

# Capítulo 1

## Introducción

*“La razón por la cual el lenguaje natural se expresa en términos difusos no es porque el pensamiento humano sea difuso, sino por que el mundo es difuso”.*

John F. Sowa, matemático norteamericano (1940-).

La mayoría de los fenómenos que se encuentran actualmente son imprecisos o inciertos, es decir, tienen implícito un *cierto grado de borrosidad o imprecisión* (difusidad) en la descripción de su naturaleza o de sus características. Esta imprecisión puede estar asociada a cualquier tipo de datos como su forma, posición, momento, color, textura, o incluso en la semántica que describe lo que son. En muchos casos, el mismo concepto puede tener diferentes significados imprecisos en diferentes contextos o tiempos. Un día cálido en invierno no es exactamente lo mismo que un día cálido en primavera, la definición exacta de cuando la temperatura va de templada a cálida es imprecisa depende del contexto. No se puede identificar un valor específico y único que identifique a cálido o templado, así que se pasa a asociar cálido o templado con un simple grado, por ejemplo, si se mide una temperatura de “25 grados” este grado se asocia a una temperatura cálida, pero, puede ser que “24 grados”, sea asociada a cálida también. Este tipo de imprecisión o difusidad esta continuamente enlazada a los fenómenos, y es común en todos los campos de estudio: Sociología, física, biología, finanzas, ingeniería, oceanografía, etc.

Una solución al planteamiento anterior es la *teoría de los conjuntos difusos*, que fue introducida por L. A. Zadeh en 1965, cuya propuesta considera que, *cada elemento tienen un grado de pertenencia a un conjunto*, este grado suele ser un valor entre 1 (pertenece totalmente) y 0 (no pertenece).

Con el fin de gestionar esta imprecisión en bases de datos, se han propuesto en los últimos años varios trabajos sobre “Bases de Datos Difusas”, que pretenden aplicar la lógica difusa a la tecnología de las bases de datos (Medina et al., 1994; Petry, 1996; Galindo et al., 1998). Sin embargo, estos investigadores se han centrado preferentemente en el diseño lógico

(relacional) y en la construcción de bases de datos difusas, descuidando una de las etapas más importantes, la fase de diseño conceptual.

Ya que como sabemos, se considera que una metodología del diseño de bases de datos se desarrolla en tres fases: diseño conceptual, diseño lógico y diseño físico. El *diseño conceptual* es la etapa en la que se analizan los requisitos que debe tener una base de datos desde un punto de vista independiente del modelo de datos que se utilice, para luego, *diseñar lógicamente* la base de datos y el respectivo *diseño físico* que se implementará.

## 1.1 Planteamiento y Justificación de la Tesis

El objetivo principal de esta tesis se centra en el diseño conceptual (específicamente en el modelo conceptual) que es la primera fase de una metodología de diseño de bases de datos. Para esta etapa se suele utilizar el Modelo ER (Entity-Relationship), y su extensión EER (Enhanced Entity-Relationship) al que se le denomina ER/EER. Como es sabido, existe una variedad de nomenclaturas del modelado en cuestión propuestas por diversos autores como De Miguel et al. (1999), Elmasri y Navathe (2000) entre otros, no existiendo una propuesta de modelado generalizada. Esta tesis ha tenido como finalidad extender casi la totalidad de las componentes del ER/EER para representar semántica y restricciones utilizando la teoría de conjuntos difusos, preferentemente hemos usado la notación de los autores anteriormente mencionados.

Originalmente, el diseño conceptual permite la utilización de tipos de datos elementales, a los que se les llaman clásicos, tradicionales o “*crisp*”, dicho de otra forma, datos precisos que pueden ser manipulados con la lógica clásica. Estos tipos de datos son comúnmente numéricos, alfanuméricos y datos (booleanos). Existen otros tipos de datos, que no son tratados en los modelos de datos conceptuales clásicos o tradicionales, que tienen un dominio *impreciso o ambiguo*, llamados “*difusos*”, “*fuzzy*”, o “*borrosos*”, asociados a la *Teoría de Conjuntos Difusos*. Considérese que las personas manejan de forma cotidiana y natural datos imprecisos y de ahí la importancia de extender las bases de datos tradicionales a difusas.

Por concepto o información difusa (*fuzzy*) se entiende la información que encierra alguna *imprecisión* o *incertidumbre* (Zadeh, 1965). En ocasiones el término “*imprecisión*” engloba varios significados que es interesante distinguir, por ejemplo: puede ser que la información que se tiene sea incompleta, o bien, que no se sabe si es cierta o no —*incertidumbre*—, o tal vez, se

desconoce totalmente la información —*UNKNOWN*— o que dicha información no es aplicable a una determinada entidad —*Undefined*— o que se ignora totalmente si es aplicable —*Null*—. Estos conceptos (*UNKNOWN*, *UNDEFINED* y *Null*) se usan en el sentido de Umano y Fukami (1994) (véase apartado III.5). Cada uno de estos términos dependerá del contexto en el que se aplique. A veces estos significados no son disyuntivos, sino que pueden unirse en una determinada información.

Se puede decir que cada día con mayor frecuencia nos enfrentamos al hecho de tener que manejar información imprecisa sobre el mundo real (utilizando medios informáticos), sobre todo a la hora de desarrollar sistemas de información. Es por ello, que se ha motivado esta investigación que contiene una propuesta de aplicación de la lógica difusa al modelado conceptual, aspecto que hasta el momento ha recibido poca atención tanto, por parte de los diseñadores de sistemas de información, como por parte de los investigadores en bases de datos difusas. El aporte se centra en modelar entidades, atributos, interrelaciones, definir el significado de grados de difusidad que tengan los atributos imprecisos o precisos, las subclases de una especialización y agregación, así como también, aplicar cuantificadores para conseguir “flexibilizar” las principales restricciones de las interrelaciones en un esquema conceptual, en este caso: participación, tipo de correspondencia, cardinalidad, razón de cardinalidad, además de las principales restricciones en especializaciones en un esquema conceptual, en este caso: completitud y disyunción, categorías (o tipos de unión) y subclase compartidas (o tipo de intersección), como un caso especial se define un tipo de atributo difuso en la especialización. En general, lo que proponemos es un modelo de datos difuso (notación y significado) al que llamamos FuzzyEER, dicho modelo extiende algunas de las herramientas de modelado de datos ER/EER expuestas, tanto en De Miguel et al. (1999), como en Elmasri y Navathe (2002).

Para validar nuestra propuesta, discutimos ampliamente dos casos de sistemas de información que en su especificación de requerimientos necesitan soportar tanto datos precisos como imprecisos. Los casos elegidos son: un sistema de información de una *agencia inmobiliaria*, y un sistema de información del *control de calidad del tipo de papel* de una empresa de Cartulina CMPC (Compañía Manufacturera de Papeles y Cartulina). Un tercer sistema de información, que su formulación todavía es incipiente, se presenta en el Apéndice V de *museos digitales en internet*.

También construimos una herramienta gráfica (FuzzyCASE) que permite modelar la mayoría de las notaciones del modelo difuso FuzzyEER, y con ello facilitar el entendimiento del

modelo propuesto, permitiendo así, que los usuarios logren entender las ventajas de FuzzyEER, y verificar que satisface las necesidades requeridas en cuanto a imprecisión de datos se refiera de un sistema de información, con el uso de la teoría de conjuntos difusos.

## **Descripción por Capítulo**

De una manera rápida y general, se describe el contenido de los apartados de esta tesis explicando brevemente cada uno de ellos:

**Capítulo uno:** Introducción, contiene una breve descripción de la temática en cuestión, planteamiento y justificación del trabajo, hipótesis y objetivos de trabajo, además del marco teórico.

**Capítulo dos:** Método de Trabajo, se describe el método de trabajo empleado en esta investigación.

**Capítulo tres:** Estado del Arte, contiene una breve introducción a los modelos conceptuales propuestos en: Batini et al. (1994), De Miguel et al. (1999) y Elmasri y Navathe (2002). Por otra parte, se introducen los principales términos y conceptos que utilizan las representaciones difusas, tales como: conjuntos difusos, modelos difusos, representación de datos imprecisos sobre un referencial ordenado y no ordenado, atributos difusos definidos en cuatro tipos diferentes con respecto a su representación. Por último, cuantificadores difusos absolutos y relativos y algunas definiciones básicas requeridas en esta investigación. Para finalizar, se presenta y discute algunas investigaciones referente a modelado de datos imprecisos, conjuntos difusos, etc., dichas investigaciones están clasificadas en tres grupos: (1) implementación de bases de datos difusas; (2) modelos de datos conceptuales difusos; (3) dependencias difusas.

**Capítulo cuatro:** Propuesta de Modelo Conceptual: FuzzyEER, muestra conceptos, notación y ejemplos de la extensión del modelo de datos difuso FuzzyEER que considera entidades, interrelaciones, atributos difusos, distintos grados difusos. A su vez, se analizan y representan las restricciones de interrelaciones difusas utilizando cuantificadores relativos y absolutos, además, se ilustran algunos ejemplos de completitud y disyunción entre otras para jerarquías. Finalmente, dos cuadros comparativos de las notaciones y tipos de atributos difusos propuestos en esta tesis con otros autores.

**Capítulo cinco:** Validación del Modelo FuzzyEER, se verifica y valida mediante la especificación de requerimiento de dos casos: agencia inmobiliaria y control de calidad del papel, utilizando la notación del modelo difuso FuzzyEER propuesto.

**Capítulo seis:** FuzzyCASE: Una Herramienta Gráfica de Modelo de Datos Difusos, de manera sucinta se muestra la implementación de una herramienta gráfica FuzzyCASE que incorpora la notación FuzzyEER.

**Capítulo siete:** Conclusiones, presenta los resultados, principales aportaciones, contrastación de resultados y líneas abiertas de investigación.

**Apéndices:** Se han considerado seis apéndices: I Manual de uso de la herramienta gráfica FuzzyCASE, II FuzzyUML, III Modelos de bases de datos difusas, IV Representación del conocimiento impreciso en bases de datos relacionales, V Museos digitales en internet: modelo EER difuso y recuperación de imágenes basada en su contenido, y VI Acrónimos.

Finalmente, las referencias bibliográficas utilizadas en este trabajo.

## 1.2 Objetivos e Hipótesis

El objetivo principal de esta investigación puede ser establecido como:

*DEFINIR CONCEPTOS Y NOTACIÓN PARA ESPECIFICAR UN MODELO CONCEPTUAL PARA BASES DE DATOS DIFUSAS.*

Y, en función del objetivo principal se tienen los siguientes objetivos específicos:

1. Estudiar las tecnologías de gestión de bases de datos difusas.
2. Analizar los modelos y notaciones existentes para modelos conceptuales de bases de datos.

3. Proponer una extensión para el modelado de imprecisión o incertidumbre a nivel conceptual.
4. Validar mediante la aplicación de casos de estudio la extensión del modelo propuesto.
5. Construir un prototipo como herramienta CASE que soporte el modelo conceptual difuso propuesto.

Por tanto, en vista de los objetivos presentados, en definitiva, se trata de poder demostrar a lo largo de nuestro trabajo que es cierta la siguiente hipótesis:

ES POSIBLE LA DEFINICIÓN DE UN MODELO CONCEPTUAL QUE SOPORTE TODAS LAS CARACTERÍSTICAS MÁS IMPORTANTES DE LAS BASES DE DATOS DIFUSAS.

Resumiendo, en la actualidad existen algunos sistemas gestores de bases de datos que tratan la imprecisión o incertidumbre (Galindo, 1999), pero no se ha estudiado el modelo conceptual de este tipo de bases de datos. También existen muchas propuestas de metodología de diseño de bases de datos, pero casi ninguna (Chen, 1998) tiene en cuenta los conceptos difusos.

Lo que se propone en esta tesis es crear un modelo conceptual que permita modelar datos con imprecisión o incertidumbre, además de aquellos que sean clásicos (precisos), así como también algunas restricciones inherentes al modelo conceptual, y que esta notación a su vez, pueda ser utilizada en los requerimientos de una base de datos difusa para un sistema de información determinado. Cada notación propuesta debe contemplar su definición y su componente gráfica, con ejemplos para facilitar su entendimiento.

### 1.3 Marco de la Tesis

Esta tesis se enmarca en el proyecto en el área de bases de datos de la *Red IberoAmericana de Tecnologías del Software para la década del 2000 RITOS 2*, que está incluida en la red VII.J en el subprograma VII (Electrónica e Informática Aplicadas) de la organización Iberoamericana CYTED (Ciencia y Tecnología para el Desarrollo), cuyo propósito es favorecer y promover el intercambio de conocimiento, la discusión de propuestas y la colaboración entre los investigadores y estudiantes en las áreas de: Bases de datos, Ingeniería del Software, Sistemas Distribuidos y Procesamiento del Lenguaje Natural.

Sin embargo, el nivel de investigación en informática de los países (latinoamericanos) que integran CYTED es deficiente y hay pocos grupos de investigación lo suficientemente consolidados para ofrecer apoyo a los numerosos grupos emergentes. En esta red se presenta una proporción de grupos consolidados y emergentes que se cree que permitirá satisfacer el objetivo de facilitar la formación de investigadores y nuevos doctores. Por otro lado, es objetivo de la red la de facilitar la formación de profesionales, ya sea mejorando la formación investigadora de los docentes, como también, mediante la organización periódica de talleres, cursos y conferencias, pues es evidente que la transferencia conocimiento a la industria comienza por la formación de los profesionales que en ella trabajan.

Esta tesis se enmarca, en el ámbito de la formación de nuevos investigadores, así como también, en facilitar la formación de nuevos doctores, por otro lado, la temática de la tesis aquí planteada, “*Definición de un modelo conceptual para bases de datos difusas*” propone una investigación en el área de las bases de datos, por tanto, este proyecto, en general, se enmarca en los requerimientos de la red CYTED VIIIJ de RITOS 2.

Hay que señalar que este trabajo contó con la colaboración del grupo Alarcos de la Universidad de Castilla-La Mancha, que tiene como uno de sus objetivos el investigar en el diseño de bases de datos, y con el apoyo del Dr. José Galindo de la Universidad de Málaga, experto en el campo de las bases de datos difusas. Además, en la Universidad de Castilla-La Mancha como en la Universidad de Málaga se disfrutaron pasantía de estancia de la AEIC (Agencia Española de Cooperación Internacional) que facilitaron el desarrollo de este trabajo reduciendo el problema de la enorme distancia existentes entre la doctoranda (Chile) y director

y codirector de tesis (España). Cabe hacer notar, que tanto el grupo Alarcos como el director y codirector de esta tesis, pertenecen a la red CYTED VIIIJ de RITOS 2.

Otro proyecto en que se enmarca esta tesis, es un I+D+D Títulado: “Soft Data Server: Modelo de Servidor de Bases de Datos Objeto-Relacional basado en Soft Computing”, CICYT N° TIC2002-00480. Entidad financiadora: Proyectos I+D del Ministerio de Educación y Ciencia. Investigador principal: Dr. Juan Miguel Medina Rodríguez, Universidad. de Granada. Fechas: 2002-2004, donde participan como investigador el Dr. José Galindo y como investigadora invitada Angélica Urrutia.