

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD #6



**IMPLEMENTACIÓN DE UN SISTEMA PARA EL CÁLCULO DE
RIESGO DE ENFERMEDADES GENÉTICAS.**

**TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE
INGENIERO EN CIENCIAS INFORMÁTICAS**

AUTORES: Daniel Ernesto Revilla Mugarra.
Enrique Sánchez Fernández.

TUTORES: Ing. Reynaldo Rosado Roselló.
Ing. Reynaldo Alvarez Luna.

CIUDAD DE LA HABANA, CUBA

Junio 2010-06-02

AÑO 52 DE LA REVOLUCIÓN

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Enrique Sánchez Fernández (Autor)

Daniel Ernesto Revilla Mugarra (Autor)

Ing. Reynaldo Rosado Rosselló. (Tutor)

Ing. Reynaldo Alvarez Luna. (Tutor)

DATOS DE CONTACTO

Autores:

Enrique Sánchez Fernández

email: esfernandez@estudiantes.uci.cu

Daniel Ernesto Revilla Mugarra

email: derevilla@estudiantes.uci.cu

Tutores:

Reynaldo Rosado Rosselló

email: rrosado@uci.cu

Reynaldo Alvarez Luna

email: rluna@uci.cu

AGRADECIMIENTOS

Agradecemos a nuestra familia que nos ha sabido guiar durante todo este tiempo para alcanzar un sueño que hoy se hace realidad.

A nuestros amigos que nos han prestado apoyo incondicional en los momentos más difíciles. En especial a Dayron, Tonyce, Bolaños (Bola), Luis, Felix, Andrés, Jorge (Russo) y Nela, en fin a todos aquellos que de alguna manera u otra han hecho su aporte en este trabajo que hoy logramos.

A nuestros profesores en especial a Reynaldo Luna y Reynaldo Rosado, los cuales han sido nuestros tutores a lo largo de todo este proceso. Les agradecemos por su paciencia, dedicación y optimismo.

A Tonyce de la Rosa el cual siempre ha estado ahí como un tutor más.

A la Revolución y en especial al Comandante en Jefe Fidel Castro, por construirnos una universidad tan maravillosa.

A todos ustedes muchas gracias

DEDICATORIA

De: Enrique Sánchez Fernández

Especialmente a mi madre y a mi tío José Antonio, los cuales me han apoyado y guiado en todo momento, no tendría palabras, ni existiría espacio no tiempo para expresarles todo el amor que siento por ellos.

A mis abuelos Antonio y Teresa, que tanto me quieren.

A mis hermanas Gretel y Yusel, las cuales quiero tanto, que a pesar de ser el más pequeño de los tres, las protegería con mi vida.

Lo mismo para mis primos José, Graciela y Yanet, más que primos para mí son como hermanos.

A mi otra madre Martica, a la cual quiero tanto que siempre se me hizo imposible llamarte tía, pues para mí representa mucho más que eso.

A mi familia en general.

A mis amigos.

De: Daniel Ernesto Revilla Mugarra

Dedico mi tesis a lo más grande que tengo en la vida, mi familia.

En especial a mi mama por estar siempre ahí cuando mas falta me hizo, por querernos tanto a mí hermana y a mi. A mi papa mi ejemplo a seguir, confidente, amigo la persona que sin su apoyo no hubiera podido llegar tan lejos en la vida.

A mi abuela la persona a la que le debo todo lo que ahora estoy logrando, por escucharme y encaminarme por el buen camino.

A mis hermanas las cuales adoro, y a pesar de la distancia siempre estamos juntos.

A mi novia la cual me apoyado tanto, por ser tan buena, por quererme, por haberme hecho reír, por siempre creer en mí, por seguir conmigo, por ser tan comprensiva y esperarme todo este tiempo.

A mi otra madre Zuzel quien siempre me ha brindado su ayuda y amor incondicional.

A mis amigos los cuales de una forma u otra forman parte de esto, a los cuales en algún momento han aportado un poco de alegría a mi vida.

En general le dedico esta tesis a todo aquel que nos ayudo en esta travesía y estuvo junto a mí para ver terminado este trabajo.

RESUMEN

La identificación del riesgo genético, así como su posibilidad de prevención y tratamiento, son algunos de los problemas fundamentales que presentan las enfermedades de origen genético en Cuba. Por estos motivos se desarrolló una versión del sistema alasARBOGEN, la cual posee entre sus funcionalidades, la representación gráfica del árbol genealógico de un individuo y su familia, así como la gestión de la información genealógica de las personas que figuran en el genograma.

El presente trabajo desempeña un papel importante en el desarrollo de la segunda versión de alasARBOGEN, pues permite incluirle funcionalidades fundamentales como es el caso del cálculo de riesgo de enfermedades genéticas como el cáncer de mama, ovario y próstata, a través de un sistema basado en reglas, el cual usa la información proveniente de los árboles genealógicos unido a los ficheros que se desarrollan en un editor diseñado para que los propios genetistas puedan editar sus reglas siempre y cuando cumplan con las normas establecidas.

PALABRAS CLAVE

Enfermedades genéticas, factores de riesgo, motor de reglas, sistemas expertos.

TABLA DE CONTENIDOS

AGRADECIMIENTOS	I
DEDICATORIA	II
RESUMEN	IV
INTRODUCCIÓN	1
CAPÍTULO 1: RESEÑA BIBLIOGRÁFICA	5
1.1 Introducción al cálculo de riesgo de enfermedades genéticas	5
1.1.1 Factores o variables de riesgo.....	8
1.1.2 Factores de riesgo del Cáncer de Mama	8
1.1.3 Factores de riesgo del Cáncer de Ovario.....	10
1.1.4 Factores de riesgo del Cáncer de Próstata.....	11
1.2 Sistemas Expertos	12
1.2.1 Componentes de un sistema experto.....	13
1.3 Primeros sistemas expertos y sus aplicaciones	19
1.4 Programas vinculados al cálculo de riesgo del Cáncer de origen genético	19
CAPÍTULO 2: MÉTODOS Y MATERIALES	22
2.1 Identificación del problema a resolver	22
2.2 Adquisición del conocimiento	23
2.3 Representación del conocimiento	23
2.4 Encadenamiento hacia atrás	25
2.5 Encadenamiento hacia adelante	25
2.6 Modelos matemáticos para el cálculo de riesgo	26
2.7. Tecnologías y herramientas para el desarrollo del sistema	27
2.7.1 Plataforma de Desarrollo y Lenguaje de Programación	28
2.7.2 Entorno de Desarrollo	28

2.7.3 Herramienta CASE (Computer Aided Software Engineering)	29
2.7.4 Lenguaje de Modelado	29
2.7.5 Drools como Motor de Reglas	30
CAPÍTULO 3: RESULTADOS Y DISCUSIÓN	32
3.1 Modelo conceptual del sistema	32
3.2 Diagrama de clases y Patrones utilizados.....	33
3.2.1 Diagrama de clases	33
3.2.2 Patrones utilizados.....	35
3.3 Descripción de la solución.....	37
3.3.1 Principales aspectos del módulo Cálculo de Riesgo.....	37
3.3.2 Principales aspectos del módulo Editor de Reglas.....	38
3.4 Interfaces del sistema	40
CONCLUSIONES.....	48
RECOMENDACIONES.....	49
TRABAJOS CITADOS.....	50
BIBLIOGRAFÍA	53
ANEXOS	54
GLOSARIO	57

INTRODUCCIÓN

La genética médica surgida en 1900 con el descubrimiento del mendelismo, alcanza un papel reconocido como disciplina central, la cual se ocupa de la variabilidad humana y al mismo tiempo desarrolla métodos que permiten nuevos enfoques a muchas enfermedades y promete brindar mucho más en un futuro inmediato.

Con más de 20 años de experiencia el Programa Nacional de Diagnóstico, Manejo y Prevención de Enfermedades Genéticas y Defectos Congénitos en Cuba, puso en marcha desde sus inicios, a instancias de la máxima dirección del gobierno de salud cubano y a través de la infraestructura del sistema nacional de salud, programas asistenciales de acceso universal y gratuito.

Con el objetivo de llevar un mejor control y asesoramiento de los problemas genéticos, ampliar las investigaciones y extender el cuidado y tratamiento al sistema primario de salud, Fidel inaugura el 5 de agosto del 2003 el Centro Nacional de Genética Médica (CNGM). Con el objetivo fundamental de localizar a todas las familias con riesgo genético de transmitir estas enfermedades a la descendencia y ofrecer el diagnóstico citogenético, molecular, metabólico prenatal y postnatal a aquellos que lo requieran. El mismo actúa como centro de referencia nacional, tiene integrada una red nacional de genética médica constituida por 184 centros, ubicados en todos los municipios y provincias del país, contribuyendo al mejoramiento de los servicios genéticos.

Con el inicio de los estudios genéticos de estas familias surgen diversas problemáticas como la de poseer una herramienta capaz de representar el árbol genealógico de un individuo. Como solución a esta problemática la Universidad de las Ciencias Informáticas (UCI) en conjunto con el CNGM deciden crear alasARBOGEN, el cual es un sistema informático que permite la representación de los árboles genealógicos y la gestión de la información de las personas que figuran en el genograma.

En el mundo existe una diversidad de herramientas que permiten la representación gráfica de árboles genealógicos como es el caso de Cyrillic, un software propietario que en su última versión 3.0 incluye entre sus funcionalidades el cálculo de riesgo para los casos genéticos de cáncer. Sin embargo dicho software no cumple con las demandas de los especialistas, ya que para su utilización es necesario el pago

de una licencia por cada copia que vaya a ser instalada en las PC de los centros donde se realizan los estudios.

Cuba dispone de 832 másteres en asesoramiento genético, es decir, uno por cada 13 228 habitantes aproximadamente, y 107 especialistas en genética clínica (según informe del MINSAP de 2009). Los primeros son los encargados de identificar el riesgo genético en la comunidad, ofrecer información relativa a estas enfermedades, sus posibilidades de prevención y tratamiento para aminorar el impacto, mientras que los segundos son galenos especializados en este tipo de diagnóstico. Esto sin dudas es una limitante, pues a pesar del gran apoyo y la prioridad que el Gobierno concede a la realización de estudios en este campo, estos especialistas no estarán disponibles en todo momento a lo largo del tiempo. Es por ello que surge la necesidad de encontrar la manera de realizar dichos estudios en todos los centros de asistencia primaria del país, ampliando así la posibilidad de una detección temprana de estas enfermedades y una utilización eficiente del personal especializado con que se cuenta. (1)

Actualmente en el país no existe una herramienta informática capaz de calcular el riesgo de enfermedades genéticas y constituye un gasto considerable adquirir este software en el mercado internacional. Por lo que se hace necesario implementar un sistema para el cálculo de riesgo de enfermedades de origen genético.

Teniendo en cuenta todos los argumentos analizados anteriormente se plantea como **Problema Científico**: ¿Cómo contribuir al cálculo de riesgos de padecimiento de cáncer de origen genético a través de la información procedente de los árboles genealógicos?

En función del problema científico identificado se determina como **Objeto de Estudio** en esta investigación, los sistemas para el cálculo de riesgo de padecimiento de cáncer de origen genético, y como **Campo de Acción** la implementación de sistemas basados en reglas para el cálculo de riesgo de padecimiento de cáncer de origen genético a partir de la información de los árboles genealógicos.

Por lo que se adopta como **Objetivo General**:

Desarrollar un sistema basado en reglas que permita calcular el riesgo de padecimiento de cáncer de origen genético a partir de la información gestionada por alasARBOGEN 2.0.

Para dar cumplimiento al objetivo general se definieron las siguientes tareas de la investigación:

- Estudio de las tendencias actuales del cálculo de riesgos de enfermedades genéticas.
- Identificación de sistemas que se utilizan para el cálculo de riesgos de enfermedades genéticas.
- Implementación de un editor de reglas propio para el sistema alasARBOGEN.
- Implementación del sistema basado en reglas para el cálculo de riesgo de padecimiento de cáncer de origen genético.
- Ejecución de pruebas funcionales al sistema.
- Integración de la solución al sistema alasARBOGEN 2.0.

El presente trabajo de diploma cuenta con tres Capítulos:

Capítulo # 1: Reseña Bibliográfica:

En el presente capítulo se hace una introducción al problema del cálculo de riesgo de enfermedades de origen genético, además brinda una descripción de algunos aspectos relacionados con los sistemas de expertos, así como: aplicaciones, tipos de sistemas de expertos en general y en específico algunos sistemas expertos en diagnóstico médico.

Capítulo # 2: Métodos y materiales:

En el presente capítulo se pone de manifiesto la identificación del problema a resolver, así como la adquisición del conocimiento y las fuentes de donde fue extraído el mismo. Brinda un ejemplo de la forma que se representa el conocimiento y muestra el mecanismo de búsqueda que se utiliza. Se presentan algunos modelos matemáticos que hay que tener en cuenta a la hora de definir alguno propio para las características de la población cubana, modelos que serian empleados en los cálculos de riesgo del cáncer genético de mama, ovario y próstata. Se explican las características favorables y desfavorables de los mismos, mostrando además las distintas herramientas utilizadas para la implementación y prueba del sistema desarrollado.

Capítulo # 3: Resultados y Discusión:

En este capítulo se presenta un modelo conceptual para lograr una mayor claridad del sistema implementado. Se presenta el diagrama de clases del diseño del sistema y los patrones de diseño empleados en la programación orientada a objetos, así como una descripción de la solución al problema planteado.

Capítulo 1: *Reseña Bibliográfica*

CAPÍTULO 1: RESEÑA BIBLIOGRÁFICA

Introducción

En el presente capítulo se hace una introducción al problema del cálculo de riesgo de enfermedades de origen genético, además brinda una descripción de algunos aspectos relacionados con los sistemas de expertos, así como: aplicaciones, tipos de sistemas de expertos en general y en específico algunos sistemas expertos en diagnóstico médico.

Antes de realizar una introducción en el cálculo de riesgo es necesario definir el concepto de Enfermedad Genética, la misma es:

Una condición patológica establecida por el efecto biológico consecuente a una alteración del genoma, es decir, son aquellas que se transmiten en los genes, de padres a hijos. Las enfermedades genéticas corresponden a un grupo heterogéneo de afecciones que en su etiología presentan un significativo componente genético. Ello puede ser alguna alteración en un solo gen (monogénicas), en varios genes (poligenes) o en muchos genes (cromosomas). La alteración genética puede producir directamente la enfermedad (por ejemplo, el caso de la Hemofilia) o interactuar con factores ambientales (como por ejemplo, la predisposición genética en la etiología del cáncer y la Hipertensión arterial). (2)

1.1 Introducción al cálculo de riesgo de enfermedades genéticas

La estimación del riesgo o cálculo de riesgo de padecimiento de enfermedades de origen genético es competencia del médico especialista y se basa en la evaluación de la historia oncológica personal y familiar de cada paciente, donde se establecen estrategias de seguimiento y se implementan medidas de prevención, disminuyendo así la mortalidad por causa de la herencia de dichas enfermedades.

Prevención representa reducción de riesgo. Desafortunadamente, no todo el mundo comprende que las medidas preventivas frecuentemente no eliminan completamente un riesgo, sólo lo reducen. Si bien la magnitud de la reducción de los riesgos es una consideración importante, no deja de ser cierto que no es aceptable cualquier intervención preventiva, debe haber un balance favorable entre riesgos, molestias, sufrimientos, secuelas y gastos, por una parte, y beneficios, por otra. Esta consideración es

Capítulo 1: *Reseña Bibliográfica*

particularmente importante si se considera que las estrategias de prevención deben aplicarse por tiempo prolongado (años, décadas) a personas sanas.

La amplia difusión de la genética y el avance en sus diferentes metodologías, permiten actualmente a las parejas de alto riesgo, prevenir el nacimiento de hijos con enfermedades de origen genético, especialmente si hablamos de algunas tales como el Cáncer de Mama, el Cáncer de Ovarios y Cáncer de Próstata.

El Cáncer de Mama, es la segunda causa de muerte por cáncer en la mujer a nivel mundial (3). En nuestro país, el cáncer de mama ocupa el segundo lugar dentro de la mortalidad general como ocurre en los países desarrollados (4), lo que es importante identificar al grupo de mujeres que está en mayor riesgo de enfrentar esta enfermedad. Se revisan los factores de riesgo genético y no genético que predisponen al desarrollo del cáncer de mama. Se analizan los modelos y estrategias más utilizados para hacer objetivos estos riesgos, entre los cuales se encuentran modelos matemáticos de predicción de riesgo de cáncer de mama, modelos de predicción de probabilidad de mutación genética y técnicas de obtención de células epiteliales de los conductos de la mama.

Hasta hace poco tiempo atrás, los esfuerzos médicos se concentraban en el diagnóstico precoz y tratamiento oportuno del cáncer de mama. La incorporación de programas de pesquisa precoz o “screening” o “tamizaje” por mamografía (5) ha permitido diagnosticar tumores cada vez más pequeños, disminuyendo en aproximadamente 30% de la mortalidad por cáncer de mama en mujeres mayores de 50 años. En la actualidad se han identificado diferentes factores de riesgo que están presentes en mujeres sanas que las hacen aumentar en distinto grado su probabilidad de desarrollar un cáncer mamario en el futuro. La ventaja de conocer estos factores, es que permite distinguir diferentes grupos de mujeres de mayor riesgo de cáncer mamario, lo que contribuye a elaborar estrategias de vigilancia en el tiempo de estas mujeres, como también permite establecer estrategias para disminuir el riesgo aumentado y de esta forma, prevenir el desarrollo de un cáncer de mama. El manejo clínico racional de mujeres de alto riesgo constituye un desafío que deberá incorporarse en todos los centros de diagnóstico y tratamiento de cáncer de mama.

Capítulo 1: *Reseña Bibliográfica*

El conocimiento del cáncer de ovario, denominado “el asesino silencioso” por no presentar síntomas o signos específicos particularmente en etapa inicial, es bastante antiguo. Estudios anatómicos y radiológicos en momias de la IV dinastía de los reyes egipcios (2900-2700 a.n.e) evidenciaron lesiones carcinomatosas. Luego en el papiro de Edwin Smith que data de la dinastía egipcia (1600 a.n.e) se establecen aspectos relacionados al cáncer y se describen diversas formas de tratamiento quirúrgico. Hipócrates (460-370 a.n.e) fue quien describió por primera vez el cáncer, utilizando los términos carcinos y carcinoma. (6)

Desde hace más de tres décadas, el cáncer constituye la segunda causa de muerte más frecuente en Cuba y la primera en personas de 15 a 49 años. La quinta parte de las defunciones registradas entre 1990 y 2009 en el país fueron por cáncer. En ese período, la tasa estandarizada de mortalidad por cáncer se elevó de 112,8 por 100000 habitantes a 116,3 por 100000 habitantes.

En Cuba, en el 2003, se reportaron 383 casos para una tasa cruda de $8,2 \times 10^5$, que aumenta hasta 19,4 en las mayores de 60 años. En el 2009, 219 mujeres murieron por cáncer de ovario; la tasa de mortalidad aumenta cada 5 años a partir de los sesenta años de edad y declina a partir de los 80. Ocupa el octavo lugar entre los cánceres que afectan a la mujer y es la décima causa de muerte entre ellas.

El cáncer de próstata es un problema trascendente en la salud de la población masculina. En nuestro país constituye la segunda causa de muerte por enfermedad neoplásica, ubicándose solo detrás del cáncer de estómago. Las tasas de defunción de estas enfermedades han crecido sostenidamente hasta alcanzar la cifra de 1.537 muertes en 2002, con una tasa cruda de 20,5 por 100.000 hombres. En la provincia de Guantánamo 48 personas de cada cien mil habitantes portan un adenocarcinoma de próstata, tercera causa de muerte por tumores malignos en el extremo oriente cubano, de acuerdo con estadísticas del Ministerio de Salud Pública. En el caso de este cáncer, existe la predisposición genética, de familiares que lo hayan padecido, hasta las conductas de riesgo como los malos hábitos nutricionales y la obesidad. Mundialmente está comprobado que mientras más temprano se diagnostiquen esos tumores oncogénicos, hay mayor posibilidad de curación. Cuba cuenta con productos de la biotecnología como vacunas terapéuticas para los pacientes declarados crónicos, que han recibido un grupo de tratamientos, y se aplican para alargar y mejorar la calidad de vida. Según el comportamiento de esta enfermedad porcentual en edades, con el envejecimiento pronosticado de la población para los próximos cinco años, se considera

Capítulo 1: *Reseña Bibliográfica*

que esta enfermedad será cada vez de mayor importancia, por el número de posibles pacientes a padecerla. (7)

1.1.1 Factores o variables de riesgo

Un factor de riesgo es aquel que cambia las probabilidades de que padezca una enfermedad como lo es el cáncer. Los distintos tipos de cáncer conllevan diferentes factores de riesgo. Por ejemplo, exponerse sin protección a la luz solar intensa es un factor de riesgo para el cáncer de la piel, mientras el fumar es un factor de riesgo para un número de cánceres. Pero los factores de riesgo no lo indican todo. Presentar uno o incluso varios factores de riesgo no significa que usted tendrá la enfermedad. Además, muchas personas que adquieren la enfermedad pueden no tener factores de riesgo conocidos.

A continuación se representarán una serie de factores, escogiendo como dominio de este trabajo el cáncer de mama, ovario y próstata como muestra representativa que pueden ser adquiridos genéticamente, presentando variables que son comunes entre ellos, mientras que otras no.

1.1.2 Factores de riesgo del Cáncer de Mama

Diversos estudios han demostrado que hay variables que aumentan el riesgo de desarrollar un cáncer de mama a las mujeres que los poseen. Entre estas variables se encuentran factores epidemiológicos, ginecobstétricos, genéticos, dietéticos y ambientales, entre otros. Es importante identificarlos y objetivarlos con la idea de poder implementar un plan de seguimiento y prevención. El hecho de tener alguno de estos factores de riesgo no significa que usted contraerá la enfermedad. Casi todas las mujeres tienen algunos factores de riesgo y en su mayoría no contraen cáncer de mama.

Identificando los factores o variables de riesgo:

Epidemiológicos y no genéticos

- **Edad:** La edad es el factor aislado más importante en el riesgo de cáncer de mama. La incidencia del cáncer de mama es distinta en los distintos grupos etarios, siendo muy baja en mujeres bajo los 25 años (8).

Capítulo 1: *Reseña Bibliográfica*

- **Raza:** Las mujeres blancas son más propensas a padecer esta enfermedad que las de raza negra, aunque la mortalidad en éstas últimas es mayor, probablemente porque a ellas se les detecta en estadíos más avanzados. Las que tienen menor riesgo de padecerlo son las mujeres asiáticas e hispanas (9).

Antecedentes ginecobstétricos

- **Edad de menarquía:** Se ha encontrado una relación entre la edad del primer ciclo menstrual y el riesgo de cáncer de mama, a menor edad, mayor es el riesgo.
- **Embarazo y edad al primer embarazo de término:** Las mujeres nulíparas o quienes presentan su primer embarazo de término después de los 35 años, están en mayor riesgo de presentar un cáncer de mama. Este riesgo se hace más evidente en las mayores de 45 años.
- **Lactancia:** Las mujeres que no han dado pecho tienen más probabilidades de sufrir esta enfermedad que aquellas que lo han dado (10).

Factores dietéticos

- **Dieta grasa:** Se ha postulado durante mucho tiempo que la dieta rica en grasas cumple un rol en el aumento del riesgo de cáncer de mama (11).
- **Alcohol:** El consumo de alcohol ha sido relacionado con un aumento del riesgo de cáncer de mama (12).

Familiares y Genéticos

Entre los factores de riesgo más significativos están los antecedentes familiares de cáncer de mama. Aproximadamente 20 a 30% de las mujeres con diagnóstico de cáncer de mama tiene al menos un familiar con la enfermedad, sin embargo sólo un 5 a 10% puede ser atribuible a mutación de los genes BRCA1 y BRCA2.

Capítulo 1: *Reseña Bibliográfica*

En 1990 se identificó un gen en el cromosoma 17q21, que actualmente se llama BRCA1 y se ha relacionado con el cáncer de mama y de ovario. El gen BRCA2 fue detectado en 1994 en el cromosoma 13q12-13. Otros autores han publicado el riesgo a lo largo de toda la vida de tener un cáncer de mama para las mujeres portadoras de mutaciones en los genes BRCA1/2 con antecedentes de familiares con cáncer de mama u ovario (13).

Factores ambientales

- **Tabaquismo:** No todos los estudios concuerdan respecto a si el tabaco aumenta o no el riesgo de cáncer de mama. Así como algunos han publicado un aumento del riesgo, otros han publicado que no influye o que incluso lo disminuye.

1.1.3 Factores de riesgo del Cáncer de Ovario

Aún cuando una persona con cáncer de ovario tiene un factor de riesgo, a menudo es muy difícil saber cuánto pudo haber contribuido ese factor de riesgo al cáncer. Los investigadores han descubierto varios factores específicos que cambian la probabilidad de una mujer de padecer de cáncer ovárico epitelial. Estos factores de riesgo no se aplican a otros tipos de cáncer ovárico menos comunes, tales como los tumores de células germinales y los tumores estromales (14).

Identificando los factores o variables de riesgo:

- **Edad:** El riesgo de padecer de cáncer de ovario aumenta con la edad. Los cánceres de ovario son poco comunes en las mujeres menores de 40 años, y la mayoría se originan después de la menopausia. La mitad de todos los cánceres de ovario se encuentran en mujeres de más de 63 años.
- **Obesidad:** Varios estudios han analizado la relación entre la obesidad y el cáncer de ovario. En general, parece que las mujeres obesas (aquellas con un índice de masa corporal de al menos 30) tienen un riesgo mayor de cáncer de ovario.

Capítulo 1: *Reseña Bibliográfica*

- **Antecedentes relacionados con la reproducción:** Una mujer que haya tenido hijos tiene un riesgo menor de cáncer de ovario en comparación con las que no han tenido hijos. El riesgo es aún menor con cada embarazo. La lactancia (amamantar al bebé) puede reducir el riesgo aún más. El uso de anticonceptivos orales (píldora anticonceptiva) también reduce el riesgo de cáncer de ovario.
- **Antecedentes familiares de cáncer de ovario, del seno o cáncer colorrectal:** El cáncer de ovario puede ser más común en algunas familias. El riesgo de cáncer de ovario aumenta si su madre, hermana o hija tiene, o tuvo, cáncer ovárico. El riesgo también aumenta mientras más familiares tengan cáncer de ovario.
- **Antecedentes personales de cáncer del seno:** Si ha tenido cáncer del seno, puede que también tenga un riesgo mayor de cáncer de ovario.

1.1.4 Factores de riesgo del Cáncer de Próstata

El cáncer de próstata es un grupo de células cancerosas (un tumor maligno) que comienza muy frecuentemente en la parte exterior de la próstata. Todavía no se conocen exactamente las causas del cáncer de próstata pero se han encontrado factores de riesgo que pueden aumentar las probabilidades de contraerlo.

Tales factores de riesgo incluyen la edad, la raza, la nacionalidad, el medio ambiente, el estilo de vida, la dieta diaria, la inactividad física, la obesidad, la historia familiar, la vasectomía, el lugar de trabajo, la actividad sexual y las ETS, el tabaco y la existencia de otros cánceres sincrónicos o metacrónicos (15).

Identificando los factores o variables de riesgo:

- **Edad:** La probabilidad de tener cáncer de próstata aumenta rápidamente después de los 50 años.
- **Raza:** El cáncer de próstata ocurre alrededor del 70% más a menudo en hombres afroamericanos que en hombres blancos americanos. Los hombres de ascendencia asiática o los habitantes de las islas del Pacífico, tienen las tasas más bajas de incidencia y mortalidad.

Capítulo 1: *Reseña Bibliográfica*

- **Dieta:** La nutrición parece tener un papel importante en el desarrollo del cáncer de próstata, como lo demuestra la distribución geográfica de la incidencia de este cáncer, siendo altamente elevada en los países industrializados o "desarrollados".
- **Inactividad física y obesidad:** La obesidad definida como un índice de masa corporal superior a 29kg/m² se asocia con el doble de riesgo de padecer cáncer de próstata que en los hombres que tienen un peso normal.
- **Historia familiar:** El cáncer de próstata parece que ocurre más frecuentemente en algunas familias, sugiriendo que existe un factor hereditario.
- **Trabajo:** Los trabajadores de las industrias del caucho y del cadmio en soldaduras y baterías, parecen tener más probabilidades de desarrollar cáncer de próstata.

En el mundo existe una variedad de programas que tienen como principal objetivo el cálculo de riesgo de enfermedades de origen genético como es el caso de Constantino, Cáncer-Gene, Cyrilic etc. Estas herramientas utilizan técnicas de inteligencia artificial para su implementación.

1.2 Sistemas Expertos

Con el desarrollo de la inteligencia artificial (IA), se ha logrado modelar la inteligencia humana en sistemas computacionales. Entre sus múltiples aplicaciones se encuentra, el área de control de sistemas, planificación automática, la habilidad de responder a diagnósticos y consultas de los consumidores, reconocimiento del habla y reconocimiento de patrones.

La IA en la medicina ha tenido grandes avances, pues ayuda a los médicos a hacer diagnósticos, supervisar la condición de los pacientes, administrar tratamientos y preparar estudios estadísticos. En Cuba, se han experimentado avances en este sentido: historias clínicas electrónicas con insospechadas posibilidades en el futuro, sistemas para tratamientos estadísticos que son capaces de ofrecer información gerencial para la toma de decisiones, procedimientos de aprendizaje, agentes inteligentes para el diagnóstico de trastornos ginecológicos; así como otras interesantes utilidades.

Capítulo 1: Reseña Bibliográfica

Los Sistemas Expertos (SE) son una rama de la IA que hace un amplio uso del conocimiento especializado para resolver problemas como un especialista humano. Es decir el especialista tiene conocimientos o habilidades especiales que la mayoría no conoce o de las que no dispone; puede resolver problemas que la mayoría no podría resolver, o los resuelve con mayor eficiencia. Un SE es un Sistema Basado en Conocimiento (SBC) al que se le incorpora conocimiento proveniente de expertos en dicho dominio. La finalidad de un SE es la resolución de problemas del dominio para el que ha sido creado, aplicando técnicas de razonamiento sobre el conocimiento que alberga su base de conocimiento.

1.2.1 Componentes de un sistema experto

A continuación se ilustra el concepto básico de un sistema experto basado en conocimiento. El usuario aporta los hechos u información al sistema experto y recibe consejo o experiencia como respuesta. En su interior, el sistema experto incluye dos componentes principales la base de conocimiento contiene el conocimiento que le permite al mecanismo de inferencia sacar conclusiones; éstas son las respuestas del sistema experto a la consulta especializada del usuario.

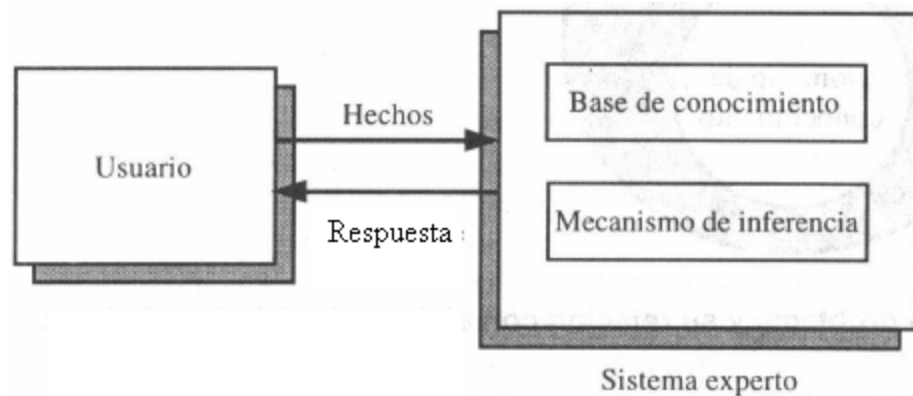


Figura 1.1 Modelo Conceptual.

La base de conocimientos

La representación del conocimiento ha sido ampliamente estudiada por la Inteligencia Artificial (IA). Su principal objetivo es la creación de formalizaciones del conocimiento que sean procesables por programas

Capítulo 1: *Reseña Bibliográfica*

informáticos y que se acerquen al modelo de razonamiento humano. Un Sistema Basado en el Conocimiento es aquel en el que aparece representado el conocimiento de un dominio determinado, de tal forma que dicha representación sea procesable por un programa informático.

La base de conocimientos es la parte más importante de un sistema experto. Contiene el conocimiento y las heurísticas del dominio en el cual el programa es competente y, en algunos sistemas, el conocimiento de control. La elección del formalismo de representación es fundamental para el rendimiento del sistema. Por ello su elección y diseño requieren de un estudio cuidadoso. La base de conocimientos debe permitir un rápido acceso a los conocimientos, debe ser modular y fácil de desarrollar y mantener. En la misma intervienen dos elementos importantes.

- El conocimiento se almacena en la base de conocimiento y consiste en un conjunto de objetos y un conjunto de reglas que gobierna las relaciones entre esos objetos.
- La información almacenada en la base de conocimiento es permanente y estática, es decir, no cambia de una aplicación a otra, a menos que se incorporen al sistema experto elementos de aprendizaje.

El motor de inferencia

Es el intérprete del conocimiento almacenado en la base de conocimientos. Sus tareas incluyen seleccionar, decidir y aplicar el conocimiento de la base de conocimientos sobre la base de hechos, con el fin de obtener la solución deseada. Sus funciones principales son la inferencia y el control. El control se refiere a la secuencia en que se examinan las primitivas de la base de conocimiento. La inferencia es el razonamiento que realiza a partir del análisis y evaluación de las primitivas de la base de conocimiento, encadenando conclusiones intermedias hasta presentar la solución final. Las soluciones pueden clasificarse en dos tipos: Las soluciones simples son las que resultan de una regla simple; las soluciones compuestas son las que resultan de más de una regla. Para obtener soluciones se utilizan diferentes tipos de reglas y estrategias de inferencia y control.

- **Modus Ponens:** es quizás la regla de inferencia más comúnmente utilizada. Se utiliza para obtener conclusiones simples. En ella, se examina la premisa de la regla, y si es cierta, la

Capítulo 1: *Reseña Bibliográfica*

conclusión pasa a formar parte del conocimiento. Como ilustración, supongamos que tenemos la regla, “Si A es cierto, entonces B es cierto” y que sabemos además que A es cierto. La regla Modus Ponens concluye que B es cierto. Esta regla de inferencia, que parece trivial, debido a su familiaridad, es la base de un gran número de sistemas expertos.

- **Modus Tollens:** se utiliza también para obtener conclusiones simples. En este caso se examina la conclusión y si es falsa, se concluye que la premisa también es falsa. Por ejemplo, supongamos de nuevo que se tiene la regla, “Si A es cierto, entonces B es cierto” pero se sabe que B es falso. Entonces, utilizando la regla Modus Ponens no se puede obtener ninguna conclusión, pero la regla Modus Tollens concluye que A es falso.
- **Encadenamiento hacia delante (forward chaining):** es una de las estrategias de inferencia más utilizadas para obtener conclusiones compuestas. Esta estrategia puede utilizarse cuando las premisas de ciertas reglas coinciden con las conclusiones de otras. Cuando se encadenan las reglas, los hechos pueden utilizarse para dar lugar a nuevos hechos. Esto se repite sucesivamente hasta que no pueden obtenerse más conclusiones. El tiempo que consume este proceso hasta su terminación depende, por una parte, de los hechos conocidos, y, por otra, de las reglas que se activan. Este algoritmo puede ser implementado de muchas formas. Una de ellas comienza con las reglas cuyas premisas tienen valores conocidos. Estas reglas deben concluir y sus conclusiones dan lugar a nuevos hechos. Estos nuevos hechos se añaden al conjunto de hechos conocidos, y el proceso continúa hasta que no pueden obtenerse nuevos hechos.
- **Encadenamiento hacia atrás (backward chaining):** Se propuso una solución hipótesis del problema, entonces el algoritmo navega hacia atrás a través de las reglas en búsqueda de una conclusión que confirme dicha hipótesis. Si no se obtiene ninguna conclusión con la información existente, entonces el algoritmo fuerza a preguntar al usuario en busca de nueva información sobre los elementos que son relevantes para obtener información sobre la hipótesis objetivo.

La interfaz de usuario

Capítulo 1: *Reseña Bibliográfica*

La interfaz de usuario facilita la comunicación entre el usuario y el sistema, la interacción entre ellos se realiza en lenguaje natural. También es altamente interactiva y sigue el patrón de la conversación de los seres humanos, aceptando datos del usuario, planteando preguntas, dando información y mostrando la conclusión final a la que ha arribado el sistema experto. Para conducir este proceso de manera aceptable por el usuario es especialmente importante el diseño del interfaz de usuario.

1.2.2 Tipos de Sistema Basado en el Conocimiento o Sistemas Expertos

Algunos tipos de Sistema Basado en el Conocimiento o Sistemas Expertos son:

- Sistemas basados en reglas.
- Sistemas basados en casos.
- Sistemas basados en probabilidades.

Sistemas basados en reglas

Un Sistema Basado en Reglas es un sistema basado en el conocimiento en el cual se realiza una representación simbólica declarativa de un dominio mediante reglas de producción o reglas condicionales. Es uno de los modelos de representación del conocimiento más ampliamente utilizados. Esto es debido a que resultan muy apropiados en situaciones en las que el conocimiento que se desea representar surge de forma natural con estructura de reglas (16).

Las reglas son la técnica más popular para representar el conocimiento. Esto puede atribuirse a que ellas tienen un formato parecido al que utilizan los expertos humanos para expresar su conocimiento condición o acción. Adicionalmente, son excelentes para representar el conocimiento heurístico, su implementación es fácil, su entendimiento es rápido y pueden usar medidas de incertidumbre.

La fórmula general de la regla es la siguiente:

SI < Condiciones > ENTONCES < Conclusiones o acciones >

Capítulo 1: *Reseña Bibliográfica*

La primera parte de la regla es una prueba condicional también llamada premisa o antecedente. Si la prueba se satisface a través de un emparejamiento con los hechos conocidos, entonces la segunda parte de la regla también llamada conclusión, acción o consecuente también se satisface.

Es interesante notar que si las reglas sólo incluyen proposiciones es decir son constantes, entonces los hechos emparejarán exactamente con ellas. Esto quiere decir que la interconexión entre las reglas y los hechos es explícita y es conocida antes de la ejecución, por lo que la base de conocimientos puede verse como un árbol de decisión y/o, “y” para las condiciones de cada regla, “o” para distintas reglas. En otros casos, si las reglas incluyen variables, se usa el emparejamiento con patrones para ligarlas con los hechos, consecuentemente las relaciones entre las reglas y los hechos varían en cada corrida del programa y sólo se pueden conocer durante el tiempo de ejecución. El primer tipo de reglas se aplica exitosamente en sistemas de diagnóstico y clasificación. El segundo tipo es útil en situaciones en las cuales es necesario probar muchos datos antes de conseguir la solución, como en el diseño, planeación y síntesis.

Los problemas con los que pueden tratar los sistemas basados en reglas pueden clasificarse en dos tipos: problemas esencialmente deterministas y problemas esencialmente estocásticos. Consecuentemente, los sistemas expertos pueden clasificarse en dos tipos principales según la naturaleza de problemas para los que están diseñados:

- **Deterministas:** Son formulados usando un conjunto de reglas que relacionan varios objetos bien definidos. Se los conoce como Sistemas basados en reglas porque obtienen conclusiones basándose en un conjunto de reglas utilizando un mecanismo de razonamiento lógico. Las reglas deterministas constituyen la más sencilla de las metodologías utilizadas en Sistemas Expertos. La base de conocimiento contiene el conjunto de reglas que definen el problema y el motor de inferencia saca las conclusiones aplicando la lógica clásica a estas reglas.
- **Estocásticos:** Se introducen algunos medios para tratar la incertidumbre en situaciones inciertas. En este aspecto, algunos sistemas expertos usan la estructura de los sistemas basados en reglas pero introducen una medida asociada a la incertidumbre de las reglas y a la de sus premisas. Una medida de la incertidumbre es la probabilidad en la que la distribución conjunta de un conjunto de

Capítulo 1: *Reseña Bibliográfica*

variables se utiliza para describir las relaciones de dependencia entre ellas. A estos sistemas se los conocen como sistemas expertos probabilísticos y la estrategia de razonamiento se conoce como razonamiento probabilístico o inferencia probabilística. Finalmente, la ventaja de las reglas es que son muy modulares, uniformes y naturales, pero tienen la desventaja de que su número crece muy rápido y a la hora de realizar su mantenimiento se debe tener mucho cuidado para no introducir reglas contradictorias o que generen encadenamiento infinito.

Sistemas basados en casos

El Sistema Basado en Casos, no es más que otro paradigma de resolución de problemas, pero es precisamente la diferencia con el resto de los acercamientos de la inteligencia artificial las que la hacen tan especial. En lugar de confiar únicamente en el conocimiento general del dominio del problema, o realizar asociaciones a lo largo de relaciones entre descripciones del problema y conclusiones, este paradigma es capaz de utilizar conocimiento específico de experiencias previas, es decir, situaciones de un problema concreto (casos). Ante el planteamiento de un problema no abordado con anterioridad, se intenta localizar un caso pasado similar y adaptar su solución a la situación del problema nuevo. De esta adaptación podemos obtener una nueva experiencia a la hora de resolver problemas con ciertas similitudes, lo que nos lleva a una segunda diferenciación del SBC con respecto al resto de tendencias, el aprendizaje incremental, ya que las nuevas adaptaciones se almacenan como nuevos casos, relacionados, y disponibles para comparaciones futuras.

Sistemas basados en probabilidades

En los comienzos de los sistemas expertos de tipo probabilístico surgieron varios obstáculos, debido a las dificultades encontradas para definir la distribución de probabilidad conjunta de las variables. Ello ha ralentizado su desarrollo. Con la introducción de los modelos de redes probabilísticas, estos obstáculos se han superado y los sistemas expertos probabilísticos han vuelto de forma espectacular durante las dos últimas décadas. Estos modelos, que incluyen las redes de Markov y las Bayesianas, se basan en una representación gráfica de las relaciones entre las variables. Esta representación conduce no solo a formas más eficientes de definir la distribución conjunta de probabilidad sino también a una propagación de incertidumbre muy eficiente, que permite sacar conclusiones.

Capítulo 1: *Reseña Bibliográfica*

1.3 Primeros sistemas expertos y sus aplicaciones

Los sistemas expertos proceden inicialmente de la inteligencia artificial a mediados de los años sesenta del siglo XX. En ese período se creía que bastaban unas pocas leyes de razonamiento junto con potentes ordenadores para producir resultados brillantes. Un intento en ese sentido fue el llevado a cabo por los investigadores Alan Newell y Herbert Simon que desarrollaron un programa denominado GPS (17) (General Problem Solver; solucionador general de problemas). Podía trabajar con criptoaritmética, con las torres de Hanoi y con otros problemas similares. Lo que no podía hacer el GPS era resolver problemas del mundo real, tales como un diagnóstico médico.

A partir de 1967 se creó lo que consideraron el primer sistema de experto, el DENTRAL que se utilizaba para identificar estructuras químicas moleculares a partir de su análisis espectrográfico. Posteriormente se desarrollaron otros sistemas como el MYCIN (para consulta y diagnóstico de infecciones de la sangre) dando conclusiones análogas a las que un ser humano daría tras largos años de experiencia, EMYCIN (MYCIN Esencial) para estudiar la función pulmonar, GUIDON (para elegir tratamientos terapéuticos) (18), HERSAY (que intentaba identificar la palabra hablada), PROSPECTOR (utilizado para hallar yacimientos de minerales).

A partir de 1980 se ponen de moda los sistemas expertos, numerosas empresas de alta tecnología investigan en este área de la inteligencia artificial, desarrollando sistemas expertos para su comercialización. Se llega a la conclusión de que el éxito de un sistema experto depende casi exclusivamente de la calidad de su base de conocimiento. El inconveniente es que codificar la pericia de un experto humano puede resultar difícil, largo y laborioso.

1.4 Programas vinculados al cálculo de riesgo del Cáncer de origen genético

Cyrillic

Cyrillic posee todas las características necesarias para dibujar líneas genéticas. Es el más completo, funcional y fácil de usar software diseñado para el manejo de datos genéticos en el sistema operativo Windows. Cada característica de Cyrillic está diseñada para cubrir todas las necesidades especiales del usuario. Todas las herramientas están dispuestas de forma que usted pueda, rápida y fácilmente, manejar

Capítulo 1: Reseña Bibliográfica

los datos de los que disponga. Una de las herramientas más útiles de Cyrillic para aquellos que trabajan con pacientes que están afectados por enfermedades familiares, es la llamada "*disease locus type*". Con ella usted puede calcular riesgos para un individuo dentro una familia en la que exista una enfermedad genética de cáncer. Como ejemplo, el programa incluye el conjunto de datos Claus, que le permite calcular los riesgos familiares para cáncer de mama (19).

BRCAPro

Programa de computación que emplea las estadísticas para predecir si una persona va a tener una mutación (cambio) hereditaria en los genes BRCA1 y BRCA2. Las personas que tienen ciertas mutaciones en estos genes tienen un riesgo más alto que el normal de cáncer de mama, de ovario, de próstata y de otros tipos de cáncer. El programa se basa en los antecedentes clínicos personales y familiares del paciente de cáncer de mama y de ovario, siendo un sistema basado en el conocimiento. No es sensible para mujeres menores de 35 años, con historia familiar sugerente de mutaciones genéticas, historia previa de cáncer de mama o para mujeres que no están en un programa de screening. No es preciso para calcular el riesgo individual de una mujer. La precisión individual ha sido estimada en 0.58 a cinco años. Sólo 44% de los cánceres que se diagnostican en estos programas tienen estimado mayor a 1.67% a cinco años (20).

CAGene ó Cáncer-Gene

Para el cálculo de la probabilidad de presentar mutación *BRCA1* y *BRCA2*, se utilizó un programa computarizado desarrollado por David Eulus en la Universidad de Texas, llamado *CAGene*, que incluye los modelos de Couch, Myriad y *BRCAPRO*. Fue diseñado para construir el árbol genealógico de las mujeres. Calcula los riesgos y probabilidades en función de datos ingresados en la construcción del pedigrí. Tiene como dificultad que calculan el riesgo en base a datos poblacionales, existe una variación individual que no se puede calcular. No usan todas las variables que se relacionan con el aumento de riesgo (21).

En Cuba hasta la actualidad no existe ningún software que permita realizar el cálculo de riesgo de tener una enfermedad de origen genético. Este tipo de herramienta es muy costosa en el mundo entero y para

Capítulo 1: *Reseña Bibliográfica*

su instalación se debe pagar una licencia por cada computadora donde se instale el programa, no contando el país con los recursos monetarios para hacerlo, sin embargo la Universidad de Ciencias Informáticas (UCI) cuenta con los recursos necesarios para su fabricación y despliegue por todo el ámbito nacional, incluyéndolo en la segunda versión del software alasARBOGEN el cual le permite obtener los factores de riesgo de cada individuo que se encuentre representado en un árbol genealógico para de esta forma realizar el cálculo de riesgo y brindar la posibilidad de prevenir males mayores. Esta herramienta ayuda al personal con poca experiencia en la solución de problemas que requieren un conocimiento de experto.

Conclusiones

En este capítulo se realizó una reseña bibliográfica sobre el cálculo de riesgo de padecimiento de enfermedades de origen genético como el cáncer de mama, ovario y próstata, explicando en qué consisten, el comportamiento de su tasa de padecimiento y algunos tratamientos, técnicas y métodos más usados en su diagnóstico (Sistemas Expertos). Se realizó una valoración sobre los programas utilizados para realizar los cálculos de riesgo a nivel mundial, mencionando las dificultades en su aplicación, llegando a la conclusión, de la importancia que tiene para el país implementar alguno que le permita a los especialistas realizar estos cálculos con mayor eficiencia y rapidez.

Capítulo 2: Métodos y Materiales

CAPÍTULO 2: MÉTODOS Y MATERIALES

Introducción

En el presente capítulo se pone de manifiesto la identificación del problema a resolver, así como la adquisición del conocimiento y las fuentes de donde fue extraído el mismo. Brinda un ejemplo de la forma que se representa el conocimiento y muestra el mecanismo de búsqueda que se utiliza. Se presentan algunos modelos matemáticos que hay que tener en cuenta a la hora de definir alguno propio para las características de la población cubana, modelos que serian empleados en los cálculos de riesgo del cáncer genético de mama, ovario y próstata. Se explican las características favorables y desfavorables de los mismos, mostrando además las distintas herramientas utilizadas para la implementación y prueba del sistema desarrollado.

2.1 Identificación del problema a resolver

El desarrollo de la genética en Cuba ha permitido tener una mayor información sobre algunos tipos de enfermedades de origen genético las cuales son algunas de las principales causas de muerte en el país. En este existen un grupo de especialistas que se dedican al estudio de estas enfermedades. Una vez detectadas, se le aplican una serie de medidas de prevención y estrategias de seguimiento a la persona que la padece, con el objetivo de mejorar las condiciones de vida de la misma.

La realización de estos estudios tiene gran importancia, pero se hace un poco difícil ya que en Cuba son pocas las personas que cuentan con el conocimiento y las habilidades para realizar este estudio el cual depende mucho de la disposición que tenga la persona a la cual se le vaya a realizar. Los estudios hoy en día se realizan de forma manual y se archivan por escrito, de ahí que no existe una comunicación entre los especialistas que están en otras localidades.

El presente trabajo de diploma permite dar solución a dicho problema, pues combinando la información proveniente de los árboles genealógicos gestionados por alasARBOGEN 2.0 y los ficheros de reglas generados mediante un editor de reglas, se realiza el cálculo de riesgo a las personas, para enfermedades de origen genético tales como: cáncer de mama, ovario y próstata.

Capítulo 2: Métodos y Materiales

2.2 Adquisición del conocimiento

El módulo de adquisición del conocimiento es el programa que permite a los expertos modificar la base de conocimientos. Es muy útil porque el conocimiento constantemente está cambiando y es necesario actualizarlo. También sirve para corregir errores que existan en dicha base.

Un buen componente de adquisición ayudará considerablemente la labor del Ingeniero del Conocimiento. Este puede concentrarse principalmente en la estructuración del conocimiento sin tener que dedicar tanto tiempo en la actividad de programación.

Requisitos o características:

- El conocimiento, es decir, las reglas, los hechos, las relaciones entre los hechos, etc., debe poder introducirse de la forma más sencilla posible.
- Posibilidades de representación clara de todas las informaciones contenidas en una base de conocimientos.
- Comprobación automática de la sintaxis
- Posibilidad constante de acceso al lenguaje de programación

Para obtener el conocimiento se implemento un editor de reglas para las condiciones y características específicas de la población cubana. El mismo genera un archivo de reglas creado por el ingeniero del conocimiento el cual obtiene el conocimiento por varias vías como es el caso de encuestas, entrevistas a expertos en el dominio, en libro ext.

2.3 Representación del conocimiento

La experiencia es un tipo de conocimiento que poseen los especialistas, es el conocimiento implícito de los especialistas que debe extraerse y hacerse explícito para que pueda codificarse en un sistema experto. La base de conocimientos contiene el conocimiento especializado extraído del experto en el dominio. Es decir, contiene conocimiento general sobre el dominio en el que se trabaja.

Capítulo 2: Métodos y Materiales

El conocimiento de un sistema experto puede representarse de varias maneras. Un método común de representar el conocimiento es en forma de reglas.

Las reglas generadas en el editor se expresan en un pseudocódigo equivalente al formato **when** y **then**, de la forma siguiente:

rule "Edad de la persona"

when

variables: VariablesMama (edadPaciente > 10)

then

variables.totalPorCiento (0.33);

end

Toda regla comienza con un nombre que la identifica **rule** "Edad de la persona" seguido de la palabra reservada **when** la cual inicia la sección de condiciones, que esta compuesta por una variable seguido de el nombre de la clases a la que se hace referencia variables: VariablesMama luego entre paréntesis se pondrían los factores de riesgo que se encuentran en la clase escogida y la comparación que debe cumplir, estas clases son separadas por las palabras reservadas **and** y **or** en dependencia de la condición que se quiera cumplir, en caso de ser verdadera esta condición se ejecutaría el **then**, este esta compuesto por la por el nombre de la clase a la que se hace alusión anteriormente, seguido de la operación que debe ejecutarse dándole el sistema una respuesta al usuario y por ultimo cierra la regla la palabra **end**.

Una regla se activa cuando se satisfacen las cláusulas indicadas por su antecedente, implicando la ejecución de su consecuente. Cuando se activan varias reglas para una misma configuración de las variables de entrada, se genera una salida por cada una de dichas reglas. Dichas salidas son hechos que actualizan el conocimiento actual almacenado en la base de hechos, es decir, que la salida de una regla puede provocar un cambio en el conocimiento que active otras reglas. A este proceso se le conoce como

Capítulo 2: Métodos y Materiales

encadenamiento de reglas. Atendiendo al sentido del encadenamiento, se diferencian dos tipos de inferencia: hacia delante y hacia atrás.

2.4 Encadenamiento hacia atrás

El encadenamiento hacia atrás no tiene como objetivo inferir el valor de una variable, sino la demostración de una hipótesis. Su desarrollo se presenta como sigue:

1. Se define una hipótesis, es decir, un valor a alcanzar en una variable de salida del sistema.
2. Si la hipótesis es un hecho comprendido en la base de conocimiento, se finaliza el proceso. En caso contrario se continúa en 3.
3. Se busca el subconjunto de reglas de la base de conocimiento cuyos consecuentes coincidan con la hipótesis.
4. Se establecen los antecedentes de las reglas del subconjunto seleccionado como nuevas hipótesis a demostrar y se vuelve a 2.

Este tipo de inferencia establece (si existe) una cadena de reglas que prueba la veracidad de una hipótesis.

2.5 Encadenamiento hacia adelante

El encadenamiento hacia adelante es típico de sistemas en los cuales se desea conocer el valor de una variable de salida atendiendo a una serie de valores de entrada. Este tipo de inferencia parte de la observación de hechos en las variables de entrada para, mediante el encadenamiento de reglas, alcanzar un hecho de salida deseado. El proceso es el siguiente:

- Se define el hecho a alcanzar, es decir, las variables de salida del sistema cuyo valor se desea inferir.

Capítulo 2: Métodos y Materiales

- Un conjunto de hechos relativo a las variables de entrada es observado por el sistema, es decir, llega una entrada al sistema.
- Se busca el subconjunto de reglas de la base de conocimiento cuyos antecedentes son satisfechos por los hechos observados.
- Si el subconjunto está vacío, se finaliza el proceso. En otro caso se continúa en 5.
- El subconjunto de reglas seleccionado se activa y da lugar a un número de hechos nuevos igual al tamaño del subconjunto.
- La base de hechos se actualiza con los nuevos hechos.
- Si se ha alcanzado el hecho de salida deseado se finaliza el proceso, en otro caso, se vuelve a 3.

2.6 Modelos matemáticos para el cálculo de riesgo

Todos los datos relacionados a los factores de riesgo que existen en la literatura y que han sido anteriormente expuestos, han permitido diseñar modelos matemáticos para la estimación del riesgo en las pacientes. Estos modelos permiten cuantificar rápidamente el riesgo de desarrollar un cáncer de mama en un período o a una edad determinada en las pacientes. El poder expresar este valor en términos objetivos y cuantitativos facilita la educación de las pacientes, al igual que también permite diseñar racionalmente una estrategia para el manejo del riesgo y eventualmente la selección de pacientes para estudios clínicos de prevención.

En 1989 se desarrolla un modelo para estimar el riesgo de cáncer in situ (no invasor) que se llamo modelo de Gail. Este modelo incluye variables como la edad, edad de menarquía y del primer embarazo de término, historia de familiares de primer grado con cáncer de mama, antecedentes de biopsias mamarias previas y la raza (22). El modelo es bueno para calcular el riesgo de un grupo de mujeres con características determinadas, pero no es tan bueno en predecir el riesgo individual de cada paciente. En general, a pesar de todas las limitaciones de este modelo, otorga en forma rápida la estimación del riesgo

Capítulo 2: Métodos y Materiales

de cáncer de mama en las pacientes. En el **anexo #1** se muestra una interfaz de un sistema informático que implementa este modelo matemático.

En 1994 es publicado el modelo de Claus que está diseñado para las mujeres con antecedentes familiares de cáncer de origen genético. A diferencia del modelo de Gail, éste incluye antecedentes familiares tanto de primer y segundo grado, y la edad de diagnóstico de ellos. El modelo calcula la probabilidad de desarrollar una enfermedad de origen genético en un período de tiempo determinado o a una edad determinada.

Estos modelos son los más usados en los sistemas que existen a nivel mundial, los cuales utilizan un grupo de factores de riesgo que dependen de las localidades donde se usan, por lo que sería bueno contar con un modelo matemático bien definido para las características propias de la población cubana, ya que estos modelos no se ajustan en su totalidad a las características étnicas de nuestra población y el avance de los estudios genéticos y oncológicos en Cuba, permite definir un modelo adecuado en este sentido.

Por ello se opta por la definición de un sistema basado en reglas con dos características fundamentales:

- Un motor de regla que trabaja con las variables definidas como riesgo y la información de los árboles genealógicos.
- Un editor de reglas que permite, en esta fase inicial del sistema, a los especialistas que diseñen sus propias reglas y desarrollen investigaciones que garanticen la construcción de un modelo matemático bien definido para nuestro país.

2.7. Tecnologías y herramientas para el desarrollo del sistema

Las tecnologías y herramientas que a continuación se presentan son las utilizadas para el desarrollo de este trabajo. Entre las cuales se encuentran las utilizadas para la modelación e implementación de la aplicación, así como las utilizadas para el entendimiento de las técnicas de IA.

Capítulo 2: Métodos y Materiales

2.7.1 Plataforma de Desarrollo y Lenguaje de Programación

En cuanto a plataforma escogida se optó por jre versión 1.6.0.

Originalmente llamado OAK por los ingenieros de Sun Microsystems, Java fue diseñado para correr en computadoras incrustadas. Sin embargo, en 1995, dada la atención que estaba produciendo la Web, Sun Microsystems la distribuyó para sistemas operativos tales como Microsoft Windows. El lenguaje mismo se inspira en la sintaxis de C++, pero su funcionamiento es más similar al de Smalltalk que a éste. Incorpora sincronización y manejo de tareas en el lenguaje mismo (similar a Ada) e incorpora interfaces como un mecanismo alternativo a la herencia múltiple de C++.

Los programas en Java generalmente son compilados a un lenguaje intermedio llamado bytecode, que luego son interpretados por una máquina virtual (JVM). Esta última sirve como una plataforma de abstracción entre la máquina y el lenguaje permitiendo que se pueda "escribir el programa una vez, y correrlo en cualquier lado". También existen compiladores nativos de Java, tanto software libre como no libre. El compilador GCC de GNU compila Java a código de máquina con algunas limitaciones al año 2002.

Con la evolución de las diferentes versiones, no sólo se han producido cambios en el lenguaje, sino que se han producido cambios mucho más importantes en sus bibliotecas asociadas, que han pasado de unos pocos cientos en Java 1.0, a más de tres mil en Java 5.0. En particular, se han añadido APIs completamente nuevas, tales como Swing y Java2D. (23) (24)

2.7.2 Entorno de Desarrollo

Eclipse es una entorno de desarrollo open source basada en Java. Es un desarrollo de IBM cuyo código fuente fue puesto a disposición de los usuarios. En sí mismo Eclipse es un marco y un conjunto de servicios para construir un entorno de desarrollo a partir de componentes conectados (plug-in). Hay plug-ins para el desarrollo de Java (JDT Java Development Tools) así como para el desarrollo en C/C++, COBOL, etc. (36). En esta investigación se utilizaron plug-ins que permiten el entendimiento y desarrollo de las técnicas de IA, así como el funcionamiento del motor de reglas Drools.

Capítulo 2: Métodos y Materiales

2.7.3 Herramienta CASE (Computer Aided Software Engineering)

Las Herramientas CASE (*Computer Aided Software Engineering*, Ingeniería de Software Asistida por Ordenador) son diversas aplicaciones informáticas destinadas a aumentar la productividad en el desarrollo de software reduciendo el coste de las mismas en términos de tiempo y de dinero. Estas herramientas nos pueden ayudar en todos los aspectos del ciclo de vida de desarrollo del software en tareas como el proceso de realizar un diseño del proyecto, cálculo de costes, implementación de parte del código automáticamente con el diseño dado, compilación automática, documentación o detección de errores entre otras.

Esta herramienta CASE utiliza “UML”: como lenguaje de modelaje. Permite además, que a partir de un modelo relacional en Sql Server, MySql, etc. Para gestionar la persistencia y el mapeo de estas clases con la base de datos utiliza Hibernate para Java y NHibernate en el caso de un proyecto .Net. Genera la documentación del proyecto automáticamente en varios formatos como Web o Pdf, y permite control de versiones, ingeniería inversa, generación de código, importación desde Rational Rose, exportación/importación XML, generador de informes, editor de figuras, integración con MS Visio, plug-in, integración IDE con Visual Studio, IntelliJ IDEA, Eclipse, NetBeans y otros.

2.7.4 Lenguaje de Modelado

El Lenguaje de Modelado Unificado (UML: Unified Modeling Language) es la sucesión de una serie de métodos de análisis y diseño orientadas a objetos que aparecen a fines de los 80's y principios de los 90s. UML es llamado un lenguaje de modelado, no un método. Los métodos consisten de ambos de un lenguaje de modelado y de un proceso. El UML, fusiona los conceptos de la orientación a objetos aportados por Booch, OMT y OOSE (Booch, G. et al., 1999). UML incrementa la capacidad de lo que se puede hacer con otros métodos de análisis y diseño orientados a objetos. Los autores de UML apuntaron también al modelado de sistemas distribuidos y concurrentes para asegurar que el lenguaje maneje adecuadamente estos dominios.

El lenguaje de modelado es la notación (principalmente gráfica) que usan los métodos para expresar un diseño. El proceso indica los pasos que se deben seguir para llegar a un diseño. (25)

Capítulo 2: Métodos y Materiales

2.7.5 Drools como Motor de Reglas

Drools son reglas de negocio manejadas por un motor de reglas de aplicación en un sistema. El motor de reglas se encuentra implementado por ReteOO, que es una base del algoritmo Rete de Charles Forgy adaptados para el lenguaje Java. Más importante aún, Drools es lo suficientemente flexible como para que coincida con la semántica del dominio de su problema en determinados idiomas, provee herramientas de edición gráfica, herramientas basadas en Web y desarrolladores de herramientas de productividad. (26)

El mismo está compuesto por:

- **Drools Expert:** Motor de inferencia de reglas empresariales, nos permite desacoplar reglas de negocio de nuestras aplicaciones y utilizar su motor de inferencia que utiliza para la deducción encadenamiento hacia adelante (forward chaining) y una extensión del Algoritmo RETE (llamada RETEOO).
- **Drools RuleFlow:** módulo de BPMWorkflow, que nos permite definir procesos de negocio utilizando una API unificada para aprovechar las ventajas y la sesión del motor de inferencia. Permitiendo de esta manera una solución elegante y completamente integrada para el modelado de problemas empresariales reales.
- **Drools Fusion:** módulo de procesamiento de eventos complejos, se agrega a la integración, este interesante módulo que permite que nuestras reglas de negocio puedan realizar deducciones basadas en datos temporales. También de una manera integrada al set de APIs.
- **Drools Guvnor:** sistema de administración de reglas empresariales que nos permitirá centralizar todos nuestros recursos de reglas, RuleFlow y consumirlos desde este repositorio.

Principales Características

- **Código abierto:** Drools tiene licencia de fuente abierta que le hace libre para descargar, usar, integrar y distribuir.

Capítulo 2: Métodos y Materiales

- **Programación Declarativa:** Permite decir “que hacer” y no decir “como hacerlo”.
- **Accesibles reglas de negocio:** Drools es amigable tanto a desarrolladores como administradores de las reglas de negocio. DSLs permitirá a los desarrolladores a escribir casi todas las reglas de negocio como un lenguaje natural. Flujo de reglas, cuadros de decisión, en hojas de cálculo son algunas de las herramientas para poder definir las reglas, incluso existe una solución Web para la gestión de reglas.
- **Lógica y los datos por separado:** Los motores de reglas es lo último en tecnología para el desacople de logia y datos. Aumenta la mantenibilidad y la calidad del código.
- **Velocidad y escalabilidad:** El algoritmo de Rete, y sus descendientes, como Drools' ReteOO, proporcionan la velocidad y la escalabilidad, más allá de que es práctico para mantener a mano.
- **Eclipse como procesador:** Hace que sea más fácil utilizar Drools, con IntelliSense, auto-completo y vistas para debug, flujo de reglas, gráficos entre otros.
- **Basado en Web:** Basado en AJAX, construye desde cero hasta avanzadas reglas de aplicación, control de versiones, y de administración.

Conclusiones

En este capítulo se pudo identificar la forma en la cual adquirimos el conocimiento para el desarrollo de la aplicación, así como su forma de representarlo. Esta forma es a través de las reglas, donde se pone además un ejemplo de código de la misma. Se identificaron dos modelos matemáticos fundamentales que se deben tener en cuenta a la hora de realizar uno propio para las características de la población cubana. Se determinaron además las herramientas principales para el modelado e implementación del sistema.

Capítulo 3: Resultados y Discusión

CAPÍTULO 3: RESULTADOS Y DISCUSIÓN

Introducción

En el presente capítulo se expone de manera clara y organizada la forma en la que se aplicaron las tecnologías y herramientas que se determinaron en la investigación para el desarrollo de la aplicación, así como los resultados alcanzados durante este proceso.

3.1 Modelo conceptual del sistema

Un modelo conceptual es un diagrama que ilustra una serie de relaciones entre ciertos factores que se cree impactan o conducen a una condición de interés, y es la herramienta más importante del análisis orientado a objetos. Es la representación de cosas del mundo real y no de componentes de software, en él no se definen operaciones. Puede representarse mediante un diagrama de estructura estático (notación UML). (27)

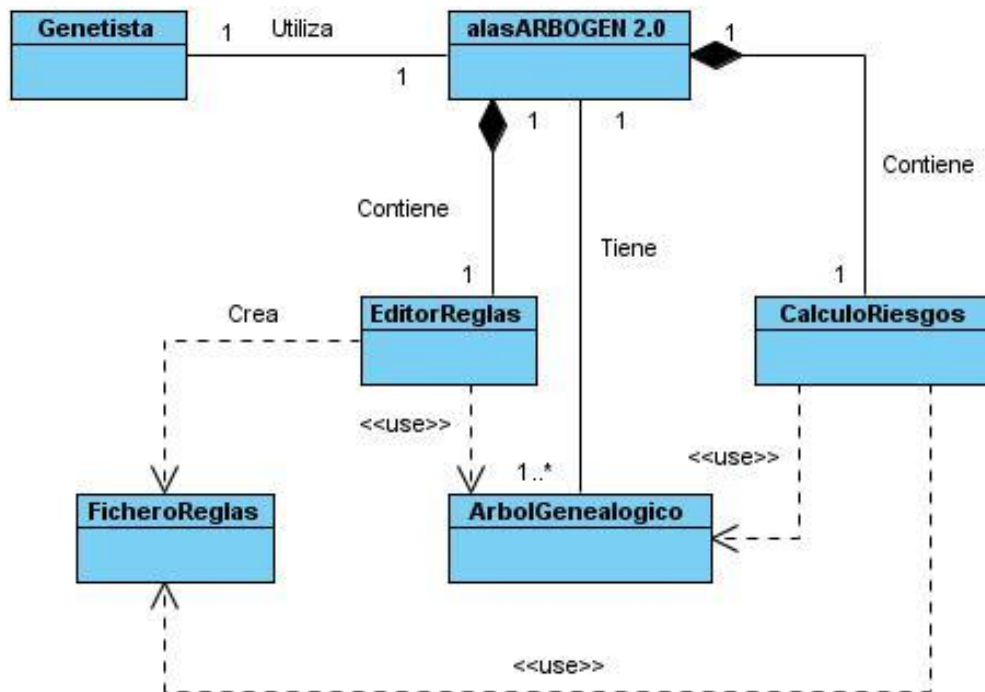


Figura 3.1 Modelo Conceptual.

Capítulo 3: Resultados y Discusión

Cada uno de los elementos que componen este modelo, representa en si un concepto específico que se tiene en cuenta al desarrollar el sistema, a continuación se brinda una explicación de todos ellos.

El **Genetista** es el usuario principal del sistema alasARBOGEN 2.0, este usuario tiene derecho a todas las funcionalidades que se brindan en dicha aplicación.

El sistema **alasARBOGEN 2.0** está destinado a la representación gráfica del árbol genealógico de un individuo y su familia, además de gestionar la información que figura en el genograma, pero además posee otras funcionalidades como son la de poder calcular el riesgo de padecimiento de enfermedades de origen genético, además de poder editar ficheros de reglas imprescindibles para estos cálculos.

El **Árbol Genealógico** es un diagrama generado por alasARBOGEN 2.0 que muestra las relaciones entre los miembros de una familia, este genograma también muestra las características de ciertos genes o enfermedades dentro de una familia.

El **Editor de Reglas** es una aplicación que le brinda la posibilidad al usuario de editar sus propias reglas utilizando la información proveniente de los árboles generados por alasARBOGEN 2.0, permitiéndole almacenar dichas reglas en un fichero .DRL el cual después será usado para realizar el cálculo de riesgo.

Un **Fichero de Reglas** es un archivo con extensión .DRL, en el cual se almacenan una o varias reglas cumpliendo con un formato definido en epígrafes anteriores.

EL **Cálculo de Riesgo** es una de las funcionalidades que brinda alasARBOGEN 2.0, para el desarrollo de este cálculo son importantes el uso de dos parámetros fundamentales, el primero está compuesto por todos los datos de la persona que se seleccione en un árbol genealógico determinado y el segundo es el archivo que se genera a partir del Editor de Reglas.

3.2 Diagrama de clases y Patrones utilizados

3.2.1 Diagrama de clases

Los diagramas de clases son utilizados durante el proceso de análisis y diseño de los sistemas, los cuales ayudan a entender la estructura de clases de los proyectos desarrollados por otras personas (o

Capítulo 3: Resultados y Discusión

desarrollados por el usuario hace mucho tiempo). Puede utilizarlos para adaptar y presentar la información del proyecto, o compartirla con otros usuarios. (28)

A continuación se muestra el diagrama de clases utilizado en nuestra aplicación.

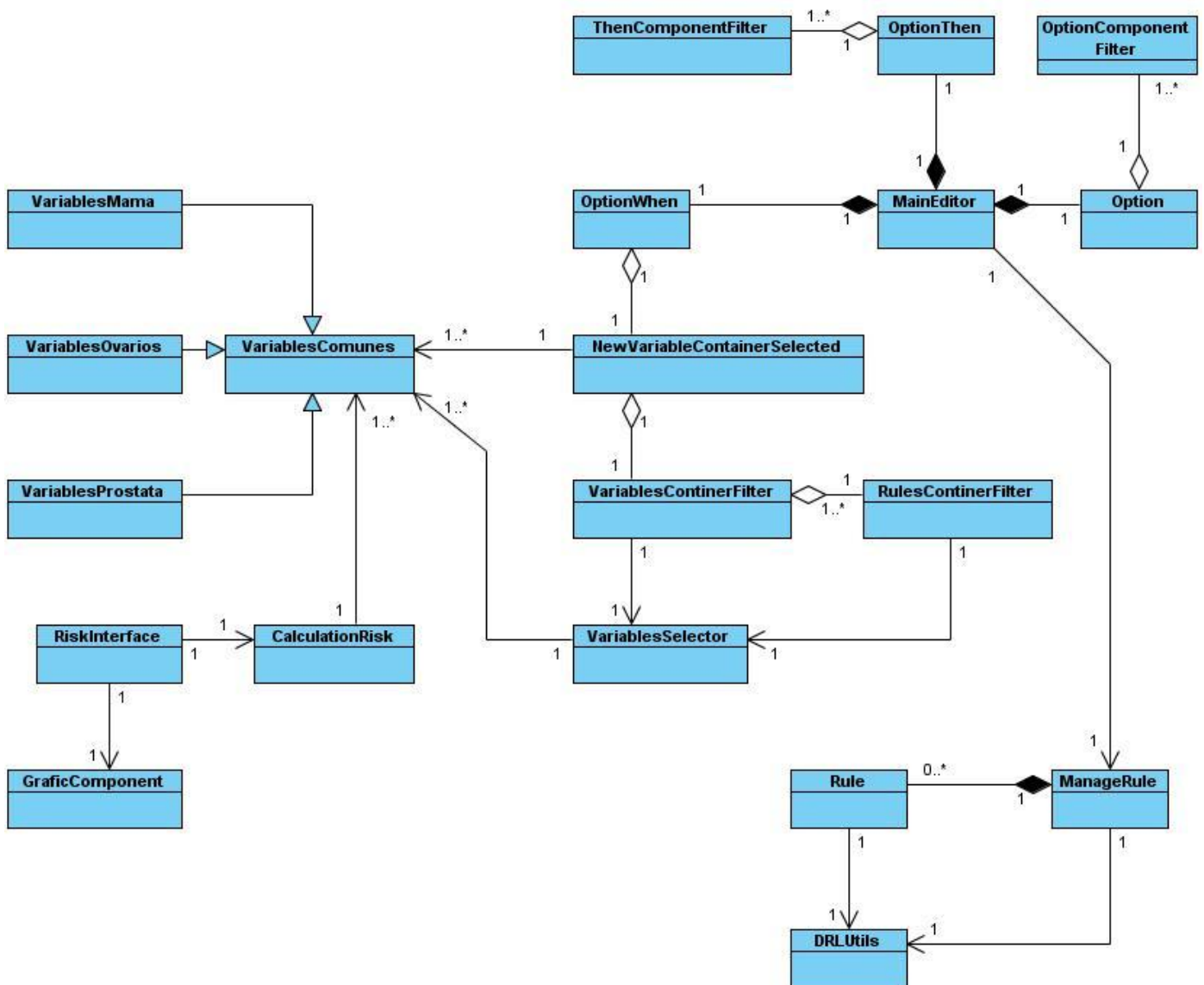


Figura 3.2 Diagrama de clases.

Capítulo 3: Resultados y Discusión

En el diagrama presentado en la **Figura 3.2** se muestran las clases más importantes de la aplicación, así como las relaciones que las unen. En él se presentan las clases **VariablesComunes**, **VariablesMama**, **VariablesOvarios** y **VariablesPróstata**, las cuales se encargan de ejecutar las operaciones para el manejo de la información que se obtiene de los árboles genealógicos. La clase **MainEditor** es la interfaz principal del editor de reglas, a partir de esta se comienza con el desarrollo de las reglas, ejecutando cada una de las otras clases que se necesiten para la edición y salva del archivo .DRL resultante en el orden que se muestra en el diagrama de acuerdo a las relaciones que unen dichas clases. Por otra parte está **RiskInterface**, esta constituye la interfaz principal del calculador de riesgo, utiliza el componente que genera la clase **GraficComponent** para mostrar el resultado del cálculo, resultado que se obtiene mediante la clase **CalculationRisk**. Esta clase contiene los métodos que vinculan el archivo de reglas con los datos gestionados en **VariablesComunes**, **VariablesMama**, **VariablesOvarios** y **VariablesPróstata**. Se decidió ocultar los atributos y métodos de cada una de las clases debido a que son muchos y dificultaría la comprensión del diagrama al aumentar este mucho más su tamaño.

3.2.2 Patrones utilizados

Patrones GRASP

En la tecnología de objetos un **Patrón** es una descripción de un problema y la solución, a la que se le da un nombre, y que se puede aplicar a nuevos contextos.

Los patrones GRASP describen los principios fundamentales de diseño de objetos para la asignación de responsabilidades. Constituyen un apoyo para la enseñanza que ayuda a entender el diseño de objeto esencial y aplica el razonamiento para el diseño de una forma sistemática, racional y explicable. (42)

En el anexo # 2 se representan una tabla con una explicación mas detallada de cada uno de los patrones usados en nuestra investigación. (29)

El patrón **Experto** se ve evidenciado en la clase **CalculationRisk**, ya que al tener un objeto de **File** como archivo de reglas, además un objeto de **VariablesComunes**, **VariablesMama**, **VariablesOvarios** y de **VariablesPróstata**, las cuales se encargan de la gestión de variables, entonces posee la información

Capítulo 3: Resultados y Discusión

necesaria para realizar el cálculo de riesgo de padecimiento de enfermedades de origen genético. De esta manera se le asigna a la clase **CalculationRisk** la responsabilidad de realizar uno o varios cálculos de riesgos

CalculationRisk
-fileRules : File = null -varComunes : VariablesComunes -varMama : VariablesMama -varOvarios : VariablesOvarios -varProstata : VariablesProstata
+CalculationRisk() +calculatorRiskVC(edadpersona : int, cantFamiliaresConCancer : int) : float [] +calculatorRiskVM(lactancia : string, origenEtnico : string, edadPMestrua : int, cantBSeno : int, edadPParto, dietaGrasa : string, consuAlcohol : string) : float[] +calculatorRiskVO(indiceOvesidad : float, lactancia : string, cantHijos : int) : float[] +calculatorRiskVP(origenEtnico : string, indiceObesidad : int, tipoTrabajo : string) : float[] +getFileRules() : File +setFileRules(fileRules : File) : void +readRule(file : File) : RuleBase

Tabla 3.3 Ejemplo de una clase experta.

La clase **ManageRule** contiene un objeto de **Rule**, por esto, el patrón **Creador** sugiere que la clase **ManageRule** es idónea para asumir la responsabilidad de crear instancias de la clase **Rule**.

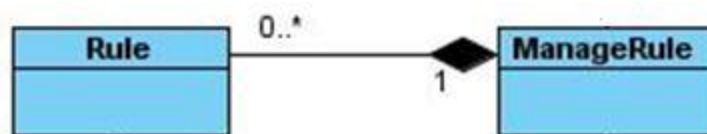


Tabla 3.4 Ejemplo de la aplicación del patrón Creador.

Capítulo 3: Resultados y Discusión

3.3 Descripción de la solución

Para la implementación del sistema como solución al problema planteado, se decidió dividir el trabajo en dos partes fundamentales, las cuales son integradas al sistema alasARBOGEN 2.0, ya que a partir los árboles genealógicos generados por dicho sistema se obtienen los datos de los pacientes a tratar, necesarios para el funcionamiento de nuestra aplicación.

Una primera parte que se encarga del **Cálculo de Riesgo** de padecimiento de enfermedades de origen genético, y una segunda parte que maneja todo lo referente a la creación de ficheros de reglas a través de un **Editor de Reglas**.

3.3.1 Principales aspectos del módulo Cálculo de Riesgo

Accesibilidad

Se accede a través del menú Cálculo de riesgo ubicado en el menú principal de la alasARBOGEN 2.0.

Funcionalidades que brinda

Le permite al genetista realizar el cálculo de riesgo para cada una de las enfermedades que se relacionan a continuación:

- Cáncer de Mama (Seno).
- Cáncer de Ovarios.
- Cáncer de Próstata.

Dichos cálculos se hacen con un período de cinco años en dependencia de la edad del paciente, es decir, el sistema muestra el resultado de la predicción en una gráfica por ciento de riesgo/edad del paciente, donde la primera barra representa el por ciento de riesgo para la edad obtenida del árbol genealógico, la segunda barra representa el por ciento pero con cinco años de diferencia, y la tercera para diez años de diferencia. Además el sistema permite imprimir esta gráfica

Funcionamiento

Capítulo 3: Resultados y Discusión

A partir de la ejecución del sistema alasARBOGEN 2.0, el usuario tendrá la oportunidad de realizar el cálculo de riesgo de padecimiento de una enfermedad de origen genético. Para esto lo primero que debe hacer es seleccionar del árbol genealógico el paciente a tratar, para así obtener los datos que en él se almacenan. Una vez seleccionado del árbol el paciente en cuestión, procede a ejecutar el menú Cálculo de riesgo, generando así una nueva área con las funcionalidades referentes al cálculo.

En esta nueva vista lo primero que se hace es seleccionar el tipo de enfermedad, de ahí procede a introducir los otros datos del paciente relacionados con cada enfermedad en específico, lo cual se realiza en una tabla de propiedades que se genera a partir de la selección del tipo de cáncer, donde también se selecciona el archivo de reglas correspondiente.

Una vez introducidos los datos correctamente procede a realizar el cálculo y luego a imprimir el resultado en caso de que así lo desee.

3.3.2 Principales aspectos del módulo Editor de Reglas

Para trabajar en el editor solamente tiene que tener conocimiento de que las reglas que ahí va a crear sean avaladas por una investigación científica aprobada, para así poder obtener un archivo de reglas válido.

Accesibilidad

Se accede a través del menú Cálculo de riesgo ubicado en el menú principal de la alasARBOGEN 2.0.

Funcionalidades que brinda

Crear un nuevo archivo el cual está compuesto por una o varias reglas de las que ahí se editan. La aplicación muestra dos áreas fundamentales, en uno se encuentra el editor en sí, donde se da formato a las reglas por la parte del cliente y en el otro se va mostrando en tiempo real el código DRL de cada una de las reglas generadas por el editor. Permite exportar todas estas reglas en un fichero con extensión .DRL.

Funcionamiento

Capítulo 3: Resultados y Discusión

Una vez ejecutada la aplicación alasARBOGEN 2.0, a partir de la cual el genetista accede al Editor, se podrá comenzar a desarrollar el proceso de elaboración de ficheros de reglas. Para esto lo primero que se debe hacer es crear un nuevo archivo de reglas, esta acción lo que indica es la dirección en donde se guardara el fichero de reglas resultante y habilita el área de edición.

El usuario comienza con agregar un nuevo contenedor de variables en la sección “Editar condiciones de la regla”, este filtro contiene todas las variables correspondientes a cada enfermedad en particular, y se podrá agregar hasta dos filtros máximo de acuerdo con la selección que haga. A partir de ahí se comienzan a introducir tantas comparaciones como el usuario crea necesarias.

Otro paso importante es la introducción del valor de porciento que arrojan estas comparaciones en caso que se cumplan, esto se hace en la sección “Editar las acciones a realizar”.

También se puede introducir el nombre del autor y la fecha de creación de cada regla, lo cual es opcional. Estas operaciones se realizan en la sección “Datos opcionales”.

Capítulo 3: Resultados y Discusión

3.4 Interfaces del sistema

Para realizar el cálculo de riesgo el usuario lo hace a partir del sistema alasARBOGEN 2.0.

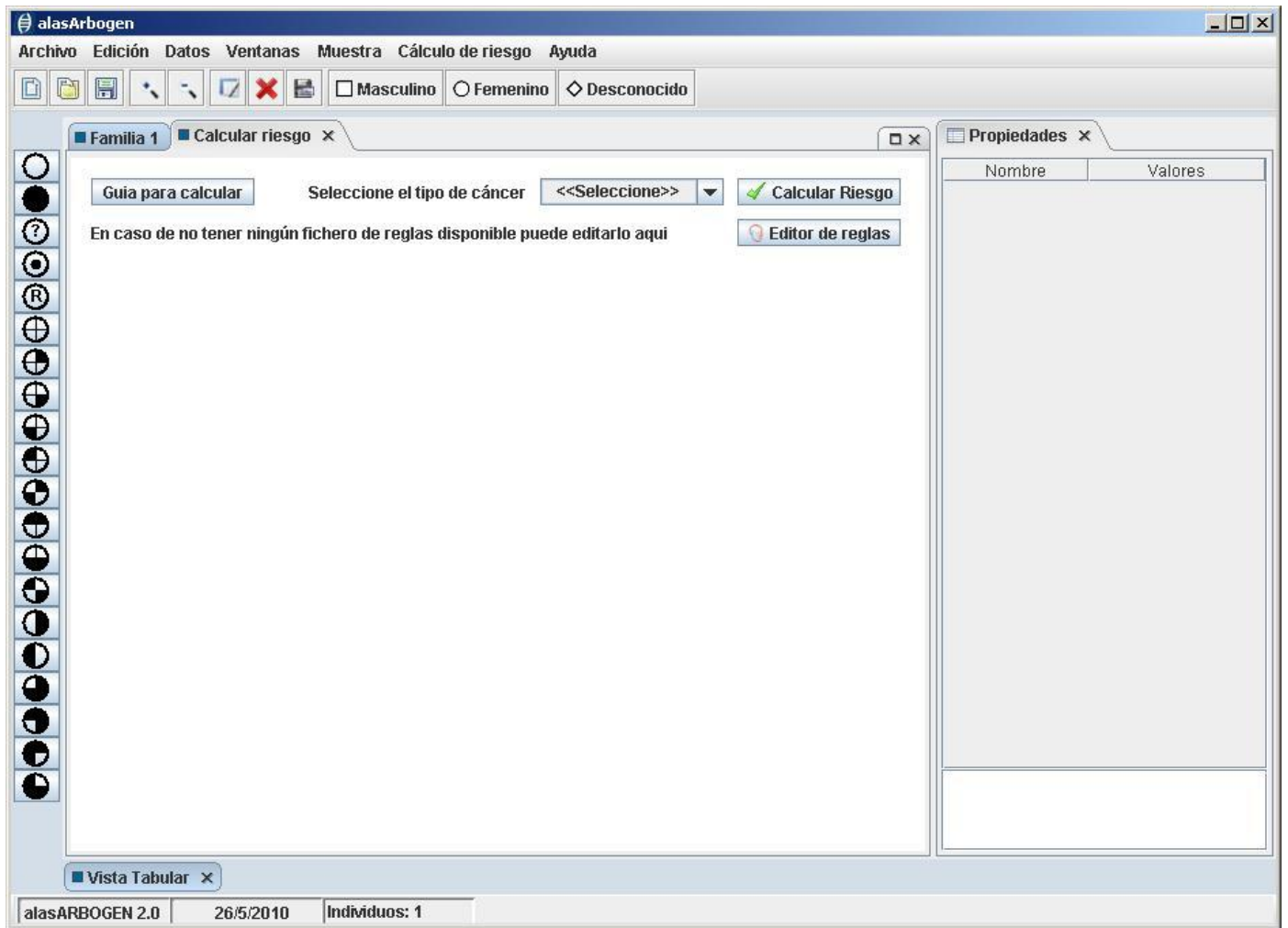


Figura 3.5 Interfaz inicial del módulo Cálculo de Riesgo.

Una vez ahí el usuario procede a la selección de fichero de reglas correspondiente al tipo de cáncer especificado.

Capítulo 3: Resultados y Discusión

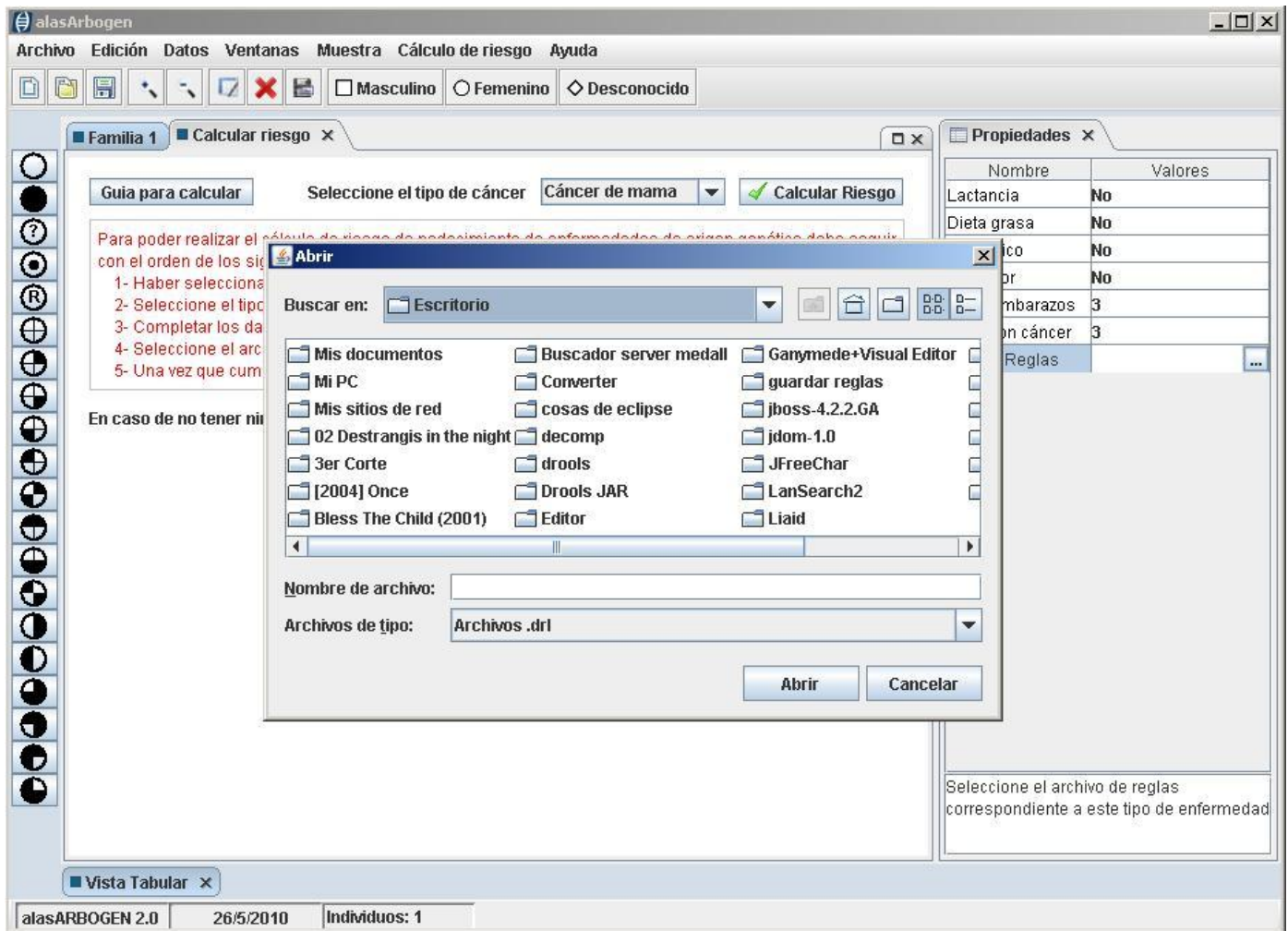


Figura 3.6 Cargando archivo de reglas del cáncer especificado.

Luego de tener el archivo de reglas seleccionado de acuerdo al cáncer que se especificó, así como haber completado todos los campos de la tabla de propiedades correspondiente, se procede a calcular el riesgo y se muestra el resultado en una gráfica de barras como se muestra a continuación.

Capítulo 3: Resultados y Discusión

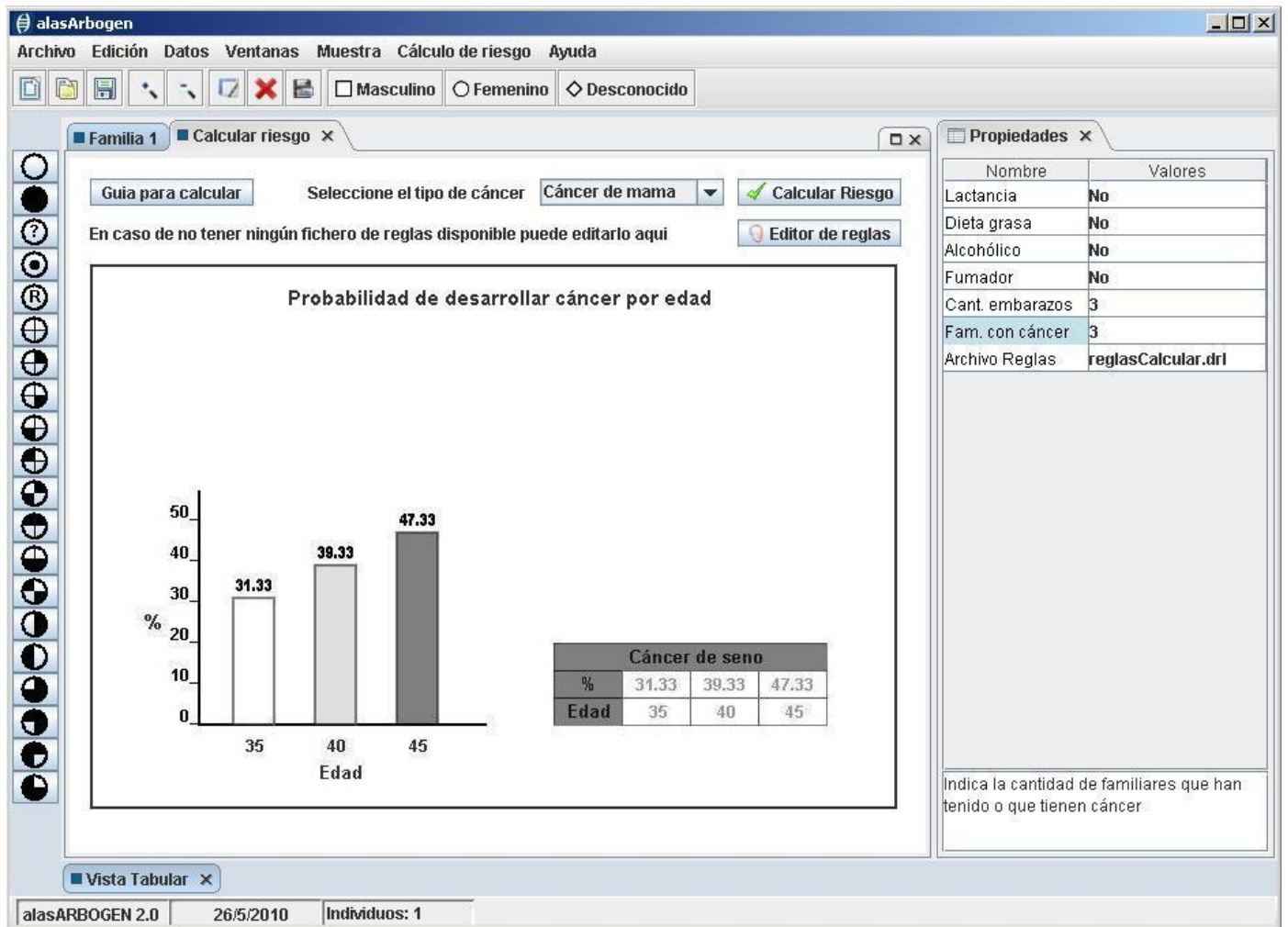


Figura 3.7 Representación gráfica de los resultados obtenidos.

Por otra parte el Editor de Reglas en cuanto a interfaces trabaja de forma independiente, es decir es una aplicación individual que se ejecuta a partir de alasARBOGEN 2.0.

Capítulo 3: Resultados y Discusión

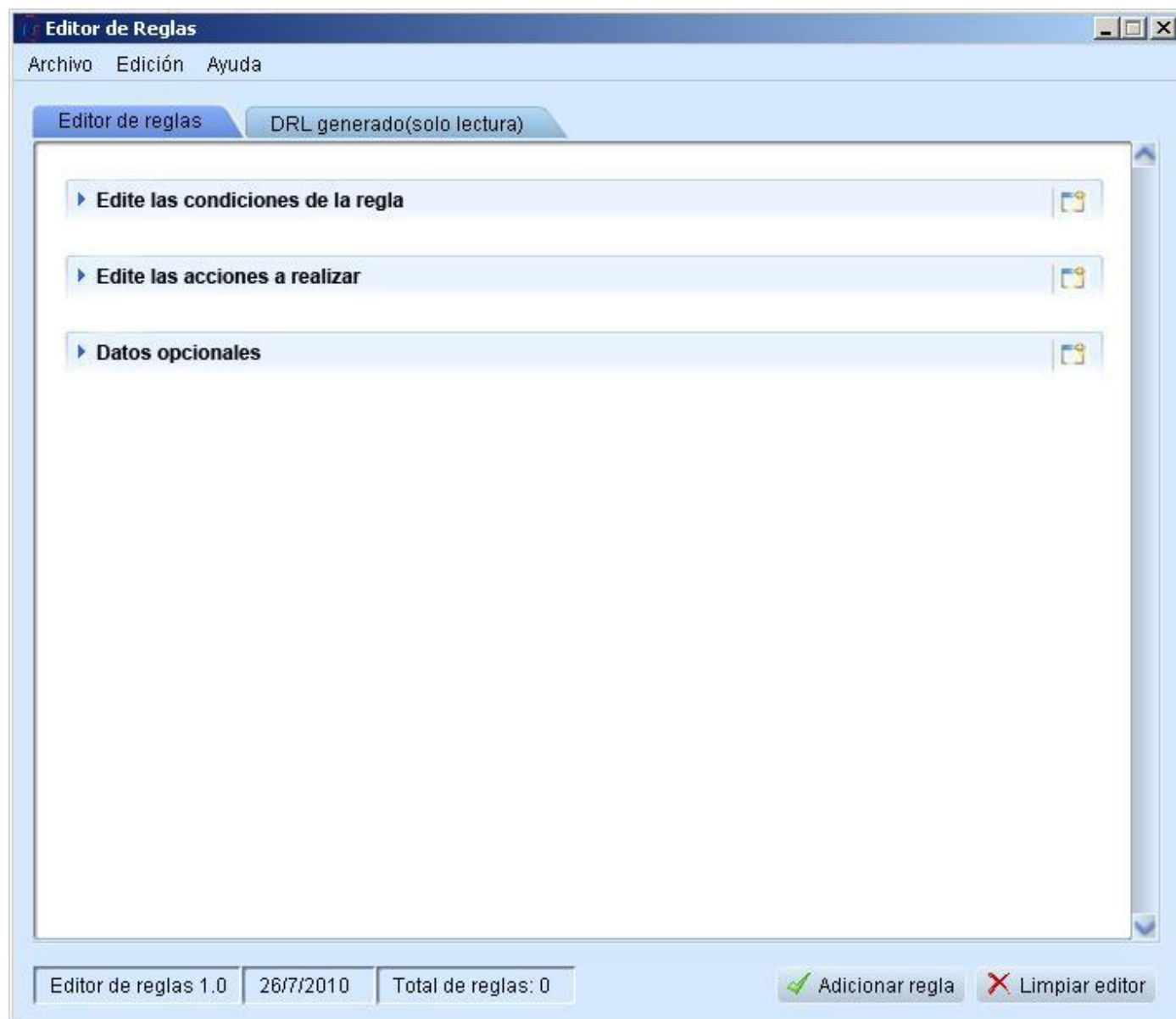


Figura 3.8 Interfaz principal del Editor de Reglas.

Una vez ejecutada este módulo, el usuario procede confeccionar las reglas de acuerdo a los pasos que brinda el editor. Luego de elaborar dichas reglas se muestra en tiempo real el código DRL que se genera, para dar paso al final a exportar el fichero .DRL.

Capítulo 3: Resultados y Discusión

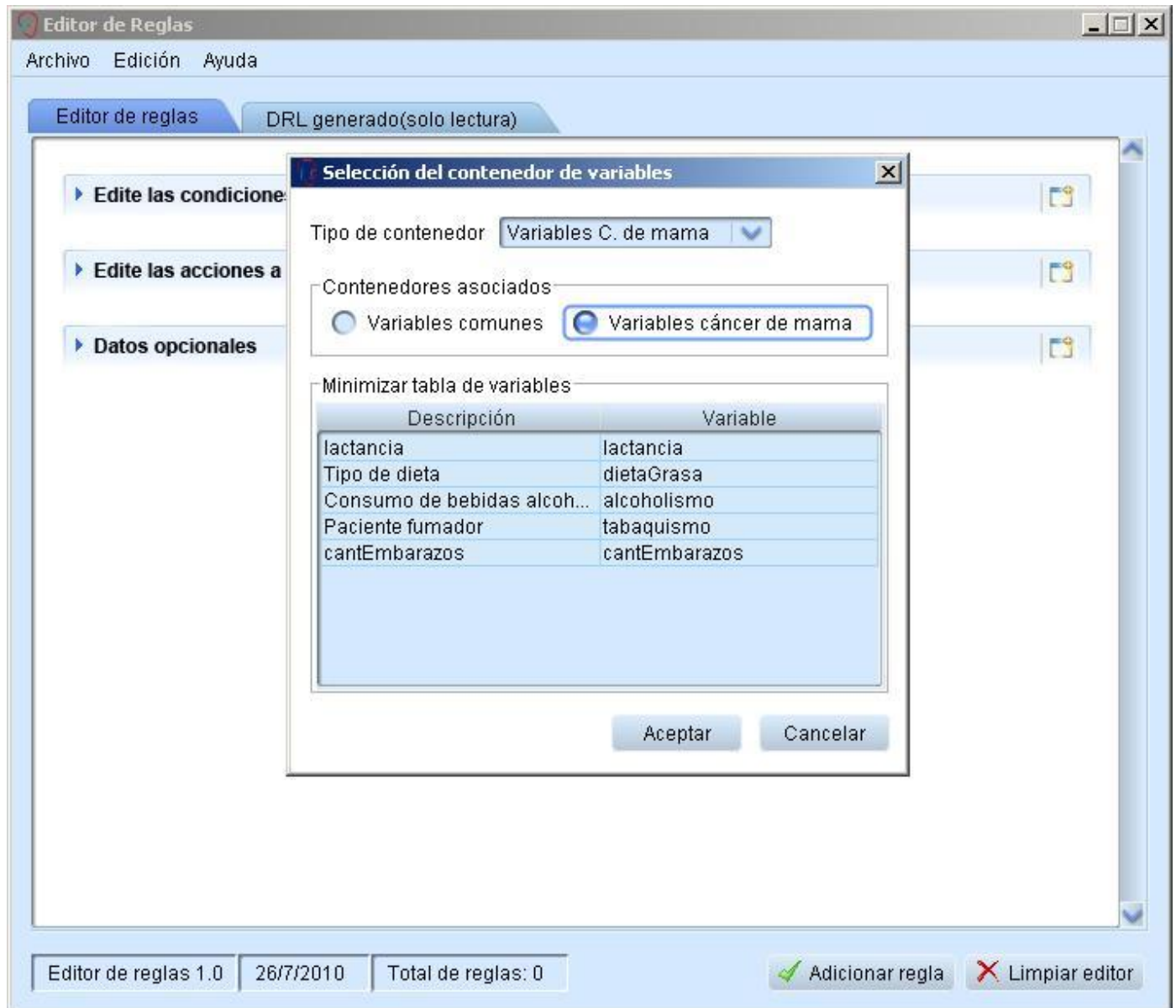


Figura 3.9 Selección de los contenedores de variables.

Capítulo 3: Resultados y Discusión



Figura 3.10 Edición de las comparaciones.

Capítulo 3: Resultados y Discusión

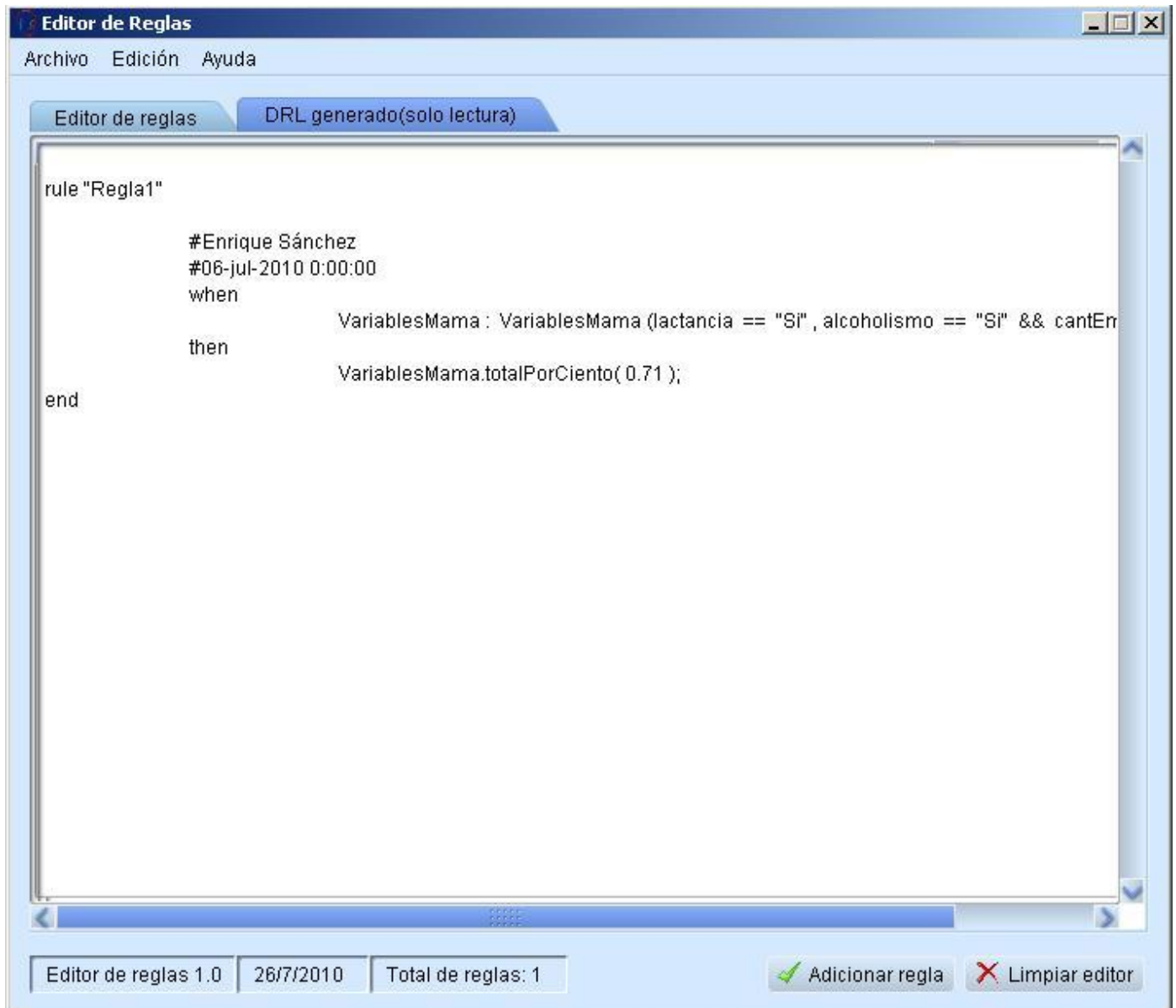


Figura 3.11 Código DRL generado.

Conclusiones

En el presente capítulo se identificaron aspectos fundamentales como son el diagrama de clases, el modelo conceptual y los patrones de diseño que nos permitieron desarrollar una aplicación amigable y

Capítulo 3: Resultados y Discusión

flexible a futuros cambios, de la cual se muestran además algunas imágenes con las principales interfaces con las que el usuario interactúa.

CONCLUSIONES

Al culminar con el desarrollo de la investigación arribamos a las siguientes conclusiones:

- Se realizó una búsqueda en la bibliografía, en cuanto al tema del cálculo de riesgo de padecimiento de enfermedades de origen genético, identificándose las herramientas y sistemas más usados a nivel mundial en este tipo de procedimiento.
- Se implementó un sistema basado en reglas, el cual será usado para realizar el cálculo de riesgo en enfermedades como el cáncer de mama, ovario y próstata, en correspondencia con la metodología y patrones escogidos, los cuales facilitaron que el sistema sea flexible a futuros cambios y abierto a la reutilización de código.
- Se implementó un editor de reglas el cual permite desarrollar ficheros de reglas que serán usados para realizar el cálculo de riesgo.
- Se integró el sistema basado en reglas y el editor a *alasARBOGEN 2.0* satisfactoriamente, permitiendo el trabajo con la información procedente de los árboles genealógicos.

RECOMENDACIONES

Para un correcto funcionamiento del sistema recomendamos que se trabaje en:

- Determinar por parte de los especialistas en genética clínica un conjunto de reglas válidas según el modelo de BRCAPRO para las condiciones y características específicas de la población cubana.
- Incluir otras reglas, en dependencia del avance de los estudios del riesgo de padecimiento de otros cánceres de origen genético y enfermedades multifactoriales en general.
- Introducir una medida asociada a la incertidumbre de las reglas y a la de sus premisas.

Referencias Bibliográficas

TRABAJOS CITADOS

1. **Riera, Lilliam.** Gobierno cubano apoya estudios en Genética Médica. *Granma*. 24 de Julio de 2009 .
2. **Santos, Dr. Manuel.** Patología Pediátrica. *Generalidades sobre Enfermedades Genéticas en Pediatría*.
3. **Marinello, Lic. Sofía González.** [aut. libro] Dr. Felipe Chávez Martínez, Dra. Teresa Romero Pérez y Juan Lence Anta Lic. Sofía González Marinello. *Riesgo de morir por cáncer de mama en Cuba*. Ciudad de La Habana, Cuba : s.n., 29 de enero de 1997.
4. **Barnard, Chester I.** Exponentes de la Escuela Matemática.: [aut. libro] Herbert Simon, James E. March Chester I. Barnard. *Exponentes de la Escuela Matemática* .
5. **Santana Tejero, Ruth.** *DENDRAL*.
6. *Revista médica de Chile.* **Eduardo Atalah S, Carmen Urteaga R, Annabella Rebolledo A.** Santiago : s.n., 2000.
7. **Jboss Community.** Jboss Community. [En línea] <http://www.jboss.org/drools/>.
8. **Levy, M.** SESAM-Diabetes, un sistema experto para la insulina que requieren educación del paciente diabético. [En línea] octubre de 1989. [Citado el: 16 de enero de 2010.] <http://translate.google.com.cu/translate?hl=es&langpair=en|es&u=http://linkinghub.elsevier.com/retrieve/pii/0010480989900372>.
9. **Doukidis, Georgios I.** El potencial para el diagnóstico asistido por ordenador de las enfermedades tropicales en los países en desarrollo: Un caso de estudio de expertos del sistema. [En línea] 20 de mayo de 2003. [Citado el: 16 de enero de 2010.] <http://translate.google.com.cu/translate?hl=es&langpair=en|es&u=http://linkinghub.elsevier.com/retrieve/pii/037722179090345C>.

Referencias Bibliográficas

10. **Cyrillic.** Cyrillic. *Cyrillic*. [En línea] [Citado el: 16 de enero de 2010.] <http://www.softwarecientifico.com/paginas/cyrillic.htm>.
11. **Instituto Nacional de Cáncer.** Instituto Nacional de Cáncer. [En línea] [Citado el: 16 de enero de 2010.] <http://directorscorner.cancer.gov/diccionario/?CdrID=657827>.
12. **UT Southwestern Medical.** UT Southwestern Medical. [En línea] [Citado el: 16 de enero de 2010.] <http://www.utsouthwestern.edu/error.html>.
13. **D Medicina.** D Medicina. [En línea] 2009. [Citado el: 20 de enero de 2010.] <http://www.dmedicina.com/enfermedades/cancer/cancer-mama>.
14. **División de Prevención y Control del Cáncer (DCPC).** División de Prevención y Control del Cáncer (DCPC). [En línea] 25 de marzo de 2010. [Citado el: 20 de enero de 2010.] <http://www.cdc.gov/spanish/cancer/breast/statistics/race.htm>.
15. **Gardner, Amanda.** La lactancia podría reducir el riesgo de cáncer de mama. [En línea] 2009. [Citado el: 20 de enero de 2010.] <http://www.healthfinder.gov/news/newsstory.aspx?docid=629914>.
16. **Martínez, Dr. Benjamín Berdión.** *Pesquisa masiva de cáncer de mama con un equipo móvil de mamografía*. Santiago de Cuba, Cuba. : s.n., 12 de enero de 1994.
17. **Mora, Dr. Rubén Elzaurdín.** Actualización bibliográfica sobre cáncer de ovario. *Formación Para Médicos*. [En línea] 10 de abril de 2009. [Citado el: 11 de enero de 2010.] <http://www.portalesmedicos.com/publicaciones/articulos/1413/1/Actualizacion-bibliografica-sobre-cancer-de-ovario.html>.
18. **Martines, Dr. Luis.** Manual Urología. *Cáncer de la próstata*. [En línea] [Citado el: 11 de enero de 2010.] <http://escuela.med.puc.cl/publ/manualurologia/CancerProstata.html>.
19. **Graide, Luis León Cárdenas.** MYCIN: Medical Diagnosis using Production Rules. *Inteligencia Artificial*. [En línea] [Citado el: 16 de enero de 2010.] <http://www.dcc.uchile.cl/~cgutierr/cursos/IA/mycin.pdf>.

Referencias Bibliográficas

20. **Szolovits, Peter.** Inteligencia Artificial y Medicina. *Inteligencia Artificial y Medicina*. [En línea] 1982. [Citado el: 16 de enero de 2010.] <http://groups.csail.mit.edu/medg/ftp/psz/AIM82/ch1.html>.
21. **RS, Kaplan.** un prototipo de sistema de dirección basado en la historia de AI. *un prototipo de sistema de dirección basado en la historia de AI*. [En línea] 1991. [Citado el: 16 de enero de 2010.] <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2247609/>.
22. **Maisonneuve, Patrick.** Alcohol aumenta riesgo de padecer cáncer de mama. [En línea] 27 de 09 de 2007. [Citado el: 20 de enero de 2010.] <http://www.terra.com/salud/articulo/html/sal16527.htm>.
23. **Siteman Cancer Center.** Siteman Cancer Center. [En línea] 2004. [Citado el: 20 de enero de 2010.] http://www.yourdiseaserisk.wustl.edu/hccpquiz.pl?lang=spanish&func=show&quiz=ovarian&page=risk_list.
24. **M., Dr Gonzalo E. Díaz.** Ecografía drGdiaz.com. [En línea] 1996. [Citado el: 20 de abril de 2010.] <http://www.drgdiaz.com/eco/prostata/cancerprostatariesgo.shtml>.
25. **Goldstein., Dr. Steven R.** Prevención en el Cáncer de Mama:. [En línea] 2000. [Citado el: 15 de abril de 2010.] <http://latina.obgyn.net/espanol/default.asp?page=/espanol/articulos/Febrero01/mama4>.
26. **M, Dr. Octavio Peralta y Mastología, Sociedad Chilena de.** Alto Riesgo de Cáncer de Mama y Estrategias de Prevención. [En línea] 2010. [Citado el: 15 de abril de 2010.] <http://www.mednet.cl/link.cgi/Medwave/ArticulosRevision/4358>.
27. **Proctor, Mark.** Drools. [En línea] [Citado el: 28 de abril de 2010.] http://downloads.jboss.com/drools/docs/4.0.7.19894.GA/html_single/index.html#d0e2364.
28. **Eugenia, María.** El Modelo Conceptual. [En línea] [Citado el: 28 de abril de 2010.] http://eisc.univalle.edu.co/materias/Material_Desarrollo_Software/MCONCEP_A4.pdf.
29. **Modelo de Clases.** Modelo de Clases. [En línea] [Citado el: 28 de abril de 2010.] <http://www.dcc.uchile.cl/~psalinas/uml/modelo.html>.

BIBLIOGRAFÍA

Zukowski, John. *Programacion Java 2 j2se 1.4 Volumen I , II , III.*

Martinez, Jairo Chapela. *Introducción al entorno de desarrollo Eclipse.*

Jboss Community. *Jboss Community.* <http://www.jboss.org/drools/>

Java-Source.net <http://java-source.net/open-source/rule-engines/openrules>

Proctor, Mark. *Drools.* <http://downloads.jboss.com/drools/docs/4.0.7.19894.GA>

[/html_single/index.html#d0e2364.](http://downloads.jboss.com/drools/docs/4.0.7.19894.GA/html_single/index.html#d0e2364)

Díaz-Faes José. *Cáncer de mama. Avances en diagnosticos, tratamientos e investigacion.*

Pastor E., Soto MA. *Cáncer de mama/Ovario Hereditario*

ANEXOS

Anexo #1 Interface del programa computacional del modelo de Gail.

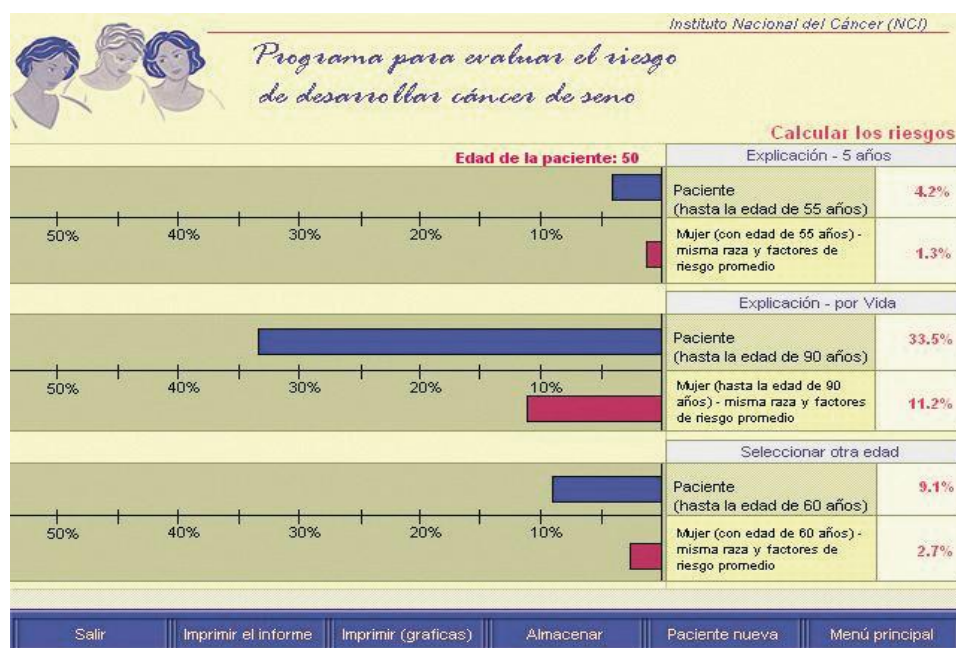
Instituto Nacional del Cáncer (NCI)

Programa para evaluar el riesgo de desarrollar cáncer de seno

Calcular los riesgos

¿Edad de su paciente?	50		Descripción Los cálculos para las mujeres asiáticas o de las Islas del Pacífico están basados en las tasas para las mujeres blancas y son inciertos. Presione Siguiente para continuar.
¿Edad a la que se desarrolló?	12		
¿Edad de la paciente al momento de su primer parto?	23		
¿Cuántas de las familiares directas de la paciente tienen o han tenido cáncer de seno?	1		
¿Ha sido la paciente sometida alguna vez a una biopsia de seno?	Sí		
¿Cuántas biopsias de seno (positivas o negativas) ha tenido la paciente?	1		
¿Ha tenido la paciente al menos una biopsia que haya revelado una hiperplasia atípica?	Sí		
Si conoce la raza de la paciente, favor indicarla.	Asiática /Isla Pacífica		

Presione "Siguiente" para continuar.



Anexo #2 Descripción de los patrones usados en nuestra investigación.

Nombre del Patrón	Problema	Solución
Expert – Experto	¿Cuál es un principio general para asignar responsabilidades a los objetos?	Asignar una responsabilidad al experto en información, la clase que tiene la información necesaria para la realización de la asignación.
Creator – Creador	¿Quién debería ser el responsable de la creación de una nueva instancia de alguna clase?	<p>Asignar a la clase B la responsabilidad de crear una instancia de clase A si se cumple uno o más de los casos siguientes:</p> <ol style="list-style-type: none"> 1. B agrega objetos de A 2. B contiene objetos de A 3. B registra instancias de objetos de A 4. B utiliza más estrechamente objetos de A. 5. B tiene datos de inicialización que se pasarán a un objeto de A cuando sea creado (por tanto, B es un Experto con respecto a la creación de A). 6. B es un creador de los objetos A.
Low Coupling – Bajo Acoplamiento	¿Cómo soportar bajas dependencias, bajo impacto del cambio e incremento de la reutilización?	Asignar una responsabilidad de manera que el acoplamiento permanezca bajo.

High Cohesion – Alta cohesión	¿Cómo mantener la complejidad manejable?	Asignar una responsabilidad de manera que la cohesión permanezca alta.
----------------------------------	---	---

GLOSARIO

A

alasARBOGEN: Sistema informático desarrollado pro la Universidad de las Ciencias Informáticas, el cual se encarga fundamentalmente de la modelación de arboles genealógicos de personas y familias, además de gestionar la información de los individuos que figuran en estos genogramas.

B

BRCA1: Gen que se identificó en 1990 en el cromosoma 17q21, el cual se relaciona con el cáncer de mama y de ovario.

BRCA2: Gen que se identificó en 1994 en el cromosoma 13q12-13, el cual se relaciona con el cáncer de mama y de ovario.

C

CNGM: Centro Nacional de Genética Médica.

Cyrilic: Es el más completo, funcional y fácil de usar software diseñado para el manejo y representación gráfica de datos genéticos en el sistema operativo Windows

S

SE: Sistema experto

SBR: Sistema basado en regla