINAL.** Un nodo que es fin de un camino.

**ONCOLOGÍA.** Especialidad en el área de medicina que se enfoca al estudio del tratamiento adecuado que debe de llevar un paciente cuando ha sido diagnosticado el cáncer.

**PROCESAMIENTO DE LENGUAJE NATURAL.** Subdominio de la Inteligencia Artificial avocado a la tarea de desarrollar programas que poseen alguna capacidad de “entendimiento” a un lenguaje natural en orden para lograr sus metas.

**RAD.** Rapid Applications Development. Herramientas de desarrollo rápido de aplicaciones, una de ellas el Borland C++ Builder.

**SISTEMA EXPERTO.** Aplicación informática que soluciona problemas complicados que de otra manera exigirían ampliamente la pericia humana. Para lograr esto, se simula el proceso de razonamiento humano mediante la aplicación específica de conocimientos y de inferencias.

**UBICACIÓN ANATÓMICA.** Se refiere al lugar exacto del cuerpo en donde el paciente siente o presenta un signo o síntoma.

## BIBLIOGRAFÍA

1. Fajardo Arturo  
Epidemiología descriptiva del cáncer en el niño.  
Publicación independiente, 1ª Edición  
México, D. F. 2000
2. Hebert Schildt  
Utilización de C en la inteligencia artificial.  
McGraw-Hill.  
1989.
3. Rolston D. W.  
Inteligencia artificial y sistemas expertos.  
Mc Graw Hill.  
1992.
4. Aguirre A.  
Enfermedades malignas de la infancia (Su detección temprana).  
Bol Med Hosp Infant Méx  
México, D. F. 1965
5. Singh GP, Yu SM  
US childhood mortality, 1950 through 1993: trends and socioeconomic differentials.  
Am J Public Health  
1996
6. Gut Benchimol, Et. Al.  
Los Sistemas Expertos en la empresa.  
Macrobit Editores.  
1990
7. David Hu  
C/C++ for expert systems.  
Mis Press.  
1989
8. Robert I Levine, ET. AL.  
AI and Expert Systems  
McGraw-Hill.  
1990

9. Salas M. M.  
Neoplasias malignas en niños  
Editorial Interamericana Mc Graw-Hill, 1ª Edición  
México, D. F. 1988
10. Chin-Hon Pui. William M Crist.  
Pediatric solid tumors  
American Cancer Society Textbook of Clinical Oncology  
Atlanta, U. S. 1991
11. Russell Stuart. Norving Peter.  
Inteligencia Artificial. Un enfoque moderno.  
Prentice Hall. 1ª Edición.  
México, 1996
12. Barnat J.  
Enciclopedia médica familiar.  
Ediciones NAUTA.  
Barcelona, España 1977
13. Flores LE., Williams J., Bell BA, O'Brien M, Ragab AH.  
Delay in the diagnosis of pediatric brain tumors.  
AJDC 1986
14. Pratt CB, Smith JW, Woerner S, and et al.  
Factors leading to delay in the diagnosis and effecting survival of children with  
head and neck rhabdomyosarcoma.  
Pediatrics 1978
15. Steuber CP, Nesbit ME.  
Clinical Assessment and differential diagnosis of the child with suspected cancer.  
Principles and practice of pediatric oncology. 3ª edición.  
Philadelphia, US. 1997
16. Martínez-Avalos A.  
El niño con sospecha de cancer. Oncofisagnosis, oncodiagnósticos tempranos en  
medicina de primer nivel.  
Farmacia Famitalia Carlo Erba.  
1994.
17. H. M. Deitel, P. J. Deitel  
Como programar en C/C++.  
Prentice may 2ª edición.  
México, 1994

18. Ignacio Pérez Aguirre

Utilización de procesamiento de lenguaje natural en el diseño de la interfaz de usuario en un sistema experto auxiliar en el diagnóstico temprano de cáncer en niños.

Tesis de licenciatura del Instituto Tecnológico de Minatitlán.  
Minatitlán, Ver., 2002.