



**Facultad 2**

**BASE DE CASOS Y MOTOR DE INFERENCIA PARA  
APOYAR EL APRENDIZAJE DEL DISEÑO  
CONCEPTUAL DE BASES DE DATOS**

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

**AUTORES:**

Francisdailin Cobas Sierra

Henry Cable Hernández

**TUTORES:**

Ing. Ariagnis Yero Guevara

Ing. Madelín Haro Pérez

La Habana, 20 de junio 2013

“Año 55 de la Revolución”

**Pensamiento**



*"Lo que sabemos es una gota de agua; lo que ignoramos es el océano."*

*Isaac Newton*

## **Declaración de autoría**

Declaramos que somos los únicos autores de este trabajo y autorizamos a la Universidad de las Ciencias Informáticas y a la Facultad 2 a hacer uso del mismo en su beneficio.

Para que así conste firmamos la presente a los \_\_\_\_ días mes de junio del año 2013

\_\_\_\_\_  
Henry Cable Hernández  
Autor

\_\_\_\_\_  
Francisdailin Cobas Sierra  
Autora

\_\_\_\_\_  
Ing. Ariagnis Yero Guevara  
Tutor

\_\_\_\_\_  
Madelín Haro Pérez  
Tutor

Agradecimientos

*Henry y Francisdailin*

*Agradecer primeramente a nuestras tutoras Ariagnis y Madelín por su  
paciencia y dedicación.*

*A todos los profesores a los cuales pedimos ayuda y consejos, que fueron unas  
cuántas veces, en especial a Dasiel, Héctor, Darian y David,*

*A los profesores de base de datos que nos apoyaron con las entrevistas.*

*A Elaine y Tito que nos brindaron su ayuda.*

*A todos aquellos que de una forma u otra aportaron su granito de arena.*

*Gracias.*

*Francisdailin*

*A mis padres por su entrega, amor y sacrificio.*

*A Jorge Ernesto por impulsarme todos los días a ser mejor persona.*

*A mis abuelos María, Sonia y Flor que a pesar de que hoy no contamos con  
su presencia física de seguro estarían orgullosos de mí.*

*A mi familia por apoyarme siempre.*

*A mi tía Melba por su preocupación y cariño.*

*A Luisa y Jorge por hacerme sentir parte de su familia.*

*A la gente del grupo 2108 y a todos con los que he compartido durante mi  
estancia en la universidad.*

*A mi compañero de tesis por tolerar mis malacrianzas sin perder la paciencia  
ni desanimarse a pesar de las vicisitudes.*

*Henry*

*A mi mamá por haberme guiado en estos años de estudios, por su comprensión  
y ternura.*

*A mi papá por haber compartido conmigo los pocos años que el destino decidió  
que estuviéramos juntos.*

*A Vilo que ha sabido ser un padre para mí, dándome su apoyo en lo que  
necesitara.*

*A mis abuelos maternos por estar atentos siempre a mi progreso en los  
estudios y siempre apoyarme en los momentos difíciles de mi vida*

*A los grupos 2101, 2204, 2205, 2301, 2404, 2504 por haberme acogido en  
su seno, ya que en estos grupos fue donde me formé como profesional.*

*A Lidiagnis que siempre ha sido una gran amiga para mí aconsejándome en  
los momentos difíciles y ayudándome en los estudios.*

*A Dayna, Danima, Leydys y otras amigas que me han ayudado mucho  
también en estos 5 años de carrera.*

*A Mason que me ha ofrecido su amistad y me ha ayudado mucho.*

Dedicatoria

*Francisdailin*

*En especial a mi abuela María.*

*A mis padres.*

*A Ernesto, mi compañero de vida durante estos cinco años.*

*A mi familia y a las personas que me quieren.*

*Henry*

*Muy especialmente a mi papá que aunque no está físicamente siempre estará  
en mi mente y corazón, donde quiera que estés, te quiero.*

*A mi mamá por saber guiarme por buen camino durante estos 23 años, por  
su paciencia y dedicación.*

*A una persona muy especial que llegó a mi vida hace 13 años que ha sabido  
ser un padre para mí, Vilo.*

*A mi hermanita Meli.*

*Y al resto de mi familia que siempre ha deseado lo mejor para mí, a Yaya,*

*Tutín, Lizi, Lili, Sandri, la pequeña pulu.*

## **Resumen**

El diseño de base de datos representa una etapa crucial en el desarrollo de cualquier sistema informático. Por esta razón las carreras que tienen entre sus objetivos el ciclo de vida de un software le atribuyen particular importancia a su estudio.

Actualmente, la carrera de Ingeniería en Ciencias Informáticas implementa un modelo de formación centrado en el aprendizaje donde el estudiante tiene cierto nivel de protagonismo en su propio aprendizaje. En la temática de base de datos en particular se utilizan métodos propios de este modelo y se apoya en herramientas y tecnologías informáticas como el Entorno Virtual de Enseñanza Aprendizaje con actividades de autopreparación y autoevaluación. En contenidos como el diseño conceptual, donde la ejecución de actividades prácticas determina una mejor aprehensión del conocimiento, aún no es suficiente los métodos y medios que se emplean. Por ese motivo se buscan alternativas que permitan lograr la personalización del proceso y la independencia del estudiante.

La presente investigación forma parte de un sistema inteligente para la asignatura Sistemas de Base de Datos 1, dirigido al aprendizaje del tema diseño conceptual de base de datos. En ella se utiliza la técnica de inteligencia artificial Razonamiento Basado en Casos para proponer la base de casos y el motor de inferencia. Como resultados del trabajo se obtiene el estudio del estado del arte sobre sistemas inteligentes para la enseñanza y de bases de datos, el modelo del diseño y la validación de la propuesta.

**Palabras claves:** Base de datos, diseño conceptual de base de datos, sistema inteligente de enseñanza, sistemas de razonamiento basado en casos

**Índice**

INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA.....	5
1.1 Introducción.....	5
1.2 Base de datos. ....	5
1.2.1 Diagrama Entidad-Relación.....	6
1.3 Inteligencia Artificial.....	7
1.3.1 Sistemas basados en el conocimiento. ....	8
1.3.2 Sistema de razonamiento basado en casos. ....	11
1.3.3 Base de casos. ....	15
1.3.4 Motor de inferencia.....	16
1.4 Análisis de soluciones existentes. ....	18
1.4.1 Panorama internacional.....	18
1.4.2 Panorama nacional. ....	20
1.5 Conclusiones parciales.....	21
CAPÍTULO 2: DISEÑO DE LA BASE DE CASOS Y MOTOR DE INFERENCIA.....	22
2.1 Introducción.....	22
2.2 Base de casos.....	23
2.2.1 Técnicas para la adquisición de conocimientos.....	23
2.2.2 Representación del conocimiento.....	24
2.3 Motor de inferencia.....	27
2.3.1 Fase de recuperación. ....	27
2.3.2 Fase de adaptación.....	30
2.3.3 Fase de revisión. ....	31
2.3.4 Fase de retención.....	32
2.4 Conclusiones parciales.....	32
CAPÍTULO 3: VALIDACIÓN DE LA SOLUCIÓN PROPUESTA .....	33
3.1 Introducción.....	33
3.2 Método Delphi. ....	33
3.3 Validación de la base de casos mediante Delphi.....	33
3.3.1 Selección de los expertos.....	34
3.3.2 Elaboración del cuestionario. ....	37
3.3.3 Cálculo de la concordancia de los especialistas.....	38
3.3.4 Desarrollo práctico y explotación de los resultados.....	40
3.3.5 Resultados de la validación de la base de casos propuesta. ....	42
3.4 Validación del motor de inferencia. ....	43
3.5 Conclusiones parciales.....	46
CONCLUSIONES .....	47
RECOMENDACIONES .....	48
TRABAJOS CITADOS .....	49

BIBLIOGRAFÍA.....	52
ANEXOS .....	56
Anexo 1: Modelo de entrevista a expertos para obtener los pesos de los rasgos predictores. ....	56
Anexo 2: Cálculo de los peso de los rasgos utilizando la técnica panel de experto. ....	56
Anexo 3: Modelo de encuesta a profesores para seleccionar expertos .....	58
Anexo 4: Modelo de entrevista a expertos para validar la propuesta de base de casos.....	59
Anexo 5: Matriz de semejanza entre casos para calcular el umbral. ....	60
Anexo 6: Casos semejantes.....	64

## **Introducción**

Las bases de datos son colecciones de datos que pueden ser manejados de manera intuitiva, sencilla y ordenada. Su objetivo es que los datos se manipulen y transformen para obtener información relevante para una organización. De ahí se deriva que las bases de datos se utilicen frecuentemente en el proceso de construcción de sistemas informáticos y que se preste especial atención a su diseño y modelación. Por ello, las carreras que gradúan profesionales que tienen como objeto de trabajo el ciclo de vida de un software y/o la modelación de la información, incluyen en su plan de estudio asignatura(s) relacionada(s) con esta área del conocimiento.

La carrera de Ingeniería en Ciencias Informáticas, de la Universidad de las Ciencias Informáticas, tiene entre sus objetivos, según su Modelo del Profesional [1], graduar profesionales capaces de ejecutar los diferentes roles asociados a la proyección, construcción y mantenimiento de un software; tanto en empresas de producción industrial de software, como en otras organizaciones que desarrollen sus propios software. Para ello, el proceso de formación debe lograr que sus graduados sean capaces de:

- Proyectar, construir y mantener software aplicado utilizando metodologías, métodos, técnicas y herramientas apropiadas.
- Administrar bases de datos

El plan de estudio de la carrera incluye en la Disciplina Ingeniería y Gestión de Software las asignaturas Sistema de Base de Datos 1 y 2 organizadas para impartirse en el cuarto y quinto semestre respectivamente. Estas materias se imparten teniendo en cuenta la precedencia con otras asignaturas que aportan en la comprensión y el aprendizaje de las bases de datos puesto que desarrollan habilidades como el pensamiento lógico y algorítmico y la solución de problemas. Entre ellas se encuentran Matemática Discreta, Introducción a la Programación y Programación 1. A su vez, en el mismo 2do año se imparte Programación 2 y 3, esta última impartida en paralelo con Sistema de Base de Datos 1.

Dentro de los temas de Sistema de Base de Datos 1 se encuentra el diseño de base de datos. Este abarca la obtención del modelo entidad-relación, el modelo relacional y su normalización. Su objetivo principal es lograr una base de datos que almacene la información sin redundancia y permita una recuperación rápida y eficiente de los datos. Dicha temática se imparte en modalidad presencial utilizando las formas organizativas conferencia, clase práctica y laboratorio, las dos últimas, potenciando el desarrollo de habilidades prácticas y técnicas en el estudiante, elemento muy necesario en asignaturas con estas características.

El curso está publicado en el Entorno Virtual de Enseñanza Aprendizaje. En él están disponibles orientaciones para el estudio de la asignatura, ejercicios de autoevaluación, bibliografía, entre otros materiales y recursos que propician su estudio. En el aula, a su vez, el profesor utiliza métodos

identificados con el modelo de enseñanza vigente en la universidad y medios como casos de estudio, videos, sistemas de bases de datos, entre otros. Un método muy aceptado por la mayoría de los estudiantes es la resolución de ejercicios en conjunto con el profesor pero tiene la desventaja de que el educador no está siempre disponible y el estudiante crea una dependencia no deseada. Más allá del esfuerzo de ambas partes: profesores y estudiantes, por explicar y comprender todo el tema, en la mayoría de los casos las dudas y preguntas de interés surgen en momentos que están fuera de los horarios de clase establecidos, por lo tanto, se debe esperar el próximo encuentro.

Por otra parte los informes semestrales [2-4] de la asignatura Sistema de Bases de Datos 1 emitidos por el Departamento Metodológico Central de Ingeniería y Gestión de Software, durante los últimos tres cursos, reflejan los resultados obtenidos por los estudiantes en diferentes evaluaciones. Los resultados durante los cursos desde el 2009 hasta la fecha tienen un porcentaje de promoción del 68,49 y de calidad de 27,83. Estos no son satisfactorios teniendo en cuenta la importancia demostrada de la asignatura para la carrera. A continuación se especifican los principales elementos que intervienen en estos resultados.

Por parte de los profesores:

- Deficiencias en la realización de actividades docentes y científico metodológicas.
- Claustro de profesores en su mayoría joven, sin mucha experiencia, que se enfrenta a la asignatura por primera o segunda vez.

Por parte de los estudiantes:

- La falta de sistematicidad y de responsabilidad en el estudio.
- Insuficiente desarrollo de habilidades y de formación en asignaturas precedentes.
- Insuficiente independencia en la búsqueda de conocimientos en el auto-estudio y auto-aprendizaje.
- Poca motivación e interés para realizar las tareas orientadas a realizar de manera independiente.
- Problemas de contenido en la elaboración del diseño conceptual y en las consultas realizadas en álgebra, cálculo y SQL.

Actualmente la universidad cuenta con pocos medios que permitan a los estudiantes, una vez impartidos determinados contenidos, medir sus conocimientos. Los recursos de aprendizaje existentes, a pesar de determinar las deficiencias de los estudiantes, no son capaces de orientarlos y proponer soluciones a sus carencias cognitivas.

Por lo expuesto anteriormente y centrando la atención en el diseño conceptual de bases de datos, se define como **problema a resolver**: ¿Cómo contribuir al proceso de aprendizaje del diseño conceptual de bases de datos fomentando la autopreparación de los estudiantes y el carácter práctico del contenido?

Para darle solución al problema anterior se propone como **objeto de estudio**: Sistemas inteligentes de apoyo al proceso de aprendizaje.

Para dar solución al problema planteado se determina como **objetivo general**: Proponer una base de casos y su motor de inferencia como parte de un sistema inteligente de apoyo al proceso de aprendizaje del diseño conceptual de bases de datos para la carrera de Ingeniería en Ciencias Informáticas. De esta manera el **campo de acción** queda enmarcado en los sistemas inteligentes de apoyo al proceso de aprendizaje del diseño conceptual de bases de datos. Para dar cumplimiento al objetivo de la investigación se plantean los siguientes **objetivos específicos**:

- Elaborar el marco teórico de la Investigación.
- Modelar la propuesta de solución donde se aplique la técnica de inteligencia artificial sistema basado en el conocimiento afín con el diseño conceptual de bases de datos.
- Validar la solución propuesta.

#### **Tareas investigativas para dar cumplimiento a los objetivos:**

- Estudio de los conceptos relacionados con el diseño conceptual de base de datos.
- Análisis de las diferentes técnicas de inteligencia artificial que apoyen el proceso de toma de decisiones utilizando estándares definidos y la experiencia acumulada.
- Caracterización de la situación existente en el mundo y el país concerniente a la enseñanza del diseño de bases de datos, inteligencia artificial y la técnica seleccionada por los investigadores.
- Selección de técnicas de inteligencia artificial basadas en la acumulación de experiencias y estándares definidos que permitan la enseñanza de conocimientos sobre el diseño conceptual de bases de datos.
- Análisis de las características y particularidades de los motores de inferencia o mecanismos de control.
- Selección de los rasgos a tener en cuenta en la enseñanza del diseño conceptual de las bases de datos.
- Determinación de la forma de representación, organización y selección de los casos.
- Selección de la función de distancia a emplear para medir la semejanza entre los casos.
- Diseño e implementación del algoritmo para el aprendizaje automático del razonador del sistema basado en casos.
- Selección y aplicación de un método que permita validar la propuesta de solución.

Durante la investigación se utilizan los siguientes **métodos científicos**:

#### **Métodos Teóricos.**

- **Analítico – sintético**: A partir de la documentación encontrada se seleccionan los elementos claves que permiten comprender y diseñar de manera exitosa la base de casos y el motor de inferencia.

- **Método Inductivo – Deductivo:** A partir de la documentación y bibliografía consultada se seleccionan las técnicas y algoritmos que permiten el desarrollo de la base de casos y el motor de inferencia.

#### **Métodos Empíricos.**

- **Entrevista:** Las entrevistas a los profesores de la asignatura ayudan a comprender y definir los elementos que caracterizan el proceso de enseñanza-aprendizaje de las bases de datos y los problemas y aspectos fundamentales.
- **Técnicas de validación:** Se utiliza para la validación de la base de casos el método Delphi, basado en el criterio de los expertos.
- **Documental:** La revisión documental permite realizar la extracción de conocimiento de diversas fuentes implicadas (libros, artículos de revistas, manuales, informes y documentos de ayuda), en formato digital o impreso. Se analizan y estudian documentos utilizados para impartir el tema de diseño de base de datos.

Se prevé que el **alcance de la investigación** sea: Diseño de una base de casos y motor de inferencia como parte de un sistema inteligente que permita apoyar el aprendizaje de los estudiantes durante el proceso de aprehensión del conocimiento perteneciente a la temática del diseño conceptual y apoyar el aprendizaje de manera personalizada.

Se prevé que **los resultados de la investigación** sean:

- La fundamentación de un sistema de razonamiento basado en casos.
- Obtención del diseño de la base de casos y el motor de inferencia.

El presente documento consta de cuatro partes fundamentales: resumen, introducción, desarrollo y conclusiones. El desarrollo está estructurado en tres capítulos los cuales están organizados de la siguiente manera:

**Capítulo 1: Fundamentación Teórica.** En este capítulo se realiza un estudio detallado de las técnicas a utilizar para el diseño de la base de casos y el motor de inferencia así como de los conceptos fundamentales a los que se hará referencia a lo largo de toda la investigación.

**Capítulo 2: Diseño de base de casos y motor de inferencia.** En este capítulo se definen los elementos que conforman el razonamiento basado en casos. Se precisa la estructura de los casos así como su almacenamiento. Se definen los algoritmos correspondientes a las distintas fases del ciclo de vida del razonamiento basado en casos.

**Capítulo 3: Validación de la Propuesta.** En este capítulo se definen los métodos para realizar la validación de la propuesta. Para la base de casos se aplica el método Delphi y para el motor de inferencia una aplicación desarrollada por los investigadores. Luego se aplican los métodos definidos y se llega a conclusiones a partir de los resultados obtenidos.

## Capítulo 1: Fundamentación Teórica.

### 1.1 Introducción.

En el presente capítulo se exponen los conceptos principales relacionados con el diseño de base de datos y un estudio del estado del arte de sistemas relacionados con las bases de datos. Así mismo se analizan otros que presentan un comportamiento inteligente, algunos de estos utilizan como técnica de inteligencia artificial los sistemas basados en el conocimiento para manejar la experiencia. Por último se presenta una investigación de la técnica de inteligencia artificial, sistemas basados en el conocimiento.

### 1.2 Base de datos.

Para comprender mejor qué es el diseño conceptual de base de datos se hace necesario definir un conjunto de conceptos fundamentales. Estos se mencionan a continuación:

**Base de datos:** Es un conjunto de datos interrelacionados entre sí, almacenados con carácter más o menos permanente en la computadora. O sea, que una base de datos puede considerarse una colección de datos variables en el tiempo.[5] Otra definición más estricta puede ser la siguiente:

Una base de datos es un conjunto de datos que tiene las siguientes propiedades implícitas:[6]

- Representa algún aspecto del mundo real, llamado mini mundo o universo de discurso. Las modificaciones del mini mundo se reflejan en la base de datos.
- Es un conjunto de datos lógicamente coherentes, con un cierto significado inherente. Una colección aleatoria de datos no puede considerarse propiamente una base de datos.
- Una base de datos se diseña, construye y puebla con datos para propósito específico. Está dirigida a un grupo de usuarios y tiene ciertas aplicaciones preconcebidas que interesan a distintos usuarios.

El diseño de una base de datos contiene las siguientes etapas:[7]

- **Diseño conceptual:** En esta etapa se obtiene una estructura de la información de la futura base de datos independiente de la tecnología que hay que emplear. Esta se enfoca únicamente en la estructuración de la información.
- **Diseño lógico:** Éste parte del resultado del diseño conceptual, que se transforma de forma que se adapte a la tecnología que se debe emplear. Más enfocado a las características del Sistema Gestor de Base de Datos (SGBD) que se vaya a utilizar.
- **Diseño físico:** En esta etapa se transforma la estructura obtenida en la etapa del diseño lógico, con el objetivo de conseguir una mejor eficiencia, además, se completa con aspectos de implementación física que dependerán del SGBD.

La presente investigación aborda el diseño conceptual de base de datos. Para comprenderlo mejor se definen los elementos y conceptos que lo describen. Estos se mencionan a continuación:

**Modelo de datos:** Es un una colección integrada de conceptos, para describir y manipular datos, las relaciones existentes entre los mismos y las restricciones aplicables a los datos, todo ello dentro de una organización. Estos se pueden clasificar como:[8]

- Basado en objetos.
- Basados en registros.
- Físicos.

Los primeros dos se utilizan para describir los datos a nivel conceptual y externo, mientras que el último se emplea para describir los datos a nivel interno. En cuanto a los modelos de datos basados en objetos, en el cual se utilizan conceptos tales como entidades, atributos y relaciones, existen cuatro tipos:[8]

- Entidad-Relación.
- Semántico.
- Funcional.
- Orientado a objetos.

El modelo entidad-relación amplía la definición de entidad para incluir no solo los atributos que describen el estado del objeto, sino que también las asociaciones con el objeto, su comportamiento. En este caso se dice que el objeto encapsula tanto el estado como el comportamiento.[8] Es relevante para la investigación el modelo entidad-relación ya que este es una manera de representar el diseño conceptual de base de datos, el cual se profundiza en el siguiente sub-epígrafe.

### 1.2.1 Diagrama Entidad-Relación.

Un Diagrama o Modelo Entidad Relación (a veces denominado por su siglas, E-R "Entity Relationship", o, "DER" Diagrama Entidad Relación), es una herramienta para el modelado de datos de un sistema de información. Estos modelos expresan entidades relevantes para un sistema de información así como sus interrelaciones y propiedades. Propuesto por Peter Chen en 1976, mediante el mismo se pretenden 'visualizar' los elementos que pertenecen a una base de datos, que reciben el nombre de entidades, las cuales se corresponden con el concepto de clase de la Programación Orientada a Objeto y donde cada tupla de una futura relación representaría un objeto de la Programación Orientada a Objetos.[9]

#### Principales elementos del diagrama entidad-relación.

Hay tres nociones básicas que emplea el modelo de datos E-R: conjuntos de entidades, conjuntos de relaciones y atributos. Estos se describen a continuación.

- Una **entidad** es una «cosa» u «objeto» en el mundo real que es distinguible de todos los demás objetos. Una entidad tiene un conjunto de propiedades, y los valores para algún conjunto de propiedades pueden identificar una entidad de forma unívoca. Una entidad puede ser concreta, como una persona o un libro, o puede ser abstracta, como un préstamo, unas vacaciones o un concepto. Una entidad se representa mediante un conjunto de atributos. La designación de un atributo para un conjunto de entidades expresa que la base de datos almacena información similar concerniente a cada entidad del **conjunto de entidades**; sin embargo, cada entidad puede tener su propio valor para cada atributo. Un **conjunto de entidades** es un conjunto de entidades del mismo tipo que comparten las mismas propiedades, o atributos.[10]
- Una **relación** es una asociación entre diferentes entidades. Un **conjunto de relaciones** es un grupo de relaciones del mismo tipo. Formalmente es una relación matemática con  $n > = 2$  de conjuntos de entidades (posiblemente no distintos).[10]
- **Campo o atributo:** Es la unidad menor de información sobre un objeto y representa una propiedad de un objeto. Los atributos describen propiedades que posee cada miembro de un conjunto de entidades. [5]

De esta manera queda definido el diseño conceptual de base de datos, el cual será representado mediante el diagrama entidad-relación.

### 1.3 Inteligencia Artificial.

El amplio campo que se conoce como inteligencia artificial trata aquellos problemas que hace un tiempo se consideraban imposibles de realizar por las máquinas. Puesto que su formulación y resolución requerían ciertas habilidades solo presentes en los seres humanos. Estos problemas en un principio parecían imposibles, intratables y difíciles de formular utilizando ordenadores. Sin embargo, el trabajo realizado en las tres últimas décadas por investigadores procedentes de varios campos, muestra que muchos de estos problemas pueden ser formulados y resueltos por máquinas. Bar y E. A. Feigenbaum, dos de los pioneros de la investigación en inteligencia artificial, definen esta como sigue: [11]

“La inteligencia artificial es la parte de la ciencia que se ocupa del diseño de sistemas de computación inteligentes, es decir, sistemas que exhiben las características que asociamos a la inteligencia en el comportamiento humano que se refiere a la comprensión del lenguaje, el aprendizaje, el razonamiento, la resolución de problemas, etcétera.” [11]

Hoy en día, el campo de la inteligencia artificial engloba varias sub-áreas tales como los sistemas expertos, la demostración automática de teoremas, el juego automático, el reconocimiento de la voz y de patrones, el procesamiento del lenguaje natural, la visión artificial, la robótica, las redes neuronales, etc. El presente trabajo está dedicado a los sistemas expertos. Aunque los sistemas expertos constituyen una de las áreas de investigación en el campo de la Inteligencia Artificial, la

mayor parte de las restantes áreas, si no todas, disponen de una componente de sistemas expertos formando parte de ellas. [11]

### 1.3.1 Sistemas basados en el conocimiento.

Los sistemas basados en el conocimiento no son más que programas para computadoras que simulan las cadenas de razonamiento que realiza un experto para resolver un problema de su dominio. Para conseguirlo, se dota al sistema de un conjunto de principios o reglas que infieren nuevas evidencias a partir de la información previamente conocida. En términos generales, un sistemas basados en el conocimiento puede ser definido como un sistema que utiliza el conocimiento acumulado sobre un dominio en específico para obtener resultados y llegar a conclusiones en problemas del mismo dominio.[12] Estos sistemas presentan ventajas y desventajas, estas se mencionan a continuación.

Ventajas: [13]

- **Mayor disponibilidad, accesibilidad y permanencia:** La experiencia se encuentra disponible todo el tiempo al almacenarla en un mismo sistema. Esto permite poder consultarla en cualquier momento.
- **Respuestas no subjetivas:** El sistema basado en conocimientos ofrece respuestas sólidas y completas que no se ven afectadas por la apreciación personal que pudiera representar una mala conclusión, como indicadores que se obviarán o el asumir valores irreales.
- **Tutoría inteligente:** Asegura que se den respuestas consistentes y que el conocimiento crezca en experiencia con el transcurso del tiempo, mejorando la calidad de sus respuestas sin necesitar inferir un nuevo conocimiento por parte de los expertos.
- **Preservación de la experticia:** Los sistemas basados en el conocimiento constituyen una memoria asociativa y poseen la capacidad para adquirir nuevo conocimiento y perfeccionar el que poseen.

Desventajas: [13]

- La respuesta no siempre es la correcta.
- Conocimiento limitado al dominio de experticias
- Ausencia de sentido común.
- No reconocen el límite de su conocimiento.

### Arquitectura de los sistemas basado en el conocimiento.

Los sistemas basados en el conocimiento pueden ser diferenciados del resto de las técnicas de la IA ya que en primer lugar, ejecutan tareas que de alguna forma resultan complicadas o simplemente difíciles para cualquier ser humano y que están reservadas solamente a unos pocos en un dominio en particular. Para ejecutar estas tareas o resolver los problemas que solucionan los expertos, los sistemas basados en el conocimiento utilizan estrategias de búsqueda que comúnmente se denominan heurísticas. Además, emplean conocimientos para razonar acerca de sus propios

métodos de inferencia y por último proporcionan explicaciones o justificación de la solución obtenida.[14] Los elementos fundamentales que componen la arquitectura de los sistemas basados en el conocimiento son: [14]

- Interfaz.
- Máquina o MI.
- Base de Conocimientos.

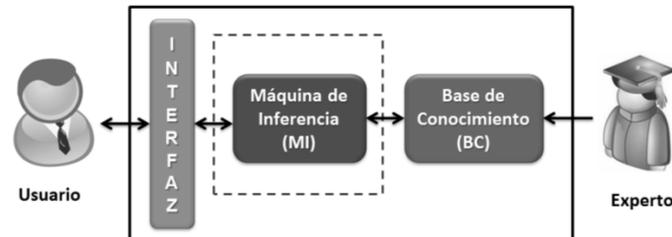


Figura 1: Arquitectura de los sistemas basados en el conocimiento

La interfaz es el componente que facilita la interacción con el sistema. La máquina de inferencia es quién realiza la obtención y transformación del conocimiento previo almacenado en la base de conocimientos por un experto, para dar solución a un nuevo problema planteado. [14]

### Tipos de sistemas basados en el conocimiento.

Entre los sistemas basados en el conocimiento se encuentran:

- **Sistemas de Razonamiento Basados en Reglas (SRBR):** Los sistemas expertos basados en reglas utilizan para el proceso de inferencia un conjunto de reglas que constituyen la base de conocimiento del experto. Este conjunto de reglas pueden ser activadas a medida que las condiciones son evaluadas positivamente y su utilización implica la creación de nuevos hechos. Este proceso permitirá a partir de unos hechos iniciales desarrollar un proceso deductivo que concluirá en el momento en que no quede ninguna otra regla por utilizar. Para realizar este tipo de tratamiento es posible hacerlo de dos maneras diferentes, por un lado realizarlo desde las evidencias hasta los objetivos o por otro lado en orden inverso que sería comenzar desde el objetivo hasta llegar al conjunto de evidencias que lo han provocado.[15] A continuación se muestran algunas ventajas e inconvenientes que presentan estos sistemas:

#### Ventajas:[15]

- ✓ Modularidad: facilita el desarrollo incremental.
- ✓ Uniformidad: facilita su manipulación.
- ✓ Naturalidad.
- ✓ Carácter declarativo.

#### Desventajas: [15]

- ✓ Ineficiencia: puede mejorarse el método descrito.
- ✓ Opacidad: es difícil seguir la traza de las deducciones.

- ✓ Cobertura del dominio: gran número de reglas.
- ✓ Dificultad de verificación de consistencia y completitud.
- **Sistemas de Razonamiento Basados en Frames (SRBF):** Un frame<sup>1</sup> es una estructura de datos compleja que contiene un agregado de información acerca de un objeto, ofreciendo una representación estructurada de este objeto. Este sistemas basados en el conocimiento aunque representa una mejora con respecto a SRBR, ya que el conocimiento está significativamente bien estructurado y organizado, permitiendo al sistema rápidamente dirigir la atención a los aspectos apropiados, cuando aparecen nuevos objetos no tiene prototipos nuevos para guiar la representación de estas instancias, inconveniente que imposibilita almacenar nuevos indicadores de proyectos en caso de requerirse para actualizar el sistema y este no llegue a ser obsoleto en algún momento. En los SRBF no se aplican mecanismos de aprendizaje automático razón por la cual no aumentan su conocimiento. [16]
- **Sistemas de Razonamiento Basados en Probabilidades (SRBP):** Este método es utilizado mayormente por sistemas donde la naturaleza de las inferencias sean probabilísticas, como lo son: inferencias por diagnóstico, inferencias causales, inferencias inter-causales e inferencias mixtas. También se utiliza en sistemas donde existe una falta de formalismo para representar cierta clase de conocimiento, o donde la información provenga de múltiples fuentes, algunas fiables y otras no tanto, o conjuntos cuyas cotas no estén definidas rigurosamente.[16]

A continuación se mencionan ventajas y desventajas de este tipo de sistemas basados en el conocimiento.

**Ventajas:[16]**

- ✓ Permiten manejar situaciones donde exista incertidumbre.
- ✓ Facilita la interpretación y obtención de conclusiones sobre el dominio en estudio por parte de las personas que lo analiza.
- ✓ Permite combinar conocimiento experto con datos (dicho conocimiento experto generalmente viene dado en forma de relaciones de causalidad).
- ✓ Permiten definir modelos y utilizarlos tanto para hacer razonamiento de diagnóstico como para hacer razonamiento predictivo.

**Desventajas:[16]**

- ✓ No son factibles para todo tipo de dominio, pues se dificulta construir las redes con ayuda de expertos humanos cuando existen carencias de conocimiento.
- ✓ No son viables para explicar el razonamiento, ya que los métodos y modelos que utiliza están aún lejos de ofrecer explicaciones comprensibles.
- **Sistemas de Razonamiento Basados en Casos (SRBC):** Representa un nuevo método para resolver problemas no estructurados, en el cual el razonamiento se realiza a partir de una memoria asociativa que usa un algoritmo para determinar una medida de semejanza entre dos objetos. En este paradigma la base del comportamiento inteligente de un sistema radica en recordar situaciones similares existentes en el pasado. Los SRBC presentan un

---

<sup>1</sup>Marco

conjunto de ventajas que al ser explotadas ofrecerán una buena visión para mejorar soluciones alcanzadas, ya que utiliza experiencias anteriores que han sido almacenadas en forma de ejemplos concretos, para llegar a nuevas soluciones o anticipar problemas. Su aprendizaje es incremental pues asimila nuevos ejemplos a medida que entrena su conocimiento sin la necesidad de un experto y ayuda a resolver problemas de forma más rápida que al hacer el análisis del problema de forma trivial por su capacidad en derivar sus respuestas a partir de otras. [16]

Luego de un análisis de las características de los diferentes sistemas basados en el conocimiento se concluye que no resulta adecuado desarrollar un SRBF pues tienen la deficiencia de que no son capaces de aumentar su base de conocimiento. Lo mismo sucede con los SRBR ya que resulta muy engorros descomponer el conocimiento del dominio y generalizarlo en reglas. También es complejo implementar un sistema probabilístico, debido a que se basa fundamentalmente en fórmulas probabilísticas. Se propone utilizar un SRBC por su capacidad de generar respuestas a partir de los conocimientos acumulados y de aumentar de manera automática su base de conocimiento. De los tipos de sistemas expertos existentes, el SRBC es el que más se asemeja al modo de pensar que tienen los seres humanos.

### 1.3.2 Sistema de razonamiento basado en casos.

Los sistemas de razonamiento basados en casos (SRBC) intentan encontrarle solución a nuevos problemas a partir de experiencias anteriores de un dominio específico. El proceso que realiza es análogo al razonamiento humano. Particularmente en el diseño de bases de datos se pone de manifiesto cuando un estudiante realiza el Diagrama Entidad-Relación de una nueva situación problemática.

Comienza cuando este busca en su memoria soluciones que presenten características similares a la actual, a partir de aquí se realizan las transformaciones necesarias para su adaptación al nuevo escenario y la obtención de una nueva solución, habilidad que se adquiere a través de la resolución de problemas de menor y mayor complejidad donde se utilizan conocimientos básicos del diseño conceptual de base de datos. Este proceso se realiza en cuestiones de segundos de manera intuitiva. A partir de esta nueva solución diferente a las demás se procede al enriquecimiento de la experiencia.

La elaboración de un sistema que emplea el RBC presenta dos problemas principales: saber cómo almacenar la experiencia de manera que esta pueda ser recuperada de forma adecuada y poder utilizar la experiencia previa para resolver un problema actual. El RBC presenta ventajas y desventajas, las cuales son:

#### **Ventajas:** [17]

- Tienen el dominio delimitado.
- Utilizan la experiencia para generar la solución a los casos nuevos.
- Adquisición de conocimiento.

- Permite proponer soluciones a problemas rápidamente.
- Propone soluciones en dominios no entendidos completamente por el sistema.
- Ofrece un medio de evaluación de soluciones cuando no se cuenta con un método algorítmico.
- Se centra en las características o partes más importantes del problema.
- El razonamiento basado en casos es aplicable a un amplio rango de problemas.
- El enfoque que utilizan los sistemas basados en casos para la adquisición de conocimiento; pues razonan desde episodios específicos, lo cual evita el problema de descomponer el conocimiento del dominio y generalizarlo en reglas.
- Otras de las ventajas de este tipo de sistema está dada en la flexibilidad para representar el conocimiento a través de los casos, la organización de la base de casos y de las estrategias de recuperación y adaptación de los casos.
- Otra ventaja es el rechazo de las soluciones previas al resolver un problema.

**Desventajas: [17]**

- Confía ciegamente en los casos previos almacenados en su memoria para intentar proponer su solución.
- Puede ser que no recupere el caso más apropiado para la solución del nuevo caso.
- El sistema no explora todo el espacio de soluciones.
- Requiere de una base de datos considerablemente grande y bien seleccionada.
- El RBC depende de una adecuada función de semejanza.

**Arquitectura y Componentes.**

Los SRBC contienen dos componentes fundamentales para su funcionamiento una base de casos y un componente encargado de solucionar los problemas. La base de casos es quien contiene todas las descripciones que han sido obtenidas previamente. En el ciclo de solución del problema se recupera un caso semejante al nuevo y la solución del problema que fue recuperado es propuesta como solución potencial del nuevo problema, luego el nuevo caso obtenido podrá formar parte de la base de casos. Esto se deriva de un proceso de adaptación en el cual se indexa la vieja solución a la nueva situación. El funcionamiento del RBC comprende cuatro actividades principales [18]. Estas cuatro actividades se muestran en la Figura 2:

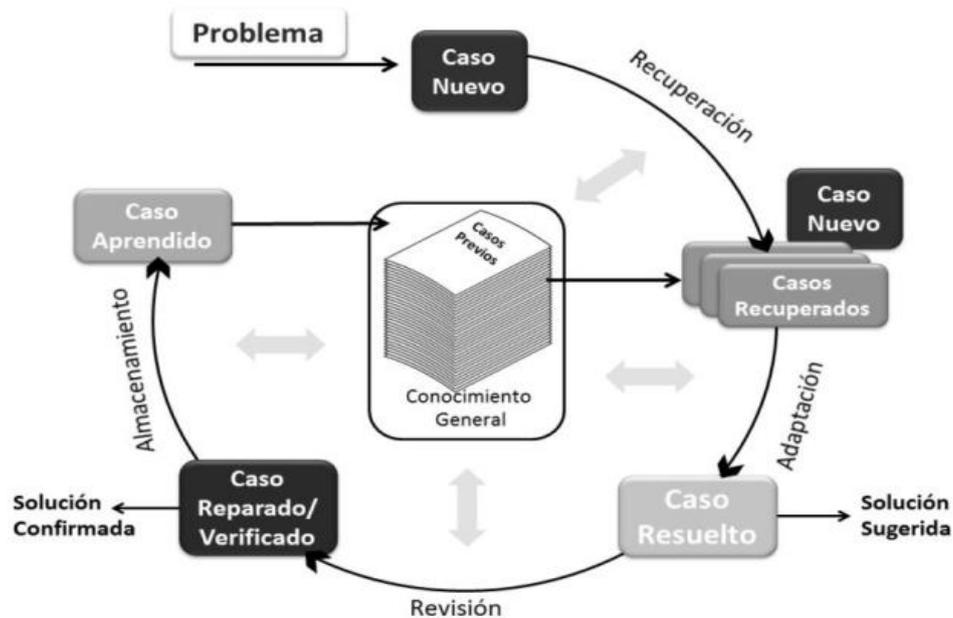


Figura 2: Ciclo básico de un sistema de RBC

Según Aamodt en su trabajo Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches [18] las cuatro fases del razonamiento basado en casos se describen como:

**Recuperación de casos:** Un caso está formado por varios atributos que dan una descripción del problema y una solución para el caso. Para poder llevar a cabo la tarea de recuperación es necesario tener un algoritmo de recuperación y una medida de similitud que permitan obtener un conjunto de casos similares.

**Reutilización de casos:** Tarea centrada en dos aspectos: las diferencias entre el caso pasado y el actual, y qué parte o partes del caso recuperado pueden ser transferidas al nuevo caso. En algunos casos la tarea de reutilización se reduce a copiar la solución pasada al nuevo caso, pero en otros casos esta solución no puede ser aplicada directamente y tiene que ser adaptada.

**Revisión de casos:** Una vez analizada la tarea de reutilización, la solución del nuevo problema tiene que ser probada; este proceso de prueba se hace durante la tarea de revisión. En este paso, la solución generada en la tarea de reutilización se evalúa, y si el resultado es satisfactorio, el nuevo caso y la nueva solución para el caso se almacenan.

**Almacenamiento de casos:** El almacenamiento es la última tarea del ciclo. En este paso el nuevo caso y la solución para este caso (obtenida en la fase de reutilización) son almacenados con vistas a un posible uso futuro. Durante este proceso de aprendizaje, el sistema tiene que seleccionar qué información del caso almacenar, la forma en la que almacenar esta información y cómo indexar el caso en la estructura de memoria para una posterior recuperación.

La calidad de un SRBC está dada por: [19]

- La experiencia que tiene.
- La habilidad para entender situaciones nuevas en términos de experiencias pasadas.
- Su capacidad de adaptación.
- Su habilidad para integrar nuevas experiencias en su memoria adecuadamente.

Para desarrollar un sistema de este tipo, se tiene que tener en cuenta la forma de almacenar la experiencia, de modo que pueda ser recuperada correctamente y lograr que esta sea utilizada en un nuevo problema. También es muy importante definir bien la función de semejanza, encontrar una estructura apropiada para representar el contenido de un caso y decidir cómo la memoria de casos debe ser organizada para un almacenamiento, recuperación y rehúso efectivo.

### **Tipos de sistemas de razonamiento basado en casos.**

Según Rafael Bello y otros en el trabajo Aplicaciones de la Inteligencia Artificial [20] existen dos tipos de sistemas con razonamiento basado en casos: sistemas interpretativos y sistemas resolvedores de problemas. En el estilo de solución de problemas se derivan las soluciones a los nuevos problemas usando las soluciones dadas a los problemas semejantes viejos. En los sistemas interpretativos las situaciones nuevas se evalúan en el contexto de las situaciones viejas.

- **Sistemas interpretativos:** La interpretación se usa cuando el problema no está bien comprendido y cuando hay necesidad de criticar una solución. Estos sistemas toman una situación o solución como entrada y su salida es un argumento fundamentando la solución. Son útiles para sistemas de clasificación, evaluación de una solución, argumentación, justificación de una solución, interpretación y predicción de los efectos de una decisión o plan.
- **Sistemas resolvedores de problemas:** El estilo de solución de problemas se caracteriza por una fuerte componente del sistema para realizar el proceso de adaptación mediante el cual se construye la solución del nuevo problema a partir de la solución dada a un problema semejante. Este es el modelo de razonamiento basado en casos más empleado para resolver tareas de diseño y planificación. Un sistema con razonamiento basado en casos se considera un resolvedor de problemas si:
  - ✓ Puede aprender de los errores, o sea, no repite el mismo error dos veces.
  - ✓ Trata de encontrar una solución mejor aun cuando ya esté disponible una.
  - ✓ No considerara utilizables los resultados inmediatamente, primero los modificará para adecuarlos a las necesidades actuales.

Para darle solución a la problemática se propone diseñar un sistema de razonamiento basado en caso del tipo resolvedor de problemas. Un sistema de este tipo es aplicable cuando una simple presentación de casos resueltos con anterioridad no es suficiente para encontrar una solución; se necesita modificar el caso (la solución dada al caso) para adecuarlo al nuevo problema. El proceso de modificación puede incluir la eliminación, adición o cambio de partes.

### 1.3.3 Base de casos.

Los casos modelan el proceso de solución de problemas representando episodios de este proceso. Un caso es un conjunto arbitrario de rasgos usados para describir un concepto particular.[21] Este está constituido por dos elementos generales: los rasgos predictores y los rasgos objetivos. Los rasgos predictores describen el problema y los objetivos la solución al problema.

#### Forma de almacenar los casos.

Los casos pueden representar distintos tipos de conocimiento y pueden ser almacenados en diferentes formatos. La información almacenada en un caso, debe ser importante tanto para el propósito del sistema como para asegurar que siempre será elegido el caso más apropiado para solucionar un nuevo problema. El formato en el que se almacenan los rasgos predictores permite determinar qué algoritmos utilizar para la recuperación acertada de los casos. Se puede elegir la implementación adecuada dependiendo del tipo de información a representar, teniendo en cuenta que no siempre se necesitan almacenar todos los casos existentes, en su lugar se sigue un criterio para decidir qué casos almacenar y cuáles descartar

Para almacenar los casos debe usarse una memoria o base de casos que posea propiedades similares a la memoria humana, tales como [22]:

- Ser ilimitada.
- En la medida en que la memoria crezca no puede hacerse más lenta.
- Debe permitir buscar directamente los elementos de memoria que sean relevantes para un problema.

#### Estructura para representar los casos.

La estructura de la base de casos puede definirse de dos maneras [20]:

- **Atributo - valor:** En esta estructura la similitud entre los casos de la base de casos y el nuevo caso sin resolver se determina comparando los valores de cada par de rasgos pertenecientes a ambos.
- **Estructurado:** Esta se basa en la similitud estructural, por ende, para recuperar un caso se necesita comparar los rasgos del nuevo caso sin resolver con los rasgos pertenecientes a la base de casos a través de las inferencias entre rasgos para llegar a la solución. Dentro de los modelos que permiten estructurar la base de casos se destacan los grafos, los árboles de decisión y las redes semánticas.

#### Modelos de memoria para organizar la base de casos.

La base de casos debe organizarse de manera que los casos sean recuperados de manera rápida. Con este propósito existen dos estructuras de memoria: la plana y la jerárquica.

En la estructura plana los casos se almacenan secuencialmente en una lista simple, un arreglo o un fichero. Esta estructura tiene como ventaja que la inserción de los casos en la memoria es poco

costoso y a diferencia de las memorias jerárquicas no almacena datos auxiliares para realizar la búsqueda minimizando el espacio que ocupan los casos. Además el procesamiento actual de las tecnologías permite que el crecimiento de la base de casos no se vea afectado en el rendimiento de la aplicación. [22]

En la estructura jerárquica los casos se agrupan en categorías para reducir el número de casos a buscar en una consulta. Se pueden encontrar de diferentes maneras: [22]

- **Redes de características compartidas.** Los casos se almacenan en un árbol o en un grafo acíclico directamente. Las características comunes ocupan nodos del grafo de donde cuelgan los casos que las comparten. La mayoría de los sistemas tienen alguna especie de umbral relativo al número de casos que debe compartir el valor de una característica, de modo que se justifique su posición en un nodo.
- **Redes de discriminación con prioridades.** En las redes de discriminación con prioridades cada nodo contiene una pregunta para la cual los sub-nodos correspondientes ofrecen respuestas alternativas. Las preguntas más importantes se formulan primero, situándose más arriba en la jerarquía. La formulación de preguntas puede implementarse más eficientemente que el emparejamiento en cada sub-nodo.
- **Redes de discriminación redundante o modelos de memoria dinámica.** El modelo de memoria dinámica fue desarrollado a partir de la teoría general de paquetes de organización de memoria, donde la memoria de casos es una estructura jerárquica de episodios generalizados (EG). Un EG contiene tres tipos de objetos: normas, casos e índices. Las normas son características comunes a todos los casos indexados bajo un EG. Los casos, que hacen referencia a la situación planteada. Los índices, que son las características que diferencian a los casos de un EG.
- **Modelo categorías-ejemplares.** La memoria se compone de una red de estructuras y categorías, semánticas, relaciones, casos y punteros índices. Cada caso se asocia a una categoría. Un índice puede apuntar a un caso o a una categoría. La búsqueda de casos se realiza recorriendo las características del problema hasta obtener los casos o categorías con mayor semejanza. La inserción de casos se realiza buscando la categoría o categorías donde debe insertarse.

#### 1.3.4 Motor de inferencia.

El motor de inferencia es el componente encargado de manejar y controlar lógicamente el manejo y utilización del conocimiento almacenado en la base de casos. El paradigma del motor de inferencia es la estrategia de búsqueda para producir el conocimiento demandado.[23] Los sistemas de razonamiento basado en casos se clasifican en uno de los tres niveles siguientes, según el conjunto de facilidades que proporcionen [23]:

- **Sistemas consejeros que tan sólo recuperan casos:** Sin las facilidades de adaptación, evaluación y reparación. El usuario describe una situación y el sistema devuelve casos previos relevantes de los que el usuario extraerá sus propias conclusiones.
- **Sistemas con recuperación y adaptación:** Sin evaluación ni reparación. El usuario describe una situación, y el sistema encuentra y adapta casos previos similares.
- **Sistemas con recuperación, adaptación y reparación:** El usuario describe una situación y el sistema encuentra y adapta casos previos relevantes. Si la solución propuesta falla, el sistema la repara y también modifica el mecanismo de recuperación para evitar el mismo error en el futuro.

Para este trabajo se propone un sistema de razonamiento basado en casos que incluya recuperación, adaptación y reparación; estas etapas son necesarias para la definición del algoritmo de aprendizaje automático que pueda aplicarse al problema planteado.

### Funcionamiento del Motor de Inferencia.

La función del motor de inferencia se puede detallar en una serie de pasos lógicos y secuenciales, empezando por la detección de casos aplicables al nuevo problema que se desea solucionar. Como puede haber más de un caso aplicable, es necesaria la elección de uno de ellos, para lo cual se establece algún criterio de selección que sea capaz de escoger el caso con mayor grado de similitud. En la adaptación se modifica el nuevo caso y se indexa la solución al problema que este plantea, quedando el caso listo para ser almacenado en la base de casos si fuese necesario, evitando siempre el establecimiento de datos redundantes, lo que vuelve ineficiente la base de casos.[15]

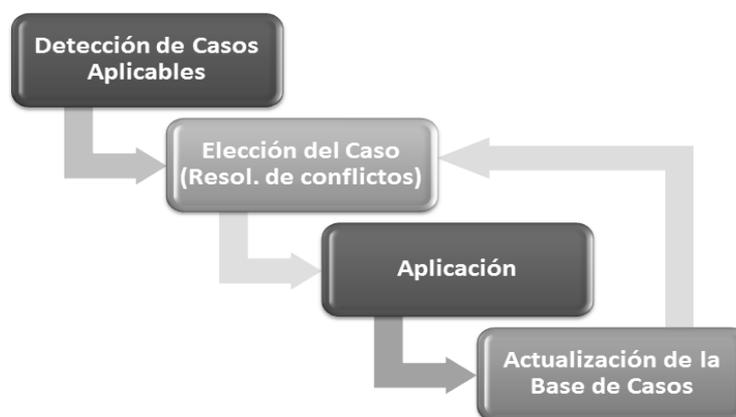


Figura 3: Ciclo base del motor de inferencia

### Pasos que incluye una inferencia basada en casos: [21]

1. Presentar como entrada al sistema una descripción del problema a resolver o problema actual (Presentación).
2. Encontrar en memoria aquellos casos relevantes que resuelven problemas similares al actual (Recuperación).
3. Determinar cuáles partes del viejo caso se deben focalizar para resolver el sub-problema actual.
4. Derivar una solución adaptando la solución previa a las restricciones del nuevo problema (Adaptación).

5. Chequear la consistencia del valor derivado con la descripción del problema a resolver, y aceptar o rechazar este (Validación).
6. Si se considera apropiada la solución validada se añade a la base de casos para ser usada en el futuro si se considera conveniente (Actualización).

#### 1.4 Análisis de soluciones existentes.

Se realiza un estudio sobre diferentes sistemas encontrados por los investigadores que mantienen relación con el dominio de la investigación los cuales pueden significar posibles soluciones a la problemática planteada.

##### 1.4.1 Panorama internacional.

Durante la investigación realizada se encontraron varios sistemas, algunos de ellos inteligentes, en el ámbito internacional. A continuación se hace mención de los mismos:

##### **Sistemas que se relacionan con el diseño de base de datos.**

**WinRDBI:** Herramienta educativa, para aprender lenguajes relacionales. Para ello, dispone de una interfaz de usuario que posibilita la creación de bases de datos, la inserción de contenidos en dichas bases de datos, y la formulación de consultas en álgebra y cálculo relacional y SQL.[24]

**WebSQL:** Es una herramienta interactiva de internet para la enseñanza del lenguaje de consulta estructurado (SQL). Facilita el aprendizaje de SQL, proporcionando una interfaz para los principales sistemas de gestión de bases de datos para la ejecución de la consulta en la web. Permite a los estudiantes aprender con bases de datos grandes y ricas, sin los problemas habituales asociados con la instalación de sistemas de bases de datos cliente / servidor en un laboratorio de campus tradicional. El tutorial SQL interactivo permite a los estudiantes invocar la interfaz de consulta WebSQL y modificar dinámicamente y ejecutar cualquier consulta de ejemplo. [25]

**SQL-Tutor:** Un sistema inteligente de enseñanza diseñado como un entorno de aprendizaje por descubrimiento guiado, que ayuda a los estudiantes a superar las dificultades en SQL. Se presentan problemas de diseño buscando hacer especial énfasis en la individualización de la enseñanza hacia un estudiante en particular. El sistema ha sido utilizado por los estudiantes de informática de alto nivel en la Universidad de Canterbury y se ha encontrado fácil de usar, eficaz y agradable. Fue desarrollado en el Allegro Common Lisp.[26]

**NORMIT:** Es un Tutor Web habilitado para la normalización de base de datos desarrollado por el Departamento de Ciencias de la Computación de la Universidad de Canterbury. Es un Sistema Tutorial Inteligente (STI) que enseña la normalización de bases de datos de los estudiantes universitarios. NORMIT utiliza modelado basado en restricciones para modelar el conocimiento del dominio y del conocimiento de sus estudiantes.[27]

**Plataforma de e-learning ACME en la docencia de bases de datos:** ACME permite asignar y corregir problemas de forma automática, siendo un complemento ideal de las clases teóricas. Al profesor le permite un seguimiento del trabajo del alumno, le facilita la labor de corrección y le ayuda en la evaluación continuada. A los alumnos les proporciona un entorno de corrección automática de problemas que les permite llevar la asignatura al día, con lo que se sienten más motivados [28].

Paquete de herramientas virtuales implementadas para apoyar la asignatura de Introducción a las Bases de Datos (IBDD) de la Facultad de Informática de la Universidad Nacional de la Plata compuesto por las siguientes herramientas:[29-31]

**CasER - Herramienta para la enseñanza de Modelado Conceptual de Bases de Datos:** CasER provee la generación asistida de un esquema conceptual. Se parte de una especificación detallada del problema, y en forma semi-automática, se genera el esquema conceptual, con el fin de facilitar el proceso de enseñanza y aprendizaje. En el comportamiento propuesto para CasER, se inicia desde la especificación general de un problema, se trabaja con el documento de requerimientos disponible, y se genera un modelo conceptual de alto nivel. Luego, a partir de un proceso de depuración y refinado a cargo del usuario, se obtiene el modelo conceptual definitivo. La herramienta CasER está basada en las notaciones para representar modelos de datos que utiliza la asignatura de IBDD.

**HEA. Herramienta de Software para enseñanza de árboles:** El propósito fundamental de HEA (Herramienta de software para la enseñanza de árboles) es brindar al alumno un asistente para la ayuda en el aprendizaje de conceptos básicos de árboles B como alternativa para implantar estructuras de índices para las base de datos. El desarrollo de HEA complementa la actividad de enseñanza y aprendizaje en lo que respecta a árboles balanceados. Así, el alumno puede analizar la resolución de un problema, generando el caso de uso y comprobando su evolución paso a paso.

#### **Otros sistemas inteligentes.**

**Sistema Tutorial Inteligente (STI) CircSim:** Es un STI para el dominio de la fisiología, fue desarrollado en conjunto por el Departamento de Ciencias de la Computación del Illinois Institute of Technology y el Departamento de Fisiología del Rush College of Medicine. CIRCSIM ha sido usado extensivamente por los estudiantes del Rush Medical College. Este tutor establece una comunicación con el estudiante con el objetivo de ayudarlo aprender y resolver problemas concernientes a las clases de fisiología cardiovascular relacionados con la regulación de la presión sanguínea. Ayuda a los estudiantes a corregir los errores conceptuales que se logran identificar a partir de sus predicciones. [32, 33]

**STI “Mentor”:** Fue desarrollado para los estudiantes de la Universidad de La Salle, Bogotá, Colombia. Es una herramienta didáctica que permite personalizar el proceso de enseñanza-aprendizaje de la solución de problemas matemáticos en los que intervienen ecuaciones lineales. Además, es capaz de seguir el desempeño del estudiante en cualquier momento, de ajustarse a sus necesidades, y ejecutar las acciones pedagógicas pertinentes, influyendo así de manera significativa

en la comprensión conceptual y el desarrollo de las habilidades cognitivas del estudiante para resolver problemas. [34]

**Sistema Tutor Inteligente ANDES:** El sistema Andes fue desarrollado por el equipo de Kurt Van Lehn de la Universidad de Pittsburg. El sistema se encarga de guiar a los estudiantes mientras estos resuelven problemas y ejercicios, en cursos de introducción a la Física. Cuando el estudiante pide ayuda en medio de un ejercicio, el sistema aporta pistas para avanzar en la solución o indica qué ha fallado en algún paso anterior. Andes fue probado con éxito durante 5 años en la Academia Naval de los Estados Unidos y puede descargarse gratuitamente. [35]

**Sistema de Razonamiento Basado en Casos para la clasificación de fallos en sistemas dinámicos:** Desarrollado en conjunto por las universidades de Valladolid y Burgos. Representa un marco computacional para resolver problemas de clasificación de fallos usando razonamiento basado en casos el cual permite la identificación de modelos de fallos en procesos continuos. El sistema razonamiento basado en casos trabaja en paralelo con un sistema de diagnóstico basado en modelos que efectúa detección y localización de fallos. Una vez que un fallo es detectado, el sistema razonamiento basado en casos proporcionará una pista sobre el fallo candidato más probable.[36]

Los sistemas y herramientas encontrados en el panorama internacional son descartados como soluciones a la problemática planteada en la investigación. La razón de esto es que los relacionados con el diseño de base de datos están enfocados en la normalización, SQL, álgebra y cálculo relacional. La utilizada para apoyar el diseño conceptual está enfocada en la notación utilizada en la asignatura IBBDD así como no permite almacenar la experiencia acumulada. Las restantes no se ajustan al dominio de las bases de datos.

#### 1.4.2 Panorama nacional.

Además, se encontraron otros sistemas en el panorama nacional entre los que se encuentran:

**Sistema Tutor Inteligente para el Diagnóstico y Tratamiento de Infecciones de Transmisión Sexual (STIITS):** Como aplicación constituye el primer STI cubano para el diagnóstico y tratamiento de ITS desarrollado en la provincia de Cienfuegos. STIITS responde a una necesidad social pero también tiene posibilidades de empleo en la docencia. Con STIITS se pretende capturar el conocimiento de los expertos en las especialidades de ginecología y obstetricia y crear interacciones en forma dinámica, para una mejor comprensión y desarrollo de estos temas por los estudiantes.[35]

**Utilización del Razonamiento Basado en Casos en las Revisiones de la Definición del Modelo de Negocio:** Desarrollado por el centro de estudios de Ingeniería de Sistemas del Instituto Superior Politécnico José Antonio Echeverría. El sistema utiliza el razonamiento basado en casos para asistir a los estudiantes y profesores en la revisión de la definición del modelo del negocio y del dominio. Permite a los alumnos de Ingeniería Informática entrenarse en el uso de las mejores prácticas en cuanto a la modelación de sistemas informáticos, específicamente en las fases iniciales aprovechando la experiencia de otros desarrolladores y las de los estudiantes en su rol de

desarrolladores de software en nuevos proyectos de las asignaturas de análisis y diseño de sistemas informáticos. [37]

**HESEI:** Es una herramienta computacional para elaborar sistemas de enseñanza-aprendizaje inteligentes, la cual utiliza técnicas de inteligencia artificial y mapas conceptuales, con el objetivo de adaptar el sistema de enseñanza-aprendizaje, a través de una interfaz visual, a las características del alumno. Permite a expertos no especialistas en computación crear sistemas de enseñanza-aprendizaje inteligentes que aborden contenidos de cualquier especialidad y nivel de enseñanza. Representa un modelo que integra el Razonamiento Basado en Casos y los SEAI (Sistemas de Enseñanza /Aprendizaje Inteligentes).[38]

**Sistema Tutorial Inteligente de apoyo al proceso de enseñanza-aprendizaje de la asignatura de Sistemas Operativos desarrollado en la Universidad de las Ciencias Informáticas:** Tiene el objetivo de apoyar el proceso de enseñanza-aprendizaje de la asignatura de Sistemas Operativos de la Facultad 2 en la Universidad de las Ciencias Informáticas ayudando a consolidar los contenidos del plan de estudio definido para la asignatura, utilizando el razonamiento basado en casos.[39]

**Sistema Web Inteligente apoyado en Mapas Conceptuales para la asignatura de Álgebra Lineal:** Incorpora mapas conceptuales adaptados al estado cognitivo particular de cada estudiante, utilizando la técnica razonamiento basado en casos. El sistema tiene el objetivo de contribuir a consolidar los conocimientos de los estudiantes en la asignatura, permitiéndoles una personalización del contenido a estudiar, logrando con esto que los estudiantes puedan determinar cuáles son sus principales deficiencias en la asignatura, brindándoles a su vez la información necesaria para la erradicación de estas deficiencias. [40]

Ninguno de estos sistemas puede ser utilizado como solución a la problemática planteada debido a que no comprenden el diseño conceptual de base de datos.

### 1.5 Conclusiones parciales.

En el presente capítulo se realiza un estudio del estado del arte de los sistemas utilizados en apoyo a la enseñanza del diseño conceptual de base de datos. Los sistemas encontrados no logran satisfacer la problemática planteada. Por lo cual se propone el diseño de una base de casos y un motor de inferencia que pueda ser utilizado como parte de un sistema inteligente que permita manejar la experiencia de los expertos en el diseño conceptual de base de datos así como apoyar el aprendizaje de los estudiantes en este tema. Para esto se estudian las técnicas de inteligencia artificial basadas en el conocimiento y se concluye que el razonamiento basado en casos es el que mejor se adapta para dar solución a la problemática planteada, ya que facilita la gestión de la experiencia que se ha adquirido en el tema tratado y además los estudiantes pueden ser representados como casos.

Capítulo 2: Diseño de la base de casos y motor de inferencia.

2.1 Introducción.

La solución que se propone en este capítulo concibe al sistema basado en casos como parte de un sistema inteligente con la cual el estudiante interactúa para realizar ejercicios propios del diseño conceptual de bases de datos. Estos ejercicios están concebidos hasta su forma de calificación pudiendo otorgar evaluaciones para cada una de las habilidades que desarrolla y conocimientos que forma. Es precisamente esta, la entrada de datos que necesita el sistema basado en casos para determinar el nivel de conocimiento que tiene el estudiante en el tema y los contenidos que aún tiene que trabajar. La siguiente figura muestra como ocurre la interacción entre el estudiante y el sistema.

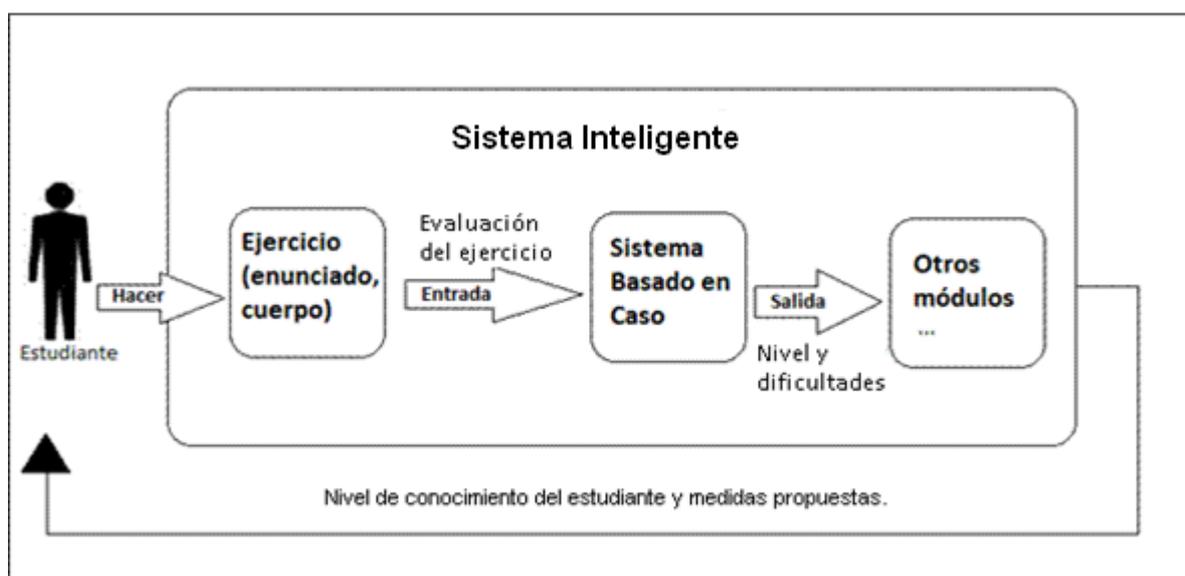


Figura 4: Interacción del estudiante con el sistema inteligente

Los resultados que se obtienen del sistema basado en casos posibilitan que el sistema trabaje en el aprendizaje del estudiante en función del nivel obtenido en el tema diseño conceptual y los contenidos donde presenta dificultad. De esta forma se pudiera, además de informarles al estudiante y al profesor su estado actual, proponer materiales a consultar, nuevos ejercicios a realizar de menor o igual complejidad y guardar una traza de su mejoramiento con vistas a mejorar la experiencia del sistema basado en casos, entre otras posibles potencialidades.

En el presente capítulo se describe el diseño de la base de casos y del motor de inferencia detallando los elementos de las fases que conforman el sistema basado en casos. Para ello se selecciona y fundamenta la estructura con que son representados los casos en la base de casos, las funciones de semejanza que utiliza el motor de inferencia en la fase de recuperación y el umbral de similitud que se emplea para escoger los casos similares al caso nuevo.

## 2.2 Base de casos.

La base de casos es el componente que contiene los conocimientos a partir del cual se infiere la solución a nuevas situaciones. Es la encargada de almacenar y organizar los casos con vista a utilizarlos en próximas ocasiones.

Para la solución un caso contiene, como información, las evaluaciones de un estudiante al resolver un ejercicio sobre diseño conceptual de base de datos. El problema a resolver en este punto está en definir la estructura de almacenamiento que debe utilizarse y que permita la recuperación de los casos. En los epígrafes siguientes se fundamenta la selección de la estructura utilizada y cuáles son los rasgos predictivos que se utilizan.

### 2.2.1 Técnicas para la adquisición de conocimientos.

El proceso de adquisición del conocimiento es una parte crucial y a la vez muy difícil a la hora de concebir un sistema basado en casos. Depende del factor humano, de la experticia y de la estructuración del conocimiento de la fuente de información. De las diferentes técnicas de adquisición del conocimiento las más utilizadas son:

**Observación directa:** es una técnica de extracción del conocimiento que consiste en la observación simple del modo en que un experto se enfrenta a los problemas del dominio.

**Entrevistas no estructuradas y semi-estructuradas:** se basa en realizar una entrevista con los expertos para investigar como representan mentalmente sus conocimientos y como analizan la información inconsistente o imprecisa.

**Conocimiento documentado:** consiste en realizar la extracción de conocimiento de diversas fuentes implicadas (libros, artículos de revistas, manuales, informes y documentos de ayuda), en formato digital o impreso, ya que cualquier documento puede ser útil para la introducción de conocimiento en la base caso.

La técnica seleccionada para identificar los elementos a almacenar es el conocimiento documentado. Los documentos utilizados son las bibliografías básicas de la asignatura Sistemas de Base de Datos 1 para el tema de diseño conceptual de base de datos y que se encuentran disponibles en el Entorno Virtual de Enseñanza Aprendizaje (<http://eva.uci.cu/mod/resource/view.php?id=26564>). Los expertos que participaron en la elaboración o selección de estos documentos son los profesores del colectivo de dicha asignatura luego de años de perfeccionamiento de la didáctica y de experiencia en la impartición del tema. Los documentos consultados son:

- *Introducción a los sistemas de bases de datos.* Elaborado como recopilación de textos por el Colectivo de profesores.
- *Modelo entidad relación (MER). Extensiones del MER.* Elaborado como recopilación de textos por el Colectivo de profesores.

- *Patrones de diseño de bases de datos.* Elaborado como recopilación de textos por el Colectivo de profesores.
1. Para la determinación de los pesos que acompañan a cada uno de los rasgos predictores considerados para la solución se utiliza la técnica entrevista estructurada la cual es aplicada en un panel de expertos conformado por los profesores de la asignatura. El modelo de la entrevista aplicada se encuentra en el Anexo 1. Para aplicar esta técnica se siguen los siguientes pasos: Anexo 1 Selección de un grupo de expertos.
  2. Selección de un conjunto de habilidades que se transforman en los atributos de entrada o rasgos que definen los casos ( $m$  = cantidad de rasgos predictores).
  3. Cada experto otorga un orden de prioridad a cada elemento que va de 1 a  $m$ . El orden de prioridad es inversamente proporcional al valor que lo identifica, siendo un orden 3 más importante que un orden 11.

Esta información es utilizada posteriormente para el cálculo de los pesos. (Ver acápite Cálculo del peso de los rasgos predictores a partir de expertos)

### 2.2.2 Representación del conocimiento.

Como se explicó anteriormente, se concibe como entrada de datos del sistemas basado en casos las evaluaciones resultantes para cada rasgo predictor. Este es el punto de partida que permite que el sistema basado en casos dé como salida el nivel de conocimiento del estudiante y las dificultades que aún debe superar. Antes de explicar la estructura de memoria a utilizar, es necesario definir los elementos que conforman un caso.

Estos elementos son: identificador, que es un numérico consecutivo; los rasgos predictores, acompañados de los pesos correspondientes y de los valores de evaluación obtenidos por el estudiante los cuales quedan definidos en el rango de 2 a 5 siendo 5 el valor máximo; y los rasgos objetivos. Los rasgos predictores estarán determinados por 18 elementos básicos pertenecientes al diseño conceptual de base de datos. Estos son:

1. Identificar entidades y sus atributos.
2. Identificar entidades débiles.
3. Identificar llave.
4. Identificar atributos multi-valorados.
5. Identificar atributos compuestos.
6. Identificar atributos derivados.
7. Identificar la cardinalidad mínima.
8. Identificar la cardinalidad máxima.
9. Identificar relaciones n-arias.
10. Identificar asociaciones recursivas.
11. Identificar relación generalización-especialización.
12. Identificar relación de agregación.

13. Aplicar patrón árboles fuertemente codificados.
14. Aplicar patrón árboles simples.
15. Aplicar patrón árboles estructurados.
16. Aplicar patrón grafo dirigido simple.
17. Aplicar patrón Modelo Entidad-Atributo-Valor.
18. Aplicar patrón grafo estructurado.

Los rasgos objetivos lo constituirán el nivel alcanzado por el estudiante según sus notas y los contenidos donde presenta dificultad. El nivel se corresponde a categorías usualmente utilizadas para clasificar el estado del conocimiento en la enseñanza universitaria. Estas categorías son: alto, medio, bajo y sin nivel. Los tres contenidos definidos son:

- Entidades y Atributos (EA): Relacionado con los rasgos del 1-6.
- Inter-relaciones (I): Relacionado con los rasgos del 7-12.
- Patrones de diseño (P): Relacionado con los rasgos del 13-18.

A continuación se representa la estructura general de un caso y los posibles valores a tomar. Nótese que en la última fila se identifican posibles combinaciones de contenidos en las que el estudiante puede presentar dificultades. Cada combinación está separada por punto y coma (;) para su identificación.

Caso		Valores a tomar
<b>Rasgos predictores</b>	Identificar entidades y sus atributos.	2, 3, 4 y 5
	Identificar entidades débiles.	
	Identificar llave.	
	Identificar atributos multi-valuados.	
	Identificar atributos compuestos.	
	Identificar atributos derivados.	
	Identificar la cardinalidad mínima.	
	Identificar la cardinalidad máxima.	
	Identificar relaciones n-arias.	
	Identificar asociaciones recursivas.	
	Identificar relación generalización-especialización.	
	Identificar relación de agregación.	
	Aplicar patrón árboles fuertemente codificados	
	Aplicar patrón árboles simples	
	Aplicar patrón árboles estructurado	
Aplicar patrón grafo dirigido simple		
Aplicar patrón Modelo Entidad-Atributo-Valor		

	Aplicar grafo estructurado	
<b>Rasgos objetivos</b>	Nivel	Alto, Medio, Bajo y Sin Nivel.
	Insuficiencias detectadas	Posibles combinaciones resultantes "EA-I-P"; "EA-I"; "EA-P"; "I-P"; "EA"; "I"; "P".

**Tabla 2.1:** Estructura de un caso: elementos que la conforman y posibles valores asociados

El valor del peso del rasgo predictor unido a la evaluación que obtiene el estudiante, permiten la comparación de casos nuevos con los ya almacenados otorgándoles prioridad de recuperación.

### Cálculo del peso de los rasgos predictores a partir de expertos.

Como se expuso en el epígrafe: Técnicas de adquisición del conocimiento, este acápite se fundamenta en el conocimiento de un panel de expertos. Para el cálculo del peso el primer paso es sumar los órdenes de prioridad dados por cada experto para cada rasgo tal como se representa en la ecuación tomada del trabajo de diploma [41]:

$$T_i = \sum_{i=1}^m O_i$$

**Fórmula 1:** Suma de las órdenes de prioridad

Siendo:

$T_i$ : Valor resultante de la suma de los órdenes.

$m$ : La cantidad de rasgos predictores.

$O_i$ : Orden de prioridad.

Con el objetivo de obtener pesos con valores entre 0 y 1 para cada atributo y teniendo en cuenta además, que a menor valor de  $T_i$  le corresponde un mayor peso, se utiliza la fórmula siguiente: [41]

$$w_i = \frac{1 - \frac{T_i}{T_i}}{m - 1}$$

**Fórmula 2:** Cálculo del peso del rasgo predictor

Donde:

$w_i$ : Peso de la variable.

$T_i$ : Suma de los órdenes de prioridad dados por el experto a cada elemento.

$T_i$  : Suma general de los órdenes de prioridad dados por el experto.

$m$ : Cantidad de rasgos predictores.

Los resultados del cálculo de los pesos se pueden apreciar en el Anexo 2.

### **Modelo de la Base de Casos.**

La selección de la estructura para almacenar los casos en la base de casos está orientada a facilitar la recuperación de estos. Teniendo en cuenta el estudio realizado en el Capítulo 1 sobre este tema y que los casos a almacenar son sencillos se ha decidido utilizar la representación plana. Esta estructura brinda mayor facilidad para insertar los nuevos casos y posibilita que el algoritmo de recuperación a implementar sea poco complejo.

### **2.3 Motor de inferencia.**

El motor de inferencia es el componente del sistema basado en casos que se encarga de determinar el nivel de conocimiento y las deficiencias que presenta un estudiante en el tema diseño conceptual de base de datos tomando en cuenta valoraciones anteriores almacenadas en la base de casos. Realizar este proceso implica la utilización de algoritmos que permitan reutilizar el conocimiento previamente almacenado y adaptarlos de manera que resuelvan la nueva situación. A continuación se describe el diseño de cada uno de estos algoritmos así como aquellos que garantizan la evaluación de la solución propuesta y el almacenamiento de nuevos conocimientos para la base de casos.

#### **2.3.1 Fase de recuperación.**

El algoritmo de recuperación exige realizar una comparación entre casos. Este proceso está determinado por las funciones de semejanzas que se utilizan al comparar los rasgos predictores, las cuales se seleccionan de acuerdo a sus valores. Para evitar la recuperación de casos que resulten irrelevantes para la solución, se utiliza un umbral de semejanza que reduzca el número de casos recuperados a solo los más parecidos. El primer paso es definir las medidas de similitud entre rasgos predictores y de semejanza entre casos.

#### **Medidas de similitud entre rasgos.**

La comparación entre los rasgos permite establecer una relación de similitud entre los valores de dominio de un rasgo del nuevo caso y el caso almacenado en la BC. El grado de similitud está determinado por la función que se utilice para calcularlo, la selección viene en correspondencia con el tipo de dato del rasgo predictor. En esta propuesta los rasgos predictores tienen como característica que todos presentan un mismo tipo de datos (ordinales) por lo que solo es necesario definir una función.

Los rasgos predictores son valores ordinales y entre estos tipos de datos no tiene sentido realizar operaciones aritméticas. Para estos casos Jiawei Han y Micheline Kamber proponen en el libro Data Mining [42] normalizar estos datos para luego aplicar una función de distancia. Para el tratamiento de estas variables se definen valores numéricos con los diferentes estados ordenados para la

clasificación. El conjunto quedaría conformado  $\{1, \dots, M\}$  c, donde M, es el último valor ordenado. Como todos los rasgos predictores toman el mismo tipo de dato, el conjunto queda definido de la siguiente manera: “2” = 1, “3”= 2, “4”= 3, “5”=4 y  $M_i=4$ . Luego a cada valor del dominio, se hace corresponder un valor real entre 0 y 1, siguiendo la siguiente fórmula:

$$x_i = \frac{v - 1}{M_i - 1}$$

**Fórmula 3:** Normalización de los valores ordinales

Donde:

$v$ : Valor del número entero asignado al valor del rasgo.

$M_i$ : Cantidad de estados ordenados.

A partir de la normalización se pueden utilizar funciones de comparación para este tipo de rasgos. Se propone utilizar la siguiente función basada en la distancia de Manhattan ajustada al rango. Esta función se muestra a continuación:

$$\delta_i O_o, O_t = 1 - \frac{X_i O_o - X_i(O_t)}{r_{máx} - r_{mín}}$$

**Fórmula 4:** Cálculo de similitud entre dos rasgos predictores

Donde:

$O_o$ : Caso a resolver.

$O_t$ : Caso existente en la base de casos.

$r_{máx}$ : Mayor valor del conjunto de valores existentes.

$r_{mín}$ : Menor valor de los rasgos.

### **Función de semejanza entre casos.**

La función de semejanza permite determinar cuán semejante es el nuevo caso respecto a los existentes en la base de casos. No existe una función única para todos los dominios, por lo que la eficiencia del proceso de recuperación de los casos semejantes radica en su correcta selección. La misma permite la recuperación de aquellos casos que presenten mayor semejanza con el nuevo.

La función puede incluir entre sus argumentos, la importancia o relevancia de cada rasgo al integrar los valores obtenidos en la comparación rasgo a rasgo, de modo que el resultado de la comparación de rasgos más relevantes, tenga más peso en el cálculo del grado de similitud entre el caso nuevo y un elemento de la memoria permanente. Con tal motivo se hace uso de la fórmula suma ponderada [43], esta incluye el peso de los rasgos predictores.

$$\beta O_o, O_t = \frac{\sum_{i=1}^n W_i \cdot \delta_{x_i O_o, x_i O_t}}{\sum_{i=1}^n W_i}$$

**Fórmula 5:** Suma ponderada

Donde:

$w_i$ : Peso asociado al rasgo  $i$ .

$O_0$ : Caso a resolver.

$O_t$ : Caso existente en la base de casos.

$n$  : Número de rasgos predictores.

$\delta x_i O_0, x_i O_t$  : Distancia entre los rasgos.

**Umbral de semejanza.**

El umbral de semejanza restringe el número de casos a recuperar a un intervalo donde solo se encuentren los casos con mayor semejanza. Un umbral muy elevado puede representar la ventaja de que solo se tengan en cuenta en la fase de adaptación los casos con un alto nivel de semejanza, pero puede presentar la desventaja de que ninguno de los casos recuperados lleguen o sobrepasen este valor, lo que traería consigo no tener al menos un caso para poder realizar la adaptación [14]. El valor del umbral se puede obtener mediante el criterio del experto, sin embargo también puede ser obtenido a partir de la construcción de una matriz cuadrada donde la fila y las columnas están representadas por los casos que se encuentran almacenados en la base de casos y en la intersección está el valor de semejanza ( $\beta$ ).

	<i>Caso 1</i>	<i>Caso 2</i>	<i>Caso 3</i>	<i>Caso n</i>
<i>Caso 1</i>	$\beta_{C_1, C_1}$	$\beta_{C_1, C_2}$	$\beta_{C_1, C_3}$	$\beta_{C_1, C_n}$
<i>Caso 2</i>	$\beta_{C_2, C_1}$	$\beta_{C_2, C_2}$	$\beta_{C_2, C_3}$	$\beta_{C_2, C_n}$
<i>Caso 3</i>	$\beta_{C_3, C_1}$	$\beta_{C_3, C_2}$	$\beta_{C_3, C_3}$	$\beta_{C_3, C_n}$
<i>Caso n</i>	$\beta_{C_n, C_1}$	$\beta_{C_n, C_2}$	$\beta_{C_n, C_3}$	$\beta_{C_n, C_n}$

**Tabla 2.2:** Matriz de semejanza entre casos

Luego de construida la matriz, mediante el cálculo de  $\beta_0$  se obtiene el valor umbral que dependerá de los valores de semejanza entre los casos contenidos en dicha matriz [44]:

$$\beta_0 = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta(O_0, O_t)$$

**Fórmula 6:** Umbral de semejanza

Donde:

$m$ : Número de casos

$i$ : Casos en filas

$j$ : Casos en columnas

$\beta (O_0, O_t)$ : Función de semejanza entre los casos  $O_0, O_t$

A continuación se definen los pasos que describen el proceso de recuperación de los casos más semejantes en la estructura plana.

### Algoritmo de acceso y recuperación.

La fase de recuperación toma como entrada el caso nuevo y los casos almacenados en la base de casos con sus respectivos pesos. Este proceso de recuperación realiza los siguientes pasos:

- Cálculo del umbral de semejanza a través de la Fórmula 6: Umbral de semejanza.
- Crea una lista  $S$  donde se almacenen los casos con semejanza mayor o igual al umbral de semejanza calculado.
- Devolver la lista  $S$  con los casos más semejantes para que puedan ser utilizados en la fase de adaptación.

### 2.3.2 Fase de adaptación.

En este módulo se transforman los casos recuperados con el objetivo de formar una nueva solución. La adaptación consiste en la modificación y combinación de las soluciones de los casos similares. Existen varios métodos de adaptación los cuales se pueden clasificar como: estructurales y derivacionales.

Se habla de métodos estructurales cuando lo que se modifica es la solución previa con el objetivo de formar una nueva solución. Estos métodos aplican las reglas de adaptación directamente a la solución almacenada. Estos pueden sustituir un solo componente de la solución, varios, o todos ellos.[36]

En los métodos derivacionales no solo se guarda la solución del caso recuperado sino también el método que la construyó. Si la diferencia entre los valores de la entrada y la recuperada afectan alguno de los valores de la antigua solución estos serán reevaluadas utilizando los valores existentes en la entrada.[36]

De estos dos tipos de métodos el más apropiado para ser usado en la presente investigación es el método adaptación estructural, debido a que para obtener las nuevas soluciones es suficiente con adaptar directamente las soluciones que proponen los casos previos almacenados en la base de casos. A continuación se enuncian las principales técnicas de adaptación descritas en el trabajo “Un estudio sobre la aplicación del razonamiento basado en casos a la construcción de programas.” [45]:

- **Adaptación nula:** Esta técnica propone no realizar ningún cambio, simplemente aplicar cualquier solución de las recuperadas al nuevo caso. La adaptación nula aparece en aquellas tareas donde el razonamiento necesario para una aplicación puede ser muy complejo, y la solución en sí misma muy simple.

- **Adaptación transformacional:** Consiste en reutilizar soluciones recuperadas después de aplicárseles determinadas transformaciones que generen la solución buscada. Estas transformaciones se pueden realizar ajustando parámetros o substituyendo valores.
- **Adaptación derivacional:** Consiste en realizar la adaptación de algún elemento de la solución que no se ajusta correctamente reapplicando, para el problema actual, el procedimiento que fue utilizado para calcular dicho elemento de la solución en el caso previo.
- **Sistemas cooperativos:** Al adaptar utilizando esta técnica se establece una cooperación entre el sistema y sus usuarios, donde el sistema únicamente presenta las soluciones relevantes y el experto humano es quien se encarga de realizar la adaptación de éstas.

### Algoritmo de adaptación

En la propuesta de solución el algoritmo de adaptación utiliza dos de las técnicas antes mencionadas: nula y transformacional. Toma como entrada los casos más semejantes obtenidos en la fase de recuperación y el nuevo caso. El algoritmo tiene en cuenta aquellos casos que fueron clasificados por el experto como fallos, o sea que la solución asignada fue errónea, para no volver a cometer el mismo error. Se define en el algoritmo los siguientes pasos:

1. Asignar a la variable  $\beta'$  el mayor valor de semejanza.
2. Guardar en la lista  $L$  los casos con semejanza igual a  $\beta'$ .
3. Si  $L$  tiene un solo elemento se aplica la adaptación nula, proponiéndole su solución al nuevo caso.
  - a. Sino, se realiza la adaptación transformacional.
    - i. Si los niveles de los casos son iguales se procede a realizar la unión de los contenidos donde presenta problemas el estudiante y se le asigna el nivel al caso nuevo.
    - j. Sino se realiza la unión de los contenidos y se asigna al nuevo problema el menor valor de nivel.
4. Guardar en la lista de casos resueltos  $E$  el caso adaptado.

Cuando el caso de mayor semejanza sea un fallo, se toma el segundo caso más parecido como candidato a adaptar.

### 2.3.3 Fase de revisión.

La fase de revisión tiene como objetivo ver si es correcta la solución propuesta. El experto es responsable de evaluar si la solución propuesta al nuevo caso es correcta, calificándolo como un caso de éxito, o no, denominándolo caso fallo. Esta fase toma como entrada la lista  $E$  con los casos adaptados. El resultado o salida de esta fase serán los casos evaluados y listos para almacenar. En el caso de que la entrada sea idéntica a la salida no se almacena pues no representa conocimiento nuevo.

#### **2.3.4 Fase de retención.**

La fase de retención es la última del ciclo del razonamiento basado en casos. Los casos clasificados como fallos se marcan como tal mientras los clasificados correctamente son guardados sin necesidad de marcarlos. El almacenamiento de los casos fallidos asegura que no se vuelva a cometer el mismo error dos veces, es decir, que cuando se recupere un caso semejante al que devolvió para este nuevo caso no le dará la misma respuesta fallida.

#### **2.4 Conclusiones parciales**

En este capítulo se definieron los componentes de la base de casos y el motor de inferencia propuestos como solución. Los resultados obtenidos son:

- Definición de los componentes de un caso: 18 rasgos predictores extraídos de los materiales utilizados para impartir la temática diseño conceptual de base de datos; y los rasgos objetivos que representan el nivel de conocimiento del estudiante y los contenidos donde tiene dificultad.
- Selección de la estructura de memoria plana como estructura de almacenamiento.
- Selección de la función de similitud tomando en cuenta que los rasgos predictores son valores ordinales.
- Selección de la función de semejanza suma ponderada para la comparación entre casos.
- Definición de las fases de recuperación, adaptación, revisión y retención.

## Capítulo 3: Validación de la solución propuesta

### 3.1 Introducción.

En el presente capítulo se realiza la validación de la base de casos y el motor de inferencia. Para validar la primera se utiliza el método Delphi, descrito por Armín González Almaguer en el trabajo [46], basado en el criterio de los expertos. A su vez, para el motor de inferencia es desarrollada una aplicación que implementa los algoritmos y fórmulas definidas.

### 3.2 Método Delphi.

El método Delphi o Delfos fue creado alrededor de los años 1963-1964 por la Ran Corporation, específicamente por Olaf Helmer y Dalkey Gordon, con el objetivo de elaborar pronósticos a largo plazo. Consiste en la utilización sistemática del juicio intuitivo de un grupo de expertos para obtener un consenso de opiniones informadas. De aquí que este método también se conozca con el nombre de Criterio de expertos [46].

A continuación se describe las características del método, estas se pueden consultar en [46].

- **Anonimato:** Los expertos contestan las preguntas sin consultarse mutua-mente (por lo que es recomendable que dos expertos no conozcan entre sí que están opinando sobre un mismo tema).
- **Retroalimentación controlada:** Después de cada ronda de preguntas, se tabulan las respuestas y se procesan antes de la siguiente ronda, para que los participantes puedan evaluar los resultados de la ronda anterior, así como las razones dadas para cada respuesta y su dispersión del promedio (esto permite que aumente el acuerdo al transcurrir varias rondas del proceso).
- **Respuesta estadística del grupo:** El procesamiento de cada ronda se realiza con métodos estadísticos. Esto es la característica más importante que diferencia a este método de otros subjetivos.

### 3.3 Validación de la base de casos mediante Delphi.

En la investigación se usa el método Delphi para la validación de la base de casos propuesta, a través de este los expertos pueden predecir los resultados a alcanzar con la propuesta elaborada. El procesamiento estadístico de la información es la parte más importante del método, que lo diferencia del resto de los métodos, debido a que la decisión final que se toma es un criterio fuertemente avalado por la experiencia de los profesores consultados. Para la aplicación práctica del Delphi hay que considerar los siguientes aspectos:

1. Selección de los expertos.
2. Elaboración del cuestionario para la validación de la propuesta.
3. Cálculo de la concordancia de los especialistas.

4. Desarrollo práctico y explotación de los resultados.

Para facilitar el trabajo se utiliza como herramienta de apoyo “Concordancia entre los expertos” (en lo adelante CEE), desarrollada por la Máster en Ciencias Sandra Hurtado de Mendoza Fernández. La principal ventaja que posee esta herramienta con respecto a otras es su enfoque a la aplicación del método Delphi para el cálculo de la concordancia entre los expertos. Ofrece la posibilidad de seleccionar los expertos a consultar, valorar los aspectos a evaluar y el cálculo de la concordancia.

**3.3.1 Selección de los expertos.**

Se considera como experto a la persona que se especializa en un área y es capaz a su vez de resolver problemáticas particulares de esta. La determinación del grupo de expertos debe garantizar la seguridad de los resultados con el mínimo de gastos. Esta seguridad depende de la cantidad de expertos y de la estructura del grupo de ellos por especialidades y, asimismo, de las características particulares de los propios expertos. Para la realización de la investigación se seleccionan como expertos los profesores de la asignatura Sistema de Base de Datos 1 de la universidad, a los cuales se les aplican entrevistas que permiten validar la solución propuesta. (Anexo 3)

La principal característica que deben poseer los expertos para asegurar la calidad de la validación es la competencia. Esta es el nivel de calificación en la rama del conocimiento objeto de indagación; lo cual no está totalmente en “línea” con su grado científico y tarea, labor o responsabilidad que desempeña.[46] Es necesario que el propio experto se auto evalúe en este sentido. Esta competencia se determina por diferentes coeficientes, entre los que se encuentra el que se muestra a continuación y que se denota por K:

$$K = \frac{1}{2} K_c + K_a$$

**Fórmula 7:** Cálculo del coeficiente de competencia

Donde  $K_c$  es el coeficiente de conocimiento o información que posee la persona acerca del problema (sobre la base de su auto valoración); sus valores están en una escala de 0 a 10 que para el cálculo se multiplica por 0,1.El cero indica que la persona no posee absolutamente ningún conocimiento de la problemática en estudio, mientras que el 10 expresa pleno conocimiento. Así, la persona solicitada deberá marcar la casilla que estime pertinente en la siguiente escala (la cual se entrega a la persona elegida):

0	1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/>										

**Tabla 3.1:** Tabla para el coeficiente de conocimiento

En las siguientes figuras puede observarse tablas que contienen el nivel y coeficiente de conocimiento de los profesores encuestados, respectivamente. Estas son generadas por la herramienta CEE.

Nombre y Apellidos	1	2	3	4	5	6	7	8	9	10
Yubismel Perdomo							X			
Yamilka Gómez León								X		
Anthony Sotolongo								X		
Luis Teijón Acosta									X	
Yaili Ledea								X		
Yorgelys González							X			
Juan Gómez Correa									X	

Figura 5: Nivel de conocimiento de los profesores encuestados

Nombre y Apellidos	Coeficiente de Conocimiento(Kc)
Yubismel Perdomo	0.7
Yamilka Gómez León	0.8
Anthony Sotolongo	0.8
Luis Teijón Acosta	0.9
Yaili Ledea	0.8
Yorgelys González	0.7
Juan Gómez Correa	0.9

Figura 6: Coeficiente de conocimientos de los profesores

El parámetro  $\mathcal{K}_a$  es el coeficiente de argumentación o fundamentación de los criterios de la persona y se obtiene del resultado de la suma de los puntos alcanzados a partir de las respuestas de la persona recogidas en la siguiente tabla patrón:

Fuentes de argumentación	Grado de influencia de cada una de las fuentes en sus criterios		
	A(Alto)	M(Medio)	B(Bajo)
Análisis teóricos realizados por usted			
Su experiencia en el tema			
Trabajos de autores nacionales consultados			
Trabajos de autores extranjeros consultados			
Su propio conocimiento del estado del problema en el extranjero			
Su intuición			

Tabla 3.2: Tabla para el coeficiente de argumentación

A la persona seleccionada se le presenta la tabla anterior y se le pide que marque con una cruz (X) cuál de las fuentes él considera que ha influido en su conocimiento de acuerdo con el grado A, M o B (se le debe pedir que responda todas las fuentes). Luego, utilizando los valores para cada casilla dados en la tabla 3.4, se calcula el valor de  $\mathcal{K}_a$ .

Fuentes de argumentación	Grado de influencia de cada una
--------------------------	---------------------------------

	de las fuentes en sus criterios		
	A(Alto)	M(Medio)	B(Bajo)
Análisis teóricos realizados por usted	0.3	0.2	0.1
Su experiencia en el tema	0.5	0.4	0.2
Trabajos de autores nacionales consultados	0.05	0.05	0.05
Trabajos de autores extranjeros consultados	0.05	0.05	0.05
Su propio conocimiento del estado del problema en el extranjero	0.05	0.05	0.05
Su intuición	0.05	0.05	0.05
<b>Total</b>	<b>1</b>	<b>0.8</b>	<b>0.5</b>

Tabla 3.3: Valores para calcular el coeficiente  $\mathcal{K}_a$

Luego de la encuesta se obtuvieron los siguientes valores de  $\mathcal{K}_a$  para cada profesor encuestado.

Profesores	$\mathcal{K}_a$	Grado de influencia
1	0.9	Alto
2	1	Alto
3	0.8	Medio
4	0.9	Alto
5	0.9	Alto
6	0.8	Medio
7	0.9	Alto

Tabla 3.4: Coeficiente de argumentación

A continuación se determina el valor de K (según la fórmula 7). Después de obtenidos los resultados se valoran de la siguiente manera:

- Si  $0.8 \leq K \leq 1$  el grado de competencia es alto.
- Si  $0.5 \leq K < 0.8$  el grado de competencia es medio.
- Si  $K < 0.5$  el grado de competencia es bajo.

El coeficiente de competencia de cada uno de los expertos encuestados se puede apreciar en la tabla siguiente:

Profesores	K	Grado de competencia
1	0.8	Alto
2	0.9	Alto
3	0.8	Alto
4	0.9	Alto

5	0.85	Alto
6	0.75	Medio
7	0.9	Alto

**Tabla 3.5:** Coeficiente K

Para la validación se escogieron todos los expertos consultados debido a que poseen grado de competencia medio y alto.

### 3.3.2 Elaboración del cuestionario.

Para la evaluación propuesta se debe realizar un cuestionario (

Anexo 4) que se ajuste a las condiciones de los expertos seleccionados, en el que se solicita sean evaluados un conjunto de indicadores, estos son características que debe cumplir la propuesta para que sea válida. Para analizar los cuestionarios realizadas a los expertos se tuvo en cuenta criterios de evaluación cualitativos (Muy adecuado, Bastante adecuado, Adecuado, Poco adecuado, Inadecuado). A cada uno de estos criterios se les otorgó una puntuación entre 1 y 5 en dependencia de los valores asignados por los expertos a las preguntas en el cuestionario para su posterior análisis.

Criterios de evaluación cualitativos	Puntuación
Muy adecuado	5
Bastante adecuado	4
Adecuado	3
Poco adecuado	2
Inadecuado	1

Tabla 3.6: Criterios de evaluación

Las preguntas aplicadas responden a los siguientes objetivos:

1. Determinar la necesidad de la base de casos propuesta para dar solución a la problemática planteada.
2. Identificar la utilidad de la base de casos.
3. Determinar el nivel de vinculación de la base de casos con el diseño conceptual de base de datos.
4. Evaluación de la base de casos propuesta.

En la tabla siguiente se puede apreciar el objetivo que se satisface con cada pregunta.

Objetivos	Preguntas
Necesidad de la base de casos propuesta	1
Utilidad de la base de casos	2
Nivel de vinculación de la base de casos	3
Evaluación de la base de casos	4

Tabla 3.7: Objetivos por preguntas

### 3.3.3 Cálculo de la concordancia de los especialistas.

Con el objetivo de darle más importancia al resultado de la validación se decide determinar la concordancia de criterios entre expertos, para ello se hace uso del coeficiente de concordancia Kendall (W). Para aplicar este criterio se introducen los valores obtenidos durante la entrevista, en una tabla de aspectos a evaluar contra expertos. Estos valores podrán ser 1, 2, 3, 4,5. La siguiente figura muestra esta información:

Aspectos a evaluar	Experto1	Experto2	Experto3	Experto4	Experto5	Experto6	Experto7
Pregunta 1	4	5	5	5	4	5	5
Pregunta 2	4	4	5	5	5	5	3
Pregunta 3	5	5	5	5	5	5	5
Pregunta 4	4	5	5	5	5	4	3

Figura 7: Datos para calcular W

Luego de introducir los datos el sistema utiliza para calcular W la siguiente fórmula:

$$W = \frac{12 * S}{K^2 N^3 - N}$$

Fórmula 8: Coeficiente de concordancia de Kendall

Para calcular a W hay que:

- Determinar la suma de los valores numéricos asignados a cada aspecto a evaluar, según el criterio dado por cada experto ( $R_j$ ).

Aspectos	$R_j$
Pregunta 1	33
Pregunta 2	32
Pregunta 3	35
Pregunta 4	31

Tabla 3.8: Valores de cada  $R_j$

$$R_{jT} = \sum_{j=1}^N R_j = 131$$

- Determinar el valor medio ( $R_j$ ) de las  $R_j$ , dado por la sumatoria de las  $R_j$  entre  $N$ , siendo  $N$  el total de aspectos a evaluar (los aspectos serán los indicadores de evaluación, en este caso  $N = 4$ ), este valor se calcula mediante la fórmula:

$$R_j = \frac{R_{jT}}{N}$$

De esta manera se obtiene que:

$$R_j = \frac{131}{4} = 32.75$$

- Determinación de la desviación media, dada por la diferencia entre cada  $R_j$  y el valor de la media.

$$D_m = R_j - R_j$$

Fórmula 9: Desviación media

Aspectos	$D_m$
Pregunta 1	0.25

<b>Pregunta 2</b>	-0.75
<b>Pregunta 3</b>	2.25
<b>Pregunta 4</b>	-1.75

**Tabla 3.9:** Desviaciones medias de cada aspecto

- Determinación de la suma de los cuadrados de las desviaciones medias (S).

$$S = \sum_{j=1}^n D_m^2$$

**Fórmula 10:** Suma de los cuadrados de S

- Determinación del cuadrado del número total de expertos  $K$ . En este caso  $K = 7$
- Determinación del cubo del número total de aspectos a evaluar  $N$ . En este caso  $N = 4$ .
- Determinación de la diferencia entre  $N^3$  y  $N$ , y su multiplicación por el cuadrado de  $K$ .

Con todos los parámetros de  $W$  calculados se procede a encontrar el valor de este. La herramienta utilizada devolvió los siguientes resultados:

<b>Parámetros</b>	<b>Valor</b>
<b>S</b>	8.8
$K^2$	49
$N^3$	64
$N^3 - N$	60
<b>W</b>	<b>0.0357</b>

**Tabla 3.10:** Cálculo de W

Los valores del  $W$  deben oscilar entre 0 y 1 ( $0 < W < 1$ ). Después de obtener el valor del  $W$  se debe probar si existe concordancia entre los expertos para lo cual se hace uso del Chi-Cuadrado, este se calcula mediante la fórmula:

$$X^2 = K N - 1 W$$

**Fórmula 11:** Cálculo del Chi-Cuadrado

Al realizar este cálculo se obtiene como resultado un valor igual a 1. Este valor se compara con el de las Tablas Estadísticas. Se busca el Chi-Cuadrado tabulado ( $X^2_{\alpha, n-1}$ ) en la tabla del percentil de esta distribución con un nivel de significación  $\alpha = 0.9$  que representa un 85% de confianza y  $n-1$  grados de libertad.

La interpretación del resultado se realiza de la siguiente manera. Si  $X^2_{real} < X^2_{\alpha, n-1}$  entonces existe concordancia en los criterios de los expertos. Como  $X^2_{real} = 1$ , se cumple que  $X^2_{real} < X^2_{0.9, 4}$  ya que  $1 < 7.78$ . Con lo realizado anteriormente se comprueba la concordancia entre los expertos en cuanto a la aceptación de la base de casos propuesta.

3.3.4 Desarrollo práctico y explotación de los resultados.

Los expertos que conformaron el panel recibieron una encuesta en la cual se incluyó un diagrama donde quedó reflejado la estructura de la base de casos propuesta como solución. Se aplican un total de cuatro preguntas que responden a los objetivos de la validación. Las entrevistas fueron aplicadas personalmente a los expertos lo cual permitió explicarles detalladamente la encuesta para un mejor entendimiento de la misma. Se realizó solo una ronda de preguntas y se analizaron los resultados de las opiniones emitidas. Luego de aplicar la encuesta se procesar toda la información recopilada. La valoración dada por los expertos a cada elemento se muestra en la siguiente tabla:

Expertos	Aspectos			
	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4
$E_1$	BA	BA	MA	BA
$E_2$	MA	MA	MA	MA
$E_3$	MA	MA	MA	MA
$E_4$	MA	MA	MA	MA
$E_5$	BA	MA	MA	MA
$E_6$	MA	MA	MA	BA
$E_7$	MA	A	MA	A

Tabla 3.11: Criterios dados por los expertos a cada aspecto

La información obtenida se procesa en la herramienta CEE, para ello se siguen los siguientes pasos:

**Paso 1:** Agrupar los resultados anteriores en una sola tabla de doble entrada: en las filas de esta se sitúan los aspectos sometidos a consideración y en las columnas cada una de las categorías evaluativas utilizadas. Estas categorías en la herramienta se ubican de modo ascendente, de la menor categoría de evaluación a la mayor.

Aspectos a evaluar	Inadecuado	Poco adecuado	Adecuado	Bastante adecuado	Muy adecuado
Pregunta 1	0	0	0	2	5
Pregunta 2	0	0	1	1	5
Pregunta 3	0	0	0	0	7
Pregunta 4	0	0	1	2	4

Figura 8: Frecuencias absolutas para cada pregunta de la encuesta

**Paso 2:** Determinar, sobre la base de la tabla anterior, la distribución de frecuencia absoluta acumulativa de cada fila. Para ello a cada valor de la fila, excepto el primero, se le suma el anterior.

Aspectos a evaluar	Inadecuado	Poco adecuado	Adecuado	Bastante adecuado	Muy adecuado
Pregunta 1	0	0	0	2	7
Pregunta 2	0	0	1	2	7
Pregunta 3	0	0	0	0	7
Pregunta 4	0	0	1	3	7

Figura 9: Frecuencias absolutas acumulativas para cada pregunta de la encuesta

**Paso 3:** Determinar, a partir de la tabla anterior, la distribución de frecuencias relativas acumulativas de cada fila y eliminar del proceso de análisis la última columna (en la que todas las frecuencias toman el valor de 1).

Aspectos a evaluar	Inadecuado	Poco adecuado	Adecuado	Bastante adecuado
Pregunta 1	0	0	0	0.28571429848671
Pregunta 2	0	0	0.142857149243355	0.28571429848671
Pregunta 3	0	0	0	0
Pregunta 4	0	0	0.142857149243355	0.428571432828903

Figura 10: Frecuencias relativas acumulativas para cada pregunta de la encuesta

**Paso 4:** Calcular, sobre la base de las frecuencias del paso 3, los percentiles de la distribución normal estándar correspondientes a cada una de las frecuencias relativas acumulativas.

Aspectos a evaluar	Inadecuado	Poco adecuado	Adecuado	Bastante adecuado
Pregunta 1	-3.09	-3.09	-3.09	-0.565948784978835
Pregunta 2	-3.09	-3.09	-1.06757049672431	-0.565948784978835
Pregunta 3	-3.09	-3.09	-3.09	-3.09
Pregunta 4	-3.09	-3.09	-1.06757049672431	-0.180012358854845

Figura 11: Percentiles correspondientes a la distribución normal

**Paso 5:** Calcular la suma algebraica de todos los percentiles anteriores y a dividirla por el producto de la cantidad de aspectos sometidos a consulta y la cantidad de categorías evaluativas empleadas. A este valor resultante le llamará N.

$$N = \frac{\sum_1^n \sum_1^m Per}{n * m}$$

Fórmula 12: Media general de los percentiles

Luego de aplicar la fórmula 12 el valor de N da como resultado -2.33982

**Paso 6:** Determinar la media de los percentiles de cada categoría evaluativa (columnas) obtenidos en el paso 4. A estos valores promedio se les llama puntos de corte. Estos se utilizan para determinar la categoría de cada criterio según la opinión de los expertos consultados.

$$PC = \frac{\sum_1^n Per}{n}$$

Fórmula 13: Puntos de corte

**Paso 7:** Determinar la media de los percentiles de cada aspecto o indicador sometido a consulta (filas). A este promedio se le llama P.

$$P = \frac{\sum_{i=1}^m Per_i}{m}$$

**Fórmula 14:** Promedio de los percentiles de cada indicador

**Paso 8:** Obtener las diferencias N-P para cada aspecto o indicador analizado (filas). Estos valores (uno por uno) se comparan con los puntos de corte y se determina en qué categoría evaluativa se encuentran.

Los resultados de aplicar los pasos del 5 al 8 se resumen en la tabla siguiente (obtenidos en la herramienta CEE):

No.	Elementos	I	PA	A	BA	P	N-P	Categoría
1	Pregunta 1	-3.09	-3.09	-3.09	-0.56595	-2.45899	0.58713	MA
2	Pregunta 2	-3.09	-3.09	-1.06757	-0.56595	-1.95338	0.08153	MA
3	Pregunta 3	-3.09	-3.09	-3.09	-3.09	-3.09	1.21815	MA
4	Pregunta 4	-3.09	-3.09	-1.06757	-0.18001	-1.8569	-0.01496	MA
<b>PC</b>		<b>-3.09</b>	<b>-3.09</b>	<b>-2.07879</b>	<b>-1.10048</b>			

**Tabla 3.12:** Resultados de aplicar los pasos del 5 al 8

Se puede apreciar en el resultado obtenido que todas las preguntas, con los criterios de los expertos en conjunto, fueron calificadas de muy adecuado.

### 3.3.5 Resultados de la validación de la base de casos propuesta.

Para concluir la validación de la BC se realiza un resumen de los principales resultados obtenidos durante todo el proceso.

#### Selección de los expertos.

A continuación se muestra una figura donde se pueden apreciar el nivel de competencia de los expertos obtenidos de la encuesta de autovaloración.

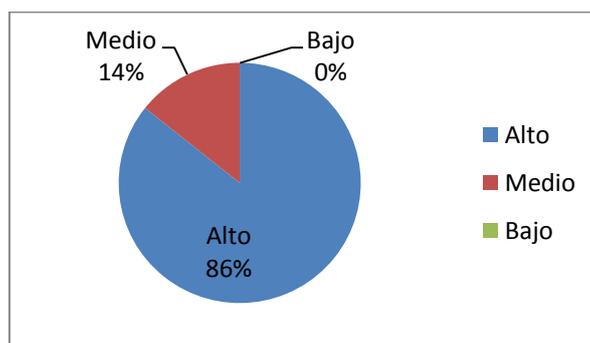


Figura 12: Competencia de los expertos

En la misma se puede constatar que el 86 % de los encuestados poseen un nivel de competencia Alto y el 14 % restante poseen un nivel medio, lo que significa que el 100 % poseen el nivel de competencia suficiente para validar la propuesta.

**Tabulación de los resultados en las preguntas.**

El resultado que se obtiene luego de analizar estadísticamente las respuestas dadas por los expertos se puede observar en la Figura 13. En esta se puede corroborar que en todas las preguntas prevalecen las calificaciones de MA y ninguna de PA e I, lo que demuestra que la BC propuesta cumple con los requerimientos necesarios para ponerla en práctica.

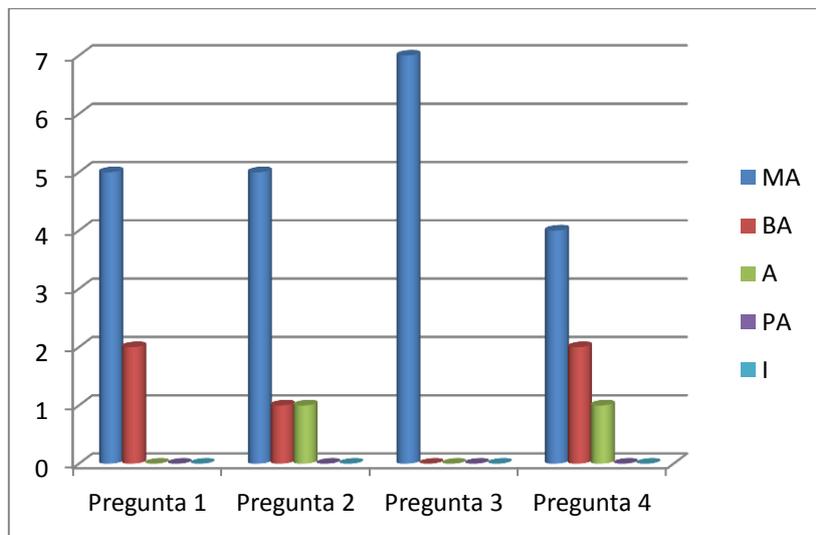


Figura 13: Evaluaciones dadas a cada pregunta por los expertos

**3.4 Validación del motor de inferencia.**

Inicialmente la base de casos contendrá información obtenida de exámenes aplicados en años anteriores, en los cuales no se evalúan todos los elementos seleccionados como rasgos predictores en la presente investigación. Para tratar esta situación existe una técnica simple, pero poco recomendable, que consiste en eliminar aquellos casos que tengan atributos sin valor. El mayor problema de esta técnica es que se podría eliminar información importante. Para mantener los casos en la base de casos sería necesario rellenar el atributo faltante con un valor válido.

Para el desarrollo de este trabajo se propone completar los valores ausentes con el valor moda. La moda es el valor que tiene mayor frecuencia de aparición y se representa por M. Se puede hallar la moda para variables cualitativas y cuantitativas. En la solución propuesta la moda sería la evaluación que más veces ha sido otorgada dentro de una misma clase.

Luego de completar los datos y a partir de la base de casos confeccionada y evaluada por el criterio de los expertos se procede a probar el funcionamiento del motor de inferencia. Para ello se escogen

cinco casos que conforman el conjunto de prueba, estos a su vez han sido clasificados por los expertos para comprobar los resultados del motor de inferencia desarrollado. Los elementos del conjunto de prueba no forman parte del conjunto de entrenamiento. Este último está formado por los casos de la base de casos. Los casos seleccionados son los siguientes:

Caso nuevo	Clasificación dada por los expertos
C1	Alto
C2	Medio
C3	Bajo
C4	Sin nivel
C5	Sin nivel

Tabla 3.13: Casos clasificados por los expertos

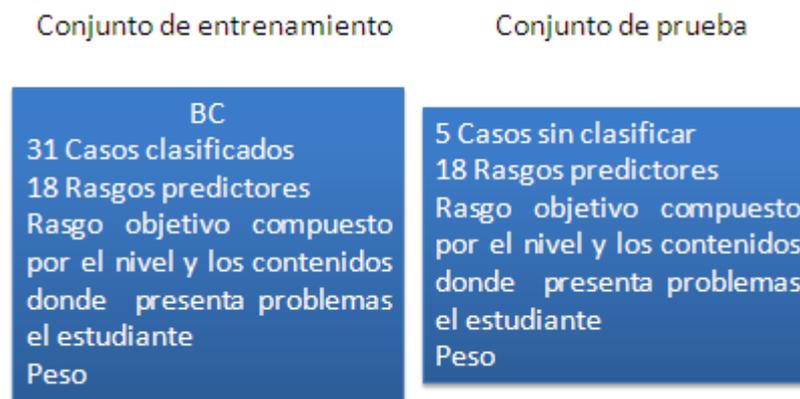


Figura 14: Conjuntos de casos definidos para validar el motor de inferencia

Para la validación del motor de inferencia se desarrolla una aplicación que implementa los algoritmos y funciones definidos en las fases del motor de inferencia. La fase de recuperación recibe, como entrada, los resultados de los estudiantes de cada ejercicio realizado. Como salida se obtiene el nivel del estudiante así como los contenidos en los cuales presenta problemas.

#### Validación de la fase de recuperación.

En esta fase se recuperan los casos más semejantes aplicando el algoritmo propuesto, el cual tiene en cuenta el umbral de semejanza. Para el cálculo del umbral se utiliza la matriz de semejanza entre los casos de la base de casos obteniéndose como resultado un valor de 0.58 (Anexo 5). A continuación son recuperados los casos cuyos valores de semejanza superan el umbral. Estos casos se pueden apreciar en el

Anexo 6.

### Validación de la fase de adaptación.

Esta fase recibe como entrada los casos recuperados. Primeramente es importante verificar si el caso más semejante es un fallo o no ya que si es fallo no se tiene en cuenta y se busca el siguiente caso más semejante. Luego se analiza que tipo de adaptación utilizar en dependencia de la situación. Después de buscar si hay más de un caso que tiene como semejanza el máximo valor de esta se pudo constatar que para los casos evaluados sólo hay un caso más semejante. Debido a esto a los casos del conjunto de prueba se les asignó como solución la del caso más semejante, realizándose así la adaptación nula. El resultado de esta fase se puede observar en la siguiente tabla:

Caso nuevo	Caso seleccionado	Nivel	Nivel dado por el experto	Contenidos a estudiar
C1	Caso 1	Alto	Alto	Ninguno
C2	Caso 1	Alto	Medio	Interrelaciones
C3	Caso 15	Bajo	Bajo	Entidades y atributos-Interrelaciones-Patrones
C4	Caso 24	Sin nivel	Sin nivel	Entidades y atributos-Interrelaciones-Patrones
C5	Caso 28	Sin nivel	Sin nivel	Entidades y atributos-Interrelaciones-Patrones

Tabla 3.14: Casos resueltos

### Validación de la fase de revisión.

Esta fase recibe los casos resueltos en el módulo de adaptación para que el experto se encargue de comprobarla solución dada. Para ello si considera que la solución propuesta no es correcta califica el caso resuelto como fallo, de esta manera otro caso con características similares a este no se le daría la misma solución, evitando así que cometa el mismo error dos veces. De ser correcta la evaluación dada el caso se mantendría tal cual llegó. Al analizar los casos resueltos se puede apreciar que el caso identificado como **C2** el nivel no es el correcto por lo que este se califica como fallo. Los demás se mantienen igual.

### Validación de la fase de retención.

En esta fase se almacenan los casos revisados. Para esto primero hay que tener en cuenta si la semejanza del caso revisado con el caso recuperado más semejante es de 1, ya que esto implicaría que sea exactamente el mismo caso por lo que no aporta información nueva y no se debe almacenar. Para el conjunto de prueba se almacenaron todos los casos debido a que ninguno cumple con la condición anteriormente mencionada.

### 3.5 Conclusiones parciales.

En el presente capítulo se valida la base de casos propuesta mediante el método Delphi, el cual permitió demostrar la existencia de concordancia entre los expertos. Para esto fueron seleccionados

como expertos los profesores que más experiencia presentan en la asignatura Sistema de Base de Datos 1, específicamente en el diseño conceptual de base de datos. La propuesta fue aceptada por la mayoría de los expertos, lo cual se evidencia en las opiniones dadas. Para el motor de inferencia fue desarrollada una aplicación con la cual se validó los algoritmos y funciones definidos para cada una de sus fases.

## Conclusiones

Con el desarrollo del presente trabajo investigativo se llegó a las siguientes conclusiones:

- Se analizó la técnica de IA sistema basado en el conocimiento y se escogió el razonamiento basado en casos como propuesta de solución.
- Fue realizado un estudio del estado del arte de la situación existente en el mundo y el país concerniente a la enseñanza del diseño de bases de datos y sistemas inteligentes que utilizan la técnica razonamiento basado en caso.
- Fueron seleccionados como rasgos predictores los elementos que componen el diseño conceptual de base de datos.
- Se estudiaron las formas de acceder a la BC y se escogió la estructura plana. Además se propuso una función de semejanza. que permite saber cuán parecido es un caso a otro teniendo en cuenta el tipo de información almacenada.
- Se validó la base de casos propuesta como solución mediante el método Delphi ,obteniendo resultados satisfactorios debido a que los expertos coinciden en que la aplicación de la base de casos puede ayudar a mejorar el proceso del enseñanza-aprendizaje del diseño de base datos en la UCI. El motor de inferencia se probó mediante una aplicación desarrollada por los investigadores, dando por resultado la acertada clasificación de los casos del conjunto de prueba.

## **Recomendaciones**

Se recomienda:

- Incluir funciones de semejanza que tengan en cuenta la incertidumbre ya que de esta manera se pueda realizar una mejor orientación del estudiante en cuanto a los contenidos en los cuales tiene problemas.
- Definir algoritmos que consigan “olvidar” la información que no se utiliza y de esta forma aligerar la cantidad de información a procesar por el razonador.
- Implementar un Sistema Inteligente que utilice la BC y MI propuestos.

**Trabajos Citados**

1. UCI, *Modelo del Profesional*.
2. UCI, *Informe Semestral. Asignatura Sistemas de Bases de Datos. Departamento de Ingeniería y Gestión de Software*. 2010.
3. UCI, *Informe Semestral. Asignatura Sistemas de Bases de Datos 1. Departamento de Ingeniería y Gestión de Software*. 2011.
4. UCI, *Informe Semestral. Asignatura Sistemas de Bases de Datos 1. Departamento de Ingeniería y Gestión de Software*. 2012.
5. García, R.M.M., *Diseño de Base de Datos*. 1999.
6. *INTRODUCCIÓN A LOS SISTEMAS DE BASES DE DATOS*.
7. *Modelos Conceptuales de Datos*.
8. Torres, G.R.y.A.R.D.; Available from: <http://www.slideshare.net/amruiz/modelos-de-datos-y-modelado-conceptual-presentation-663634>.
9. *MODELO ENTIDAD RELACIÓN (MER). EXTENSIONES DEL MER*.
10. Abraham Silberschatz, H.F.K., S. Sudarshan, *FUNDAMENTOS DE BASES DE DATOS*. 2002.
11. Arrúa, L.y.M.F., Eduardo, *Inteligencia Artificial. Sistemas Expertos y Redes Neuronales*. 2003.
12. Gálvez, D., *Curso de Sistemas Basados en el Conocimiento*. 1998.
13. Cordero, D.y.T., Yoanny, *Sistema Inteligente de Mitigación de Riesgos para el Centro ISEC*. 2011, UCI: La Habana.
14. Reyna Pupo, K., *Propuesta de motor de inferencia de una base de casos para la estimación de la duración de un proyecto de desarrollo de software*. 2012, UCI: La Habana.
15. Laura Lozano, J.F., *Razonamiento Basado en Casos, en Una Visión General*. 2006, Universidad de Valladolid. p. 59.
16. Felgar, *Optimización de redes bayesianas basado en técnicas de aprendizaje por inducción*. 2005, Universidad de Buenos Aires.
17. Robles, D.M., *Diseño de un sistema inteligente para la prevención del delito robo en domicilios*. 2012, UCI: Habana. p. 28-29.
18. Aamodt, A.y.P., E, *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*. *AI Communications*. 1998. **7:1**: p. 39-59.
19. Bergmann, R., *Developing Industrial Case Based Reasoning Applications. The INRECA Methodology*. 1999: Berlin.
20. Rafael Esteban Bello Pérez, Z.Z.G.V., María M. García Lorenzo, Antonio Reynoso Lobato, *Aplicaciones de la Inteligencia Artificial*. Primera edición, ed. 2002, Guadalajara.Jalisco.México: Ediciones de la Noche.

21. Gálvez Lio, D., *Curso Sistemas Basados en el Conocimiento (SBC)*, in *Departamento de Ciencia de la Computación* 1998, Universidad Central de Las Villas: Santa Clara.
22. Abreu Jiménez, P.C.y.D.S.C., Jeny, *Base de Casos para estimar la duración de un proyecto de desarrollo de software*. 2012, UCI: La Habana.
23. Turban, E., *Decision Support Systems and Intelligent Systems*. 5 ed. 1998, New Jersey.
24. Dietrick S.W., E.E., Piscator K, *WinRDBI: a Windows-based relational database educational tool*. , in *28th Technical Symposium on Computer Science Education. SIGCSE*. 1997.
25. A, G., *WebSQL: An Interactive Web Tool for Teaching Structured Query Language*, in *Conference on Information Systems*. 2000.
26. A, M., *Learning SQL with a computerized tutor*, in *29th Technical Symposium on Computer Science Education. SIGCSE*. 1998.
27. A, M., *NORMIT: a Web-Enabled Tutor for Database Normalization.*, in *International Conference on Computers in Education*. 2002.
28. Josep Soler, F.P., Imma Boada, Jordi Poch, *Utilización de una plataforma de e-learning en la docencia de bases de datos.*, in *Dpto. de Informática y Matemática Aplicada.*, Universidad de Girona.
29. R. Bertone, P.T., S. Antonetti, A. Miglio, *Herramienta para la enseñanza de Modelado Conceptual de Bases de Datos*, in *XV Congreso Argentino de Ciencias de la Computación*. 2009: Universidad Nacional de Jujuy.
30. Rodolfo Bertone, E.N., Pablo Thomas., *HEA: Herramienta de Software para enseñanza de árboles*. , in *XVI Congreso Argentino de Ciencias de la Computación*. 2010: Universidad de Morón.
31. Rodolfo Bertone, P.T., Durán Alejandra, Rius María Florencia., *CasER 2.0: Herramienta para la enseñanza de Bases de Datos*. , in *XVII Congreso Argentino de Ciencias de la Computación*. 2011: Universidad Nacional de la Plata.
32. W. Evens, M.B., Stefan; Chang, Ru-Charn; Freedman, Reva; Glass, Michael; Hee Lee, Yoon; Seop Shim, Leem; Woo Woo, Chong; Zhang, Yuemei; Zhou, Yujian; A. Michael, Joel; A. Rovick, Allen, *CIRCSIM-Tutor: An Intelligent Tutoring System Using Natural Language Dialogue*, in *12th Midwest AI and Cognitive Science Conference*. 2001: Oxford OH. p. 16-23.
33. Cataldi, Z.y.L., Fernando J, *SISTEMAS TUTORES INTELIGENTES ORIENTADOS A LA ENSEÑANZA PARA LA COMPRENSIÓN*, in *EDUTECH. Revista Electrónica de Tecnología educativa*. 2009.
34. Rojas Correa, Y.A.y.A.M., T, *MENTOR: Sistema Tutorial Inteligente para el desarrollo de habilidades en la solución de problemas matemáticos*, in *Revista de Investigación Universidad La Salle*. 2007. p. 235-246.

35. González Daniry, O.R.V., Reinaldo, *Análisis y Diseño del Tutor Virtual de Evaluación para el Aprendizaje Autónomo de Idiomas*. 2010, UCI: La Habana.
36. Bregón, A., y otros, *Un sistema de razonamiento basado en casos para la clasificación de fallos en sistemas dinámicos*, in *III Taller Nacional de Minería de Datos y Aprendizaje*. 2005: Valladolid. p. 203-211.
37. Dapena, M.D.D., *Utilización del Razonamiento Basado en Casos en las Revisiones de la Definición del Modelo de Negocio*, in *Centro de Estudios de Ingeniería de Sistemas (CEIS)*, Instituto Superior Politécnico José Antonio Echeverría (ISPJAE). Habana.
38. Martínez Sánchez, N., y otros, *Representación y organización del conocimiento en Software educativo*. 2009.
39. Morales Rodríguez, L.I.y.d.A.F., Yarisleydi, *Sistema Tutorial Inteligente de apoyo al proceso de enseñanza-aprendizaje de la asignatura de Sistemas Operativos*. 2012, UCI: La Habana.
40. Yordi Chaveco Bustamante, E.R.M., *Sistema Web Inteligente apoyado en Mapas Conceptuales para la asignatura de Álgebra Lineal*. 2012, UCI: Habana.
41. Gonzáles Rodríguez, Y.y.R.B., Roberto, *Modelo Basado en Casos para la planificación de proyectos de software*. 2009, UCI: La Habana.
42. Han, J.y.K., Micheline, *Data Mining: Concepts and Techniques*, in *Department of Computer Science*. 2001, University of Illinois: .
43. González, Y.G., *Base de conocimiento y motor de inferencia de un Sistema Basado en Conocimiento para la obtención de los requisitos de software en la línea temática de atención y tratamiento de las emergencias en el centro ISEC*. 2012, UCI: La Habana.
44. Bergmann, R.a.G., Mehmet (1999) *Developing Industrial Case Based Reasoning Applications. The INRECA Methodology*.
45. Múndiz Noguera, I., *UN ESTUDIO SOBRE LA APLICACION DEL RAZONAMIENTO BASADO EN CASOS A LA CONSTRUCCION DE PROGRAMAS*, in *Departamento de Informática y Automática*. 1992, UNIVERSIDAD COMPLUTENSE DE MADRID Facultad de Ciencias Fisicas: Madrid.
46. Almaguer, A.G., *EL MÉTODO DELPHI Y EL PROCESAMIENTO ESTADÍSTICO DE LOS DATOS OBTENIDOS DE LA CONSULTA A LOS EXPERTOS*, ISP "José de la Luz y Caballero".

**Bibliografía**

Aamodt, A., & Plaza, E. (1994). *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*. *AI Communications* (Vol. 7:1).

Abreu Jiménez, P. C., & Del Sol Chang, J. (junio de 2012). Base de Casos para estimar la duración de un proyecto de desarrollo de software. *Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas*. Habana.

Abreu Jiménez, P. C., & Del Sol Chang, J. (2012). Base de Casos para estimar la duración de un proyecto de desarrollo de software. *Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas*. Habana.

Altho, K., & Bartsch-Spörl, B. (1996). *Decision Support for Case-Based Applications*.

BAÑOS SANCHO, J. (2007). LA PLATAFORMA EDUCATIVA MOODLE. *MANUAL DE CONSULTA PARA EL PROFESORADO*. Getafe.

Bello, R. (2002). *Aplicaciones de la Inteligencia Artificial*. Guadalajara.Jalisco.México: Ediciones de la Noche.

Bergmann, R. (1999). *Developing Industrial Case Based Reasoning Applications. The INRECA Methodology*. Berlin, Alemania.

Bregón, A., Simón, A., Alonso, C., Pulido, B., Moro, I., & Rodríguez, J. J. (2005). Un sistema de razonamiento basado en casos para la clasificación de fallos en sistemas dinámicos. 203-211. Valladolid.

Carlos Leopoldo. (7 de Octubre de 2007). *techtástico*. Obtenido de [techtastico.com/post/software-para-disenar-bases-de-datos/](http://techtastico.com/post/software-para-disenar-bases-de-datos/)

Cataldi, Z., & Lage, F. J. (2009). SISTEMAS TUTORES INTELIGENTES ORIENTADOS A LA ENSEÑANZA PARA LA COMPRENSIÓN. (28).

Cepeda Mateos, R. D., Marco De Francisco, C., & Serrano Gordillo, T. (2009). Obtenido de <http://eprints.ucm.es/9741/1/DBCASE.pdf>

Cepeda Mateos, Rodrigo Denís y Marco de Francisco, Cristina y Serrano Gordillo, Tello. (2009). Obtenido de E-Prints Complutense: <http://www.eprints.ucm.es/9741>

*Conferencia #2 de Reconocimiento de Patrones*. (s.f.). Recuperado el 10 de abril de 2013, de [out.uclv.edu.cu/fcs/materiales/Conferencia%202%20RP.doc](http://out.uclv.edu.cu/fcs/materiales/Conferencia%202%20RP.doc)

Díaz Moreno, J. M. (1998). *Introducción a la topología de los espacios métricos*. Servicio de publicaciones de la Universidad de Cádiz.

Durán, A. (2002). Recuperado el 2013, de <http://www.lsi.us.es/~informes/lsi-2000-10.pdf>.

Gálvez Lio, D. (1998). Curso Sistemas Basados en el Conocimiento (sistemas basados en el conocimiento).

García Martínez, R., & López Nocera, M. U. (2009). *Descubrimiento de conocimiento basado en la integración de algoritmos de explotación de la información*. Buenos Aires.

González Rodríguez, Y., & Rodríguez Bombalé, R. (2009). Modelo Basado en Casos para la planificación de proyectos de software. *Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas*. La Habana.

González Daniry, Ortega; Rodríguez Veitia, Reinaldo. (2010). Análisis y Diseño del Tutor Virtual de Evaluación para el Aprendizaje Autónomo de Idiomas. Habana.

Guerra Hernández, M. (2012). Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas. *Sistema de Investigación Criminal para la creación de perfiles forenses. Modelamiento de la base de casos*. La Habana.

Gutiérrez Martínez, I., Tellería Rodríguez, A., & Bello Pérez, R. E. (2002). UN SISTEMA BASADO EN CASOS PARA LA TOMA DE DECISIONES EN CONDICIONES DE INCERTIDUMBRE. 23(2) . Santa Clara: REVISTA INVESTIGACION OPERACIONAL.

Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques*. USA: Academic Press.

Leopoldo, C. ( 7 de Octubre de 2007). Obtenido de <http://www.techastico.com/post/software-para-disenar-bases-de-datos>

Lozano, L. (2011). *Razonamiento Basado en Casos: Una Visión General*. Valladolid, España.

Lozano, L., & Fernández, J. (2005). *Razonamiento Basado en Casos: Una Visión General*.

Martínez Sánchez, N. (2009). Modelo para diseñar Sistemas de Enseñanza-Aprendizaje Inteligentes utilizando el Razonamiento Basado en Casos. *Tesis presentada en opción del grado científico de Doctor en Ciencias Técnicas*. Santa Clara.

Martínez Sánchez, N., Ferreira Lorenzo, G., García Lorenzo, M. M., & García Valdivia, Z. (2008). El Razonamiento Basado en Casos en el ámbito de la Enseñanza/Aprendizaje. 5 (Revista de Informática Educativa y Medios Audiovisuales).

Martínez Sánchez, N., G.V, Z., León Espinosa, M., de Medina Sotolongo, D., Lezcano Brito, M., Ferreira Lorenzo, G., y otros. (2009). Representación y organización del conocimiento en Software educativo. (*VIR 167*) .

Móndiz Noguera, I. (1992). *UN ESTUDIO SOBRE LA APLICACION DEL RAZONAmotor de inferenciaENTO BASADO EN CASOS A LA CONSTRUCCION DE PROGRAMAS*. UNIVERSIDAD COMPLUTENSE DE MADRID Facultadde Ciencias Fisicas, Departamento de Informática y Automática, Madrid.

Morales Rodríguez, L. I., & de Avila Fernández, Y. (Junio de 2012). Sistema Tutorial Inteligente de apoyo al proceso de enseñanza-aprendizaje de la asignatura de Sistemas Operativos. *Trabajo de diploma para optar por el título de Ingeniero en Ciencias Informáticas* . Habana.

Morales, E. (1999). *Razonamiento Basado en Casos (CBR)*. Obtenido de <http://aprenderonline.freesevers.com/cbr.html>

Ortega González, D., & Reinaldo, R. V. (2010). Tesis de Diploma para optar por el Título de Ingeniero en Ciencias Informáticas. *Análisis y Diseño del Tutor Virtual de Evaluación para el Aprendizaje Autónomo de Idiomas* . La Habana, Cuba.

Reyna Pupo, K. (Junio de 2012). Propuesta de motor de inferencia de una base de casos para la estimación de la duración de un proyecto de desarrollo de software. *Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas* . La Habana.

Rojas Correa, Y. A., & A. M., T. (julio-diciembre de 2007). MENTOR: Sistema Tutorial Inteligente para el desarrollo de habilidades en la solución de problemas matemáticos. *Revista de Investigación Universidad La Salle* , 235-246.

Schulz, G. (2007). *Un ambiente integrado de clasificación, selección y ponderación de reglas basado en sistemas inteligentes*. Buenos Aires.

Shulcloper, J. R., Alba Cabrera, E., & Lazo Cortés, M. (1995). Introducción al Reconocimiento de Patrones (Enfoque Lógico-Combinatorio). ( *Serie Verde No. 51* ) . México: CINVESTAV-IPN.

Simmons, R. (1988). A Theory of Debbuging. En J. Kolodner, *Proceedings of the First Case-Based Reasoning Workshop* (págs. 388-401). Los altos, CA: Morgan Kaufmann Publishers.

Simpson, R. (1985). *A Computer Model of Case-Based Reasoning in Problem Solving:An Investigation in the Domain of Dispute Mediation*. School of Information and Computer Science, Georgia Institute of Technology. Atlanta: Technical Report GIT-ICS-85/18.

Soria Francis, S. (2010). Modelo para diseñar Mapas Conceptuales Inteligentes utilizando el Razonamiento Basado en Casos. *TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE INGENIERO EN CIENCIAS INFORMÁTICAS* . La Habana.

Sussman, G. (1975). *A Computer Model of Skill Acquisition* (Vol. 1). Estados Unidos: AmericanElsevier.

Turban, E. (1998). *Decision Support Systems and Intelligent Systems* (Quinta ed.). New Jersey, Estados.

Valle, J. (2005). *monografias.com*. Recuperado en Enero de 2013, de <http://www.monografias.com/trabajos24/herramientas-case/herramientas-case.shtml#ejempl>

Villena Román, J., Crespo García, R. M., & García Rueda, J. J. *Inteligencia en Redes de Comunicaciones. Sistemas basados en el conocimiento*. Madrid.

W. Evens, M., Brandle, S., Chang, R.-C., Freedman, R., Glass, M., Hee Lee, Y., y otros. (Marzo de 2001). *CIRCSIM-Tutor: An Intelligent Tutoring System Using Natural Language Dialogue* , 16-23.

Watson, I. (1997). *Applying Case-Based Reasoning: Techniques for Enterprise Systems*.Morgan Kaufmann Publishers.

## Anexos

### Anexo 1: Modelo de entrevista a expertos para obtener los pesos de los rasgos predictores.

La presente encuesta tiene como objetivo obtener el peso de los rasgos predictores para ello le solicitamos que ordene los elementos en cuanto al orden de prioridad del 1-18 que le daría a estos dentro del diseño de base de datos:

Elementos	Prioridad
Identificar entidades y sus atributos.	
Identificar entidades débiles.	
Identificar llave.	
Identificar atributos multi-valuados.	
Identificar atributos compuestos.	
Identificar atributos derivados.	
Identificar la cardinalidad mínima.	
Identificar la cardinalidad máxima.	
Identificar relaciones n-arias.	
Identificar asociaciones recursivas.	
Identificar relación generalización-especialización.	
Identificar relación de agregación.	
Aplicar patrón árboles fuertemente codificados	
Aplicar patrón árboles simples.	
Aplicar patrón árboles estructurados.	
Aplicar patrón grafo dirigido simple.	
Aplicar patrón Modelo Entidad-Atributo-Valor.	
Aplicar grafo estructurado.	

### Anexo 2: Cálculo de los peso de los rasgos utilizando la técnica panel de experto.

Orden de prioridad dado por los expertos a cada elemento.

Elementos	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$Total(T_i)$
Identificar entidades y sus atributos.	1	1	1	2	1	1	1	8
Identificar entidades débiles.	12	9	8	4	8	4	8	53
Identificar llave.	2	2	3	1	9	5	4	26
Identificar atributos multi-valuados.	11	5	4	8	10	6	9	53
Identificar atributos compuestos.	10	6	12	11	11	7	10	67
Identificar atributos derivados.	9	14	15	12	12	8	11	81

Identificar la cardinalidad mínima.	8	11	9	10	3	11	18	<b>70</b>
Identificar la cardinalidad máxima.	7	3	2	9	4	12	3	<b>40</b>
Identificar relaciones n-arias.	3	7	5	3	2	9	2	<b>31</b>
Identificar asociaciones recursivas.	6	8	6	5	5	10	6	<b>46</b>
Identificar relación generalización-especialización.	5	4	7	7	6	2	5	<b>36</b>
Identificar relación de agregación.	4	12	10	13	7	3	15	<b>64</b>
Aplicar patrón árboles fuertemente codificados	16	13	13	18	13	13	14	<b>100</b>
Aplicar patrón árboles simples.	13	15	11	17	14	14	7	<b>91</b>
Aplicar patrón árboles estructurados.	17	16	14	15	15	15	13	<b>105</b>
Aplicar patrón grafo dirigido simple.	14	17	17	16	16	16	12	<b>108</b>
Aplicar patrón Modelo Entidad-Atributo-Valor.	18	10	16	6	17	17	17	<b>101</b>
Aplicar grafo estructurado.	15	18	18	14	18	18	16	<b>117</b>

De esta manera con los valores de  $T_i$  calculados para cada rasgo se puede proceder a calcular el peso de cada uno de estos rasgos.

Elementos para calcular  $W_i$ .

Elementos	$T_i$	$T_i$	$\frac{T_i}{T_i}$	$\frac{1 - T_i}{T_i}$	$W_i$
Identificar entidades y sus atributos.	8	1197	0.00668	0.99332	<b>0.05843</b>
Identificar entidades débiles.	53	1197	0.04428	0.95572	<b>0.05622</b>
Identificar llave.	26	1197	0.02172	0.97828	<b>0.05755</b>
Identificar atributos multi-valuados.	53	1197	0.04428	0.95772	<b>0.05622</b>
Identificar atributos compuestos.	67	1197	0.05597	0.94403	<b>0.05553</b>
Identificar atributos derivados.	81	1197	0.06767	0.93233	<b>0.05484</b>
Identificar la cardinalidad mínima.	70	1197	0.05848	0.94152	<b>0.05538</b>
Identificar la cardinalidad máxima.	40	1197	0.03342	0.96658	<b>0.05686</b>
Identificar relaciones n-arias.	31	1197	0.02590	0.97410	<b>0.05730</b>
Identificar asociaciones recursivas.	46	1197	0.03843	0.96157	<b>0.05656</b>
Identificar relación generalización-especialización.	36	1197	0.03008	0.96992	<b>0.05705</b>
Identificar relación de agregación.	64	1197	0.05347	0.94653	<b>0.05568</b>
Aplicar patrón árboles fuertemente codificados.	100	1197	0.08354	0.91646	<b>0.05391</b>
Aplicar patrón árboles simples.	91	1197	0.07602	0.92398	<b>0.05435</b>
Aplicar patrón árboles estructurados.	105	1197	0.08772	0.91228	<b>0.05366</b>
Aplicar patrón grafo dirigido simple.	108	1197	0.09023	0.90997	<b>0.05352</b>
Aplicar patrón Modelo Entidad-Atributo-Valor.	101	1197	0.08438	0.91562	<b>0.05386</b>
Aplicar grafo estructurado.	117	1197	0.09774	0.90226	<b>0.05307</b>

### Anexo 3: Modelo de encuesta a profesores para seleccionar expertos

La presente encuesta tiene como finalidad la evaluación de especialistas en el tema del diseño de base datos con el objetivo de seleccionar los que finalmente serán los expertos para validar los indicadores propuestos. Gracias por su colaboración.

Nombre y Apellidos: \_\_\_\_\_

Facultad \_\_\_\_\_

Categoría Docente: \_\_\_\_\_

Cargo que ocupa: \_\_\_\_\_

Años de experiencia en la materia: \_\_\_\_\_

Nombre de la Tesis: Base de casos y motor de inferencia para apoyar la enseñanza del diseño conceptual de bases de datos.

Autores: Henry Cable Hernández y Francisdailin Cobas Sierra

Tutores: Ing. Ariagnis Yero Guevara y Ing. Madelín Haro Pérez

Valore su conocimiento en el diseño de base datos, marque con una X en una escala del 0 al 10.

0	1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/>										

Marque con una X el grado que crea que han influenciado las siguientes fuentes de conocimiento.

Fuentes de argumentación	Grado de influencia de cada una de las fuentes en sus criterios		
	A(Alto)	M(Medio)	B(Bajo)
<b>Análisis teóricos realizados por usted.</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Su experiencia en el tema.</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Trabajos de autores nacionales consultados.</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Trabajos de autores extranjeros consultados.</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Su propio conocimiento del estado del problema en el extranjero.</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Su intuición.</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Anexo 4: Modelo de entrevista a expertos para validar la propuesta de base de casos.**

## Encuesta a expertos para la validación del modelo

Compañero(a):

La presente encuesta forma parte de la aplicación del Método de Valoración de Expertos. Se solicita su valiosa colaboración para evaluar si la propuesta cumple con los objetivos propuestos. Para darle cumplimiento a este objetivo se han elaborado un conjunto de preguntas que permiten medir la eficacia del modelo. Se le garantiza a usted que nadie podrá saber quién es el encuestado además se garantiza que sus opiniones se tendrán en cuenta para la posterior aplicación del modelo propuesto. Valore el grado de factibilidad de cada pregunta:

1. ¿En qué medida usted considera que hay necesidad de una base de casos para dar solución a la problemática planteada?

Muy adecuado.  Bastante adecuado.  Adecuado.  Poco adecuado.  Inadecuado.

2. ¿En qué medida usted cree útil la base de casos propuesta?

Muy adecuado.  Bastante adecuado.  Adecuado.  Poco adecuado.  Inadecuado.

3. ¿Qué nivel de vinculación tiene la base de casos con el diseño conceptual de base de datos?

Muy adecuado.  Bastante adecuado.  Adecuado.  Poco adecuado.  Inadecuado.

4. ¿Cómo usted evalúa de manera general la base de casos propuesta?

Muy adecuado.  Bastante adecuado.  Adecuado.  Poco adecuado.  Inadecuado.

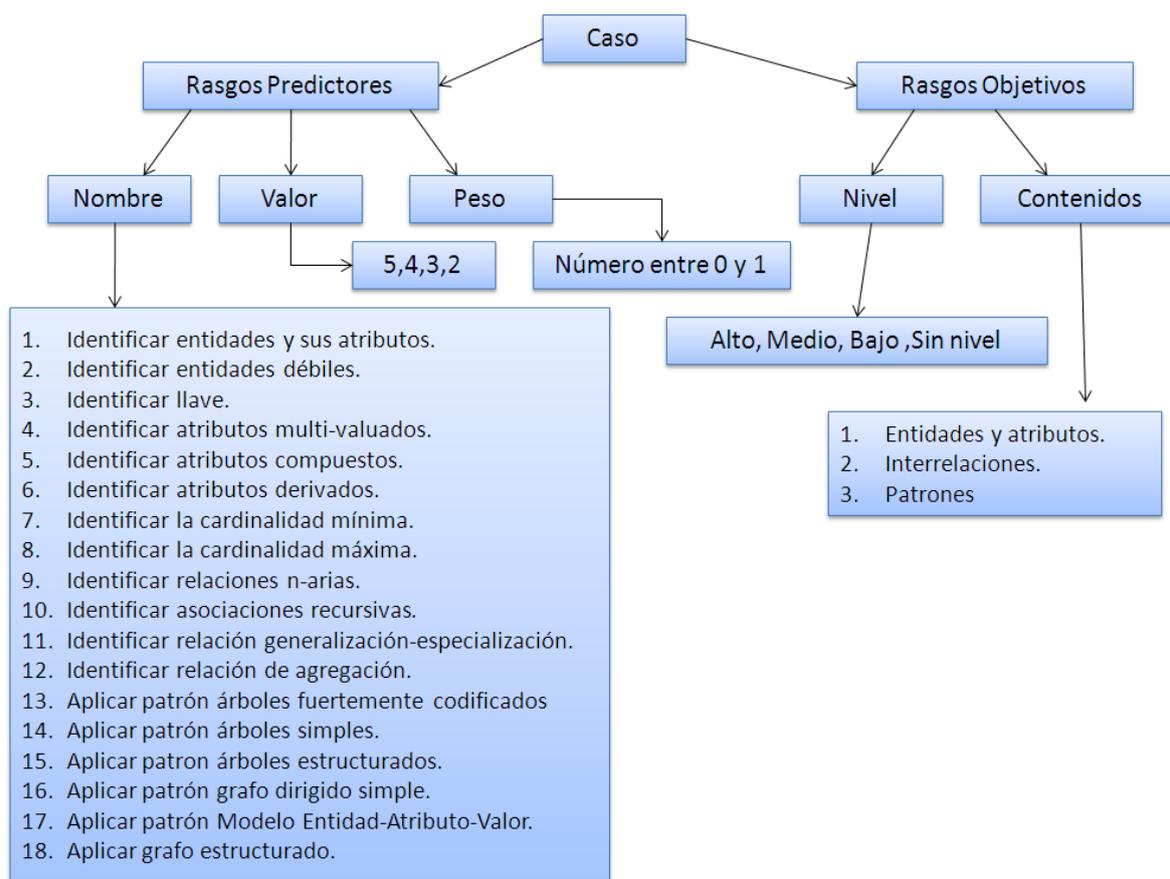


Figura 15: Modelo de la base de casos

**Anexo 5: Matriz de semejanza entre casos para calcular el umbral.**

	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$	$C_9$	$C_{10}$	$C_{11}$
$C_1$	1	0,9436	0,6667	0,9072	0,7918	0,8871	0,8868	0,8311	0,9423	0,9643	0,9253
$C_2$	0,9436	1	0,6851	0,9257	0,8113	0,8307	0,8304	0,7748	0,8859	0,9079	0,869
$C_3$	0,6667	0,6851	1	0,7235	0,648	0,6671	0,6674	0,6482	0,6861	0,7024	0,7413
$C_4$	0,9072	0,9257	0,7235	1	0,7759	0,7943	0,833	0,7383	0,8885	0,8715	0,8325
$C_5$	0,7918	0,8113	0,648	0,7759	1	0,8664	0,905	0,7729	0,7731	0,7561	0,7172
$C_6$	0,8871	0,8307	0,6671	0,7943	0,8664	1	0,9613	0,9065	0,8294	0,8514	0,8125
$C_7$	0,8868	0,8304	0,6674	0,833	0,905	0,9613	1	0,8679	0,8681	0,8511	0,8122
$C_8$	0,8311	0,7748	0,6482	0,7383	0,7729	0,9065	0,8679	1	0,7735	0,7954	0,7565
$C_9$	0,9423	0,8859	0,6861	0,8885	0,7731	0,8294	0,8681	0,7735	1	0,9066	0,9059
$C_{10}$	0,9643	0,9079	0,7024	0,8715	0,7561	0,8514	0,8511	0,7954	0,9066	1	0,8896
$C_{11}$	0,9253	0,869	0,7413	0,8325	0,7172	0,8125	0,8122	0,7565	0,9059	0,8896	1

$C_{12}$	0,7433	0,687	0,6685	0,6505	0,6853	0,7429	0,7426	0,6869	0,7238	0,779	0,7069
$C_{13}$	0,5942	0,5378	0,6317	0,6121	0,6888	0,7071	0,7073	0,6136	0,6136	0,6299	0,6308
$C_{14}$	0,6115	0,5551	0,6112	0,5905	0,666	0,7244	0,6857	0,6693	0,592	0,6115	0,611
$C_{15}$	0,7222	0,6659	0,6491	0,7013	0,7015	0,8351	0,7965	0,8541	0,7028	0,7223	0,6858
$C_{16}$	0,7067	0,6503	0,6669	0,6498	0,611	0,7446	0,706	0,6512	0,7254	0,7424	0,7062
$C_{17}$	0,7967	0,7403	0,6485	0,7039	0,8134	0,8712	0,8709	0,7777	0,7772	0,7967	0,7602
$C_{18}$	0,6468	0,6663	0,6108	0,6278	0,7019	0,7213	0,721	0,7389	0,6273	0,6111	0,6475
$C_{19}$	0,8883	0,8319	0,7042	0,7955	0,6801	0,7754	0,7751	0,7194	0,907	0,8525	0,9629
$C_{20}$	0,6872	0,6308	0,5754	0,7052	0,7071	0,6493	0,688	0,5933	0,6685	0,7229	0,6126
$C_{21}$	0,8145	0,833	0,7785	0,9073	0,7209	0,74	0,7403	0,684	0,7958	0,8144	0,7758
$C_{22}$	0,3333	0,3897	0,6667	0,4261	0,4665	0,4087	0,409	0,4652	0,391	0,3691	0,408
$C_{23}$	0,2974	0,241	0,52	0,2435	0,3908	0,4103	0,4106	0,4662	0,3168	0,3331	0,372
$C_{24}$	0,2607	0,2412	0,5177	0,2797	0,393	0,3352	0,3739	0,3528	0,3184	0,2964	0,3353
$C_{25}$	0,1469	0,2033	0,4803	0,2397	0,3176	0,2598	0,2601	0,3158	0,2046	0,1827	0,2216
$C_{26}$	0,4427	0,4233	0,6291	0,4617	0,5361	0,5556	0,5559	0,6116	0,5004	0,4428	0,4802
$C_{27}$	0,5743	0,5179	0,6101	0,5174	0,4786	0,6122	0,5736	0,6682	0,593	0,5386	0,6118
$C_{28}$	0,168	0,1486	0,4635	0,2229	0,3004	0,2809	0,2812	0,3369	0,2257	0,2038	0,2427
$C_{29}$	0,3913	0,4098	0,6511	0,4123	0,4475	0,4668	0,467	0,4862	0,449	0,4271	0,466
$C_{30}$	0,3567	0,3004	0,5413	0,3747	0,3749	0,4321	0,4324	0,4502	0,4144	0,3925	0,4314
$C_{31}$	0	0,0564	0,3333	0,0928	0,2082	0,1129	0,1132	0,1689	0,0577	0,0357	0,0746

	$C_{12}$	$C_{13}$	$C_{14}$	$C_{15}$	$C_{16}$	$C_{17}$	$C_{18}$	$C_{19}$	$C_{20}$	$C_{21}$	$C_{22}$
$C_1$	0,7433	0,5942	0,6115	0,7222	0,7067	0,7967	0,6468	0,8883	0,6872	0,8145	0,3333
$C_2$	0,687	0,5378	0,5551	0,6659	0,6503	0,7403	0,6663	0,8319	0,6308	0,833	0,3897
$C_3$	0,6685	0,6317	0,6112	0,6491	0,6669	0,6485	0,6108	0,7042	0,5754	0,7785	0,6667
$C_4$	0,6505	0,6121	0,5905	0,7013	0,6498	0,7039	0,6278	0,7955	0,7052	0,9073	0,4261
$C_5$	0,6853	0,6888	0,666	0,7015	0,611	0,8134	0,7019	0,6801	0,7071	0,7209	0,4665
$C_6$	0,7429	0,7071	0,7244	0,8351	0,7446	0,8712	0,7213	0,7754	0,6493	0,74	0,4087
$C_7$	0,7426	0,7073	0,6857	0,7965	0,706	0,8709	0,721	0,7751	0,688	0,7403	0,409
$C_8$	0,6869	0,6136	0,6693	0,8541	0,6512	0,7777	0,7389	0,7194	0,5933	0,684	0,4652

$C_9$	0,7238	0,6136	0,592	0,7028	0,7254	0,7772	0,6273	0,907	0,6685	0,7958	0,391
$C_{10}$	0,779	0,6299	0,6115	0,7223	0,7424	0,7967	0,6111	0,8525	0,7229	0,8144	0,3691
$C_{11}$	0,7069	0,6308	0,611	0,6858	0,7062	0,7602	0,6475	0,9629	0,6126	0,7758	0,408
$C_{12}$	1	0,6667	0,7253	0,6878	0,8148	0,7984	0,6495	0,6698	0,722	0,6689	0,5181
$C_{13}$	0,6667	1	0,6865	0,7595	0,7406	0,7613	0,5767	0,6296	0,7579	0,6294	0,5927
$C_{14}$	0,7253	0,6865	1	0,7079	0,7594	0,7075	0,6286	0,6099	0,6318	0,6832	0,537
$C_{15}$	0,6878	0,7595	0,7079	1	0,7254	0,8518	0,7046	0,6487	0,6636	0,6469	0,5378
$C_{16}$	0,8148	0,7406	0,7594	0,7254	1	0,7618	0,6488	0,7433	0,5727	0,6671	0,5529
$C_{17}$	0,7984	0,7613	0,7075	0,8518	0,7618	1	0,7053	0,7232	0,7037	0,6112	0,4259
$C_{18}$	0,6495	0,5767	0,6286	0,7046	0,6488	0,7053	1	0,6104	0,4449	0,5351	0,5007
$C_{19}$	0,6698	0,6296	0,6099	0,6487	0,7433	0,7232	0,6104	1	0,5755	0,7387	0,445
$C_{20}$	0,722	0,7579	0,6318	0,6636	0,5727	0,7037	0,4449	0,5755	1	0,7236	0,4636
$C_{21}$	0,6689	0,6294	0,6832	0,6469	0,6671	0,6112	0,5351	0,7387	0,7236	1	0,5188
$C_{22}$	0,5181	0,5927	0,537	0,5378	0,5529	0,4259	0,5007	0,445	0,4636	0,5188	1
$C_{23}$	0,4816	0,5567	0,5416	0,4675	0,4441	0,4649	0,4285	0,3709	0,4277	0,3362	0,6292
$C_{24}$	0,4092	0,5566	0,3904	0,4258	0,4432	0,464	0,4277	0,3724	0,4283	0,334	0,7027
$C_{25}$	0,3679	0,4079	0,426	0,3889	0,4043	0,277	0,5001	0,2227	0,3149	0,3324	0,7777
$C_{26}$	0,4102	0,4849	0,5036	0,613	0,4843	0,5383	0,4621	0,4814	0,3907	0,481	0,5932
$C_{27}$	0,5033	0,4293	0,5575	0,6339	0,6525	0,5579	0,5912	0,613	0,2974	0,499	0,536
$C_{28}$	0,4247	0,5379	0,4469	0,4458	0,4254	0,3714	0,4075	0,2438	0,4431	0,3156	0,686
$C_{29}$	0,5001	0,6118	0,4849	0,5228	0,5388	0,5572	0,3746	0,5031	0,4443	0,4296	0,6486
$C_{30}$	0,4293	0,6861	0,4475	0,5586	0,5039	0,4863	0,413	0,4685	0,5174	0,392	0,6485
$C_{31}$	0,2567	0,4058	0,3885	0,2778	0,2933	0,2033	0,3532	0,1117	0,3128	0,1855	0,6667

	$C_{23}$	$C_{24}$	$C_{25}$	$C_{26}$	$C_{27}$	$C_{28}$	$C_{29}$	$C_{30}$	$C_{31}$
$C_1$	0,2974	0,2607	0,1469	0,4427	0,5743	0,168	0,3913	0,3567	0
$C_2$	0,241	0,2412	0,2033	0,4233	0,5179	0,1486	0,4098	0,3004	0,0564
$C_3$	0,52	0,5177	0,4803	0,6291	0,6101	0,4635	0,6511	0,5413	0,3333
$C_4$	0,2435	0,2797	0,2397	0,4617	0,5174	0,2229	0,4123	0,3747	0,0928
$C_5$	0,3908	0,393	0,3176	0,5361	0,4786	0,3004	0,4475	0,3749	0,2082

$C_6$	0,4103	0,3352	0,2598	0,5556	0,6122	0,2809	0,4668	0,4321	0,1129
$C_7$	0,4106	0,3739	0,2601	0,5559	0,5736	0,2812	0,467	0,4324	0,1132
$C_8$	0,4662	0,3528	0,3158	0,6116	0,6682	0,3369	0,4862	0,4502	0,1689
$C_9$	0,3168	0,3184	0,2046	0,5004	0,593	0,2257	0,449	0,4144	0,0577
$C_{10}$	0,3331	0,2964	0,1827	0,4428	0,5386	0,2038	0,4271	0,3925	0,0357
$C_{11}$	0,372	0,3353	0,2216	0,4802	0,6118	0,2427	0,466	0,4314	0,0746
$C_{12}$	0,4816	0,4092	0,3679	0,4102	0,5033	0,4247	0,5001	0,4293	0,2567
$C_{13}$	0,5567	0,5566	0,4079	0,4849	0,4293	0,5379	0,6118	0,6861	0,4058
$C_{14}$	0,5416	0,3904	0,426	0,5036	0,5575	0,4469	0,4849	0,4475	0,3885
$C_{15}$	0,4675	0,4258	0,3889	0,613	0,6339	0,4458	0,5228	0,5586	0,2778
$C_{16}$	0,4441	0,4432	0,4043	0,4843	0,6525	0,4254	0,5388	0,5039	0,2933
$C_{17}$	0,4649	0,464	0,277	0,5383	0,5579	0,3714	0,5572	0,4863	0,2033
$C_{18}$	0,4285	0,4277	0,5001	0,4621	0,5912	0,4075	0,3746	0,413	0,3532
$C_{19}$	0,3709	0,3724	0,2227	0,4814	0,613	0,2438	0,5031	0,4685	0,1117
$C_{20}$	0,4277	0,4283	0,3149	0,3907	0,2974	0,4431	0,4443	0,5174	0,3128
$C_{21}$	0,3362	0,334	0,3324	0,481	0,499	0,3156	0,4296	0,392	0,1855
$C_{22}$	0,6292	0,7027	0,7777	0,5932	0,536	0,686	0,6486	0,6485	0,6667
$C_{23}$	<b>1</b>	0,7389	0,7045	0,6316	0,5745	0,7611	0,719	0,7207	0,7026
$C_{24}$	0,7389	<b>1</b>	0,7388	0,7029	0,535	0,7976	0,6447	0,7221	0,7393
$C_{25}$	0,7045	0,7388	<b>1</b>	0,6285	0,5344	0,7958	0,5366	0,5353	0,8531
$C_{26}$	0,6316	0,7029	0,6285	<b>1</b>	0,7959	0,6515	0,6896	0,6143	0,5573
$C_{27}$	0,5745	0,535	0,5344	0,7959	<b>1</b>	0,5937	0,6334	0,5587	0,4257
$C_{28}$	0,7611	0,7976	0,7958	0,6515	0,5937	<b>1</b>	0,6669	0,7394	0,8319
$C_{29}$	0,719	0,6447	0,5366	0,6896	0,6334	0,6669	<b>1</b>	0,7776	0,6086
$C_{30}$	0,7207	0,7221	0,5353	0,6143	0,5587	0,7394	0,7776	<b>1</b>	0,6433
$C_{31}$	0,7026	0,7393	0,8531	0,5573	0,4257	0,8319	0,6086	0,6433	<b>1</b>

**Anexo 6: Casos semejantes.**

Los casos semejantes a cada uno de los casos del conjunto de prueba, que superaron el umbral, se pueden apreciar en las siguientes figuras.

Caso	Semejanza	Nivel
C1	0,98051	A
C9	0,96179	M
C10	0,94479	M
C4	0,92667	M
C2	0,92415	A
C7	0,90629	M
C11	0,90587	M
C19	0,8688	B
C6	0,86763	M
C21	0,83397	B
C8	0,81166	M
C5	0,81131	M
C17	0,7772	B
C12	0,72384	B
C20	0,70669	B
C15	0,70276	B
C16	0,68721	B
C3	0,68614	M
C18	0,62734	B
C13	0,61364	B
C14	0,59204	B

**Figura 16:** Casos semejantes al Caso 1

Caso	Semejanza	Nivel
C2	0,77931	A
C4	0,77879	M
C17	0,77677	B
C15	0,77605	B
C11	0,76244	M
C3	0,761	M
C10	0,7609	M
C1	0,76085	A
C9	0,74138	M
C19	0,72537	B
C12	0,72388	B
C16	0,72315	B
C18	0,70368	B
C8	0,70352	M
C21	0,68609	B
C6	0,68545	M
C7	0,68516	M
C27	0,6313	S
C5	0,62809	M
C20	0,59131	B

**Figura 17:** Casos semejantes al Caso 2

Caso	Semejanza	Nivel
C15	0,76217	B
C16	0,74211	B
C14	0,72137	B
C11	0,70557	M
C19	0,70444	B
C13	0,7035	B
C3	0,70335	M
C8	0,6878	M
C18	0,68593	B
C6	0,66885	M
C10	0,66665	M
C27	0,65277	S
C17	0,65026	B
C1	0,63093	A
C7	0,63019	M
C21	0,62723	B
C30	0,61372	S
C9	0,61145	M
C4	0,60994	M
C12	0,5951	B
C23	0,59462	S
C22	0,59185	B

Figura 18: Casos semejantes al Caso 3

Caso	Semejanza	Nivel
C24	0,79759	S
C23	0,75768	S
C31	0,68636	S
C25	0,68574	S
C30	0,66779	S
C22	0,64877	B
C26	0,64719	S
C28	0,63209	S
C29	0,62794	S

Figura 19: Casos semejantes al Caso 4

Caso	Semejanza	Nivel
C28	0,84929	S
C23	0,79938	S
C31	0,79279	S
C25	0,75895	S
C30	0,73811	S
C22	0,72188	B
C29	0,7035	S
C24	0,68281	S

Figura 20: Casos semejantes al Caso 5