Restricciones de Participación y Tipo de Correspondencia Difusa en un Modelo Conceptual

Angélica Urrutia[1] **José Galindo**[2] **Mario Piattini**[3]

[1] Dpto. de Computación e Informática, Univ. Católica del Maule, Chile, aurrutia@spock.ucm.cl
[2] Dpto. Lenguajes y Ciencias de la Computación, Universidad de Málaga, ppgg@lcc.uma.es
[3] Escuela Superior de Informática, Universidad de Castilla la Mancha, mpiattin@inf_cr.uclm.es

Abstract. In this article our aim is to get flexible the constraints which can be expressed in a conceptual model using fuzzy logic. To do so we will use the fuzzy quantifiers which have been widely studied in the context of fuzzy sets and of fuzzy consultation systems for data bases, all this has not been widely studied. We will also examine the representation of these constraints in an EER model and their practical repercussions. The constraints proposed are as follows: restriction of fuzzy participation, constraints in types of correspondences restriction of fuzzy cardinality. We also demonstrate how fuzzy notation (min, max) can substitute fuzzy cardinality constraints. All these fuzzy constraints have a novel meaning and offer greater expressiveness in conceptual design.

Keywords: Conceptual Database Design, Fuzzy Conceptual Modelling, Fuzzy Constraints, Extended (or Enhanced), Entity Relationship Model, Fuzzy Databases.

Resumen. En este artículo presentamos una propuesta de aplicación de la lógica difusa al modelado conceptual. Para ello, utilizamos los cuantificadores difusos ampliamente estudiados en el contexto de los conjuntos difusos y de los sistemas de consulta difusa a bases de datos, temática muy poco estudiada en el modelado de datos conceptual. Conseguimos "flexibilizar" las principales restricciones en un modelo EER, participación y restricción de tipo de correspondencia, ambas es un contexto difuso, además, se muestra la notación (min,max) difusa. Todas estas restricciones difusas tienen un novedoso significado y aportan una mayor expresividad al modelo conceptual tratadas como lógica difusa.

Palabras claves: Diseño Conceptual de bases de datos, Modelo Conceptual Difuso, Restricciones difusas, Modelo de bases de datos difusas, Modelo EER difuso.

1 Introducción

En la actualidad, gran parte de la información que manejamos acerca del mundo real es incompleta, imprecisa, incierta o vaga. Para representar y manejar este tipo de datos se puede utilizar la teoría de la lógica difusa, que fue introducida por Zadeh (1965). Este autor observó que la lógica clásica no representaba datos tales como "aproximadamente dos", "casi todos" o "mediana edad", términos típicos del razonamiento humano. La lógica tradicional al ser bivaluada sólo podía trabajar con los conceptos: sí o no, blanco o negro, todo o nada, por lo que permitía una representación muy limitada del conocimiento.

Con el fin de gestionar esta incertidumbre 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; Galindo, 1999).

Sin embargo, estos investigadores se han centrado preferentemente en el diseño lógico (relacional) y en la construcción de bases de datos difusas, descuidando la fase de modelado conceptual. Por otro lado, las principales metodologías de diseño de bases de datos (Connolly et al., 1998; Elmasri y Navathe, 2000; De Miguel et al., 2000) tampoco han prestado atención a este tema; a pesar de que en el intento de modelar el mundo real, la imprecisión está rara vez ausente.

Esta escasa atención al modelado conceptual difuso supone que no existen muchos trabajos y que los existentes suelen ser bastante parciales, centrándose en un tema y descuidando otros. Así, ninguna publicación define el uso de restricciones difusas usando las ventajas de la teoría de conjuntos difusos. Algunas publicaciones interesantes sobre el modelado difuso son: Zvieli y Chen (1986), que proponen un modelo que soporta atributos difusos en las entidades y las interrelaciones; Chaudhry et al., (1994) proponen una metodología para diseñar bases de datos difusas, siguiendo el modelo anterior y prestando especial interés en convertir bases de datos clásicas (*crisp*) en difusas; La propuesta de Vert et al., (2000) se basa en la notación usada por Oracle y utiliza la teoría de conjuntos difusos para tratar colecciones de objetos difusos, aplicando el resultado a sistemas de información geográfica. Además, utilizando la orientación a objeto en Marín et al., 2000 presenta una forma de modelar imprecisión. En algunos trabajos Chen y Kerre definen una extensión difusa para los conceptos más importantes de EER (superclase, subclase, generalización, especialización, categoría subclase compartida) sin incluir representaciones gráficas (Kerre et al., 2000). Un resumen sobre algunos modelos propuestos puede encontrarse en Kerre et al. (1995). Finalmente Ma et al. (2001) trabajan con el trabajo de Zvieli y Chen como base e introducen el modelo que llaman *Fuzzy Extended Entity-Relationship* (FEER) para tratar con la imperfección y con objetos complejos del mundo real a nivel conceptual, sin tratar el tema de las restricciones.

En este artículo pretendemos flexibilizar el modelado de las restricciones en el modelo entidad/interrelación (ER), proponiendo una representación gráfica y definiciones, además de analizar sus repercusiones prácticas. En la siguiente sección, resumimos los conceptos básicos de la lógica difusa, prestando especial atención a los cuantificadores difusos que utilizaremos en el modelado de las restricciones. En la sección 3, presentamos las restricciones relativas a las relaciones (Participación, Tipo de correspondencia y Cardinalidad), para cada una de las restricciones difusas mostramos su formalización matemática. Finalmente, exponemos algunas conclusiones y algunas líneas para posibles trabajos futuros.

2 Conjuntos difusos: Cuantificadores Difusos

Dos conceptos básicos que son aplicados a las restricciones difusas de interrelaciones propuestas aquí, son: conjuntos difusos y cuantificadores difusos, las que son definidas a continuación.

2.1 Conjuntos difusos

L.A. Zadeh (1965) definió el concepto de conjunto difuso, basándose en la idea de que existen conjuntos en los que no está claramente determinado si un elemento pertenece o no al conjunto. A veces, un elemento pertenece al conjunto con cierto grado, llamado grado de pertenencia. Por ejemplo, el conjunto de las personas que son altas es un conjunto difuso, pues no está claro el límite de altura que establece a partir de que medida una persona es alta o no lo es. Ese límite es difuso y, por tanto, el conjunto que delimita también lo será.

Un conjunto difuso A se define como una función de pertenencia μ_A que hace corresponder los elementos de un dominio o Universo de discurso (U) con elementos del intervalo [0,1]: μ_A: $U \rightarrow [0,1]$

Cuanto más cerca esté μ_A (u) del valor 1, mayor será la pertenencia del objeto u $\in U$ al conjunto A. Los valores de pertenencia varían entre 0 (no pertenece en absoluto) y 1 (pertenencia total). Un conjunto difuso A puede representarse como un conjunto de pares de valores: Cada elemento u con su grado de pertenencia μ_A (u):

$$A = \{\mu_A (u) /u, u \in U\} \tag{1}$$

Un número difuso es un conjunto difuso, donde U es un dominio numérico (normalmente los números reales R). En la Figura 1 se representa la función de pertenencia del número difuso "Aproximadamente n". El valor margen indica los límites del conjunto difuso. Es fácil observar que cuanto más cerca esté un número del valor n, su grado de pertenencia a "Aproximadamente n" será mayor.

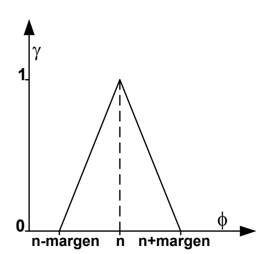

Figura 1: Función "Aproximadamente n" (n +/– margen).

A partir de este sencillo concepto se ha desarrollado toda una teoría matemática e informática que facilita la resolución de problemas (Petry, 1996; Pedrycz y Gomide, 1998) tales como: control de sistemas, simulación, reconocimiento de patrones, sistemas de información o conocimiento, visión por ordenador, vida artificial...

2.2 Cuantificadores Difusos

Los cuantificadores difusos o lingüísticos (Yager, 1983; Zadeh, 1983; Galindo, 1999; Galindo et al., 2001) permiten expresar cantidades o proporciones difusas para dar una idea aproximada del número de elementos de un subconjunto (o que cumplen cierta condición), o de la proporción de ese número en relación con el total de elementos posibles. Los cuantificadores pueden ser absolutos o relativos:

- Los **cuantificadores absolutos** expresan cantidades sobre el número total de elementos de un determinado conjunto, diciendo si este número es "grande", "pequeño", "muchos", "pocos", "muchísimos", "aproximadamente entre 5 y 10"... En estos casos se observa que la verdad del cuantificador depende de una única cantidad. Por eso, la definición de los cuantificadores difusos absolutos es, como veremos, muy similar a los números difusos.
- Los **cuantificadores relativos** expresan mediciones sobre el número total de elementos que cumplen cierta característica dependiendo del total de elementos posibles, por lo que la verdad del cuantificador depende de dos cantidades. Este tipo de cuantificadores se usa en expresiones como "la mayoría", "la minoría", "aproximadamente la mitad"... En estos casos, para evaluar la verdad del cuantificador necesitamos hallar la cantidad total de elementos que cumplen la condición y ponderarla respecto a la cantidad total de elementos que podría cumplirla (incluyendo los que la cumplen y los que no la cumplen).

Los cuantificadores difusos absolutos son definidos como conjuntos difusos (Zadeh, 1983) en el intervalo $[0,+\infty)$ y los relativos como conjuntos difusos en el intervalo $[0,1]$. O sea, un cuantificador Q es representado como una función Q, cuyo dominio depende de si es absoluto o relativo:

$$Q_{\mathrm{abs}} : \mathrm{R}^{+} \quad \rightarrow [0,1] \tag{2}$$

$$Q_{\mathrm{rel}} : [0,1] \quad \rightarrow [0,1] \tag{3}$$

Donde el dominio de Q_{rel} es $[0,1]$ porque en ese intervalo toma valor la división del número de elementos que cumplen cierta condición entre el número total de elementos existentes.

Para saber en que grado se cumple el cuantificador sobre los elementos que cumplen cierta condición, se aplica la función Q del cuantificador a un valor de cuantificación Φ:

- Si Q es **absoluto**, el valor Φ es el número de elementos que cumplen cierta condición.

- Si Q es **relativo**, Φ es la división del número de elementos que cumplen cierta condición entre el número total de elementos existentes (o posibles).

Si la función del cuantificador (absoluto o relativo), $Q(\Phi)$, toma el valor 1 indica que dicho cuantificador se satisface completamente. El valor 0 indica, por el contrario, que el cuantificador no se cumple en absoluto. Cualquier valor intermedio indica un grado de cumplimiento intermedio del cuantificador.

Ejemplo 1: Un cuantificador difuso absoluto es "aproximadamente 8", definido como muestra la Figura 1, con n=8 y margen=3, por ejemplo. Un cuantificador difuso relativo es "casi todos", definido como muestra la Figura 2. Este último cuantificador revela que si la condición es satisfecha por el 90% o más de los elementos posibles, el cuantificador "casi todos" se cumplirá con grado 1. Ese grado de cumplimiento será 0 si se cumple por el 40% o menos de los elementos posibles. Como muestra la Figura 2, cualquier valor intermedio generará valores intermedios.

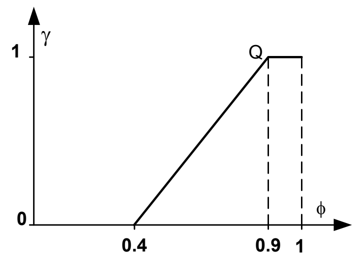

Figura 2: Cuantificador difuso relativo "casi todos": $x \in [0.4, 0.9] \leftrightarrow y = 2(x - 0.4)$.

Por supuesto, que el cuantificador Q debe estar definido por su función Q en el diccionario de datos que acompaña al modelo conceptual de datos.

En general, si Q es un cuantificador difuso relativo entonces, la Ecuación (2) expresa que la restricción debe satisfacerse en la base de datos en un *porcentaje* mínimo de $100Q^{-1}(\gamma)$. En ese caso, también será posible expresar ese *porcentaje* en vez del cuantificador Q y del umbral γ. Aunque esto pueda parecer más simple, hay que hacer notar que se pierde la expresividad intuitiva y natural del cuantificador y que esto no es válido con cuantificadores difusos absolutos ni para cuantificadores relativos que no sean crecientes.

Además, puede establecerse otro valor opcional δ, superior al grado de cumplimiento γ, de la siguiente forma: $Q[\gamma,\delta]$, tal que $\gamma<\delta$. El valor δ es más restrictivo y establece que cuando se incumpla la restricción con ese valor superior, o sea entre el intervalo $[\gamma,\delta]$ se deberá advertir al usuario, pero se permitirá la tarea correspondiente (véase Figura 3). El valor δ permite que el sistema avise de que alguna restricción difusa está cerca de su punto de incumplimiento. Esto es especialmente útil porque permite que el usuario de la base de datos (o DBA) tome las medidas necesarias para que la restricción no llegue a incumplirse.

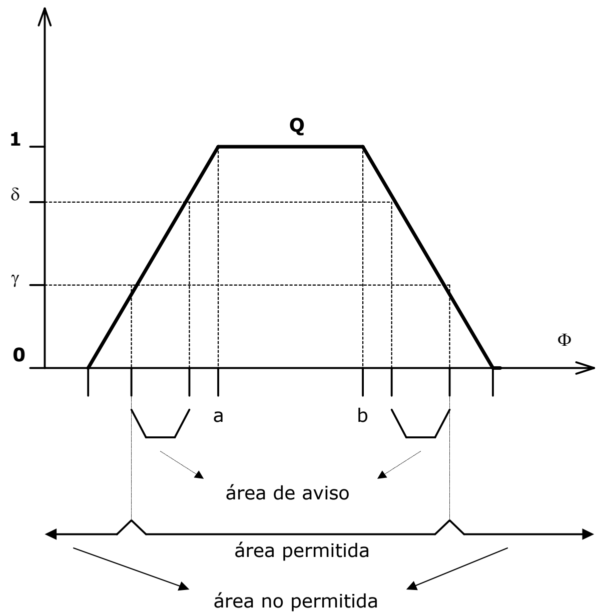

Figura 3: Cuantificador difuso con áreas de restricción $Q[\gamma,\delta]$.

Aplicados en el contexto de las bases de datos, los cuantificadores difusos permiten expresiones sobre el número de instancias que satisfacen una determinada condición o la proporción respecto del total. Su utilidad está demostrada por la flexibilidad que ofrecen para efectuar consultas que involucren estos cuantificadores, como por ejemplo, en la utilización de la operación de división del álgebra relacional en bases de datos difusas o clásicas (Galindo et al., 2001).

3 Restricciones Difusas del Modelo ER

En este apartado se mostrarán la extensión de las Restricciones difusas de Participación, Tipo de correspondencia y Cardinalidad, cada una de ellas con su respectivo concepto aplicado y su representación para un modelo conceptual difuso.

3.1 Restricción de Participación Difusa

La Participación de una entidad en una interrelación puede ser total (si cada entidad debe de relacionarse forzosamente con la otra u otras entidades de la interrelación) o parcial (si dicha interrelación no es obligatoria para todas las entidades pertenecientes a ese tipo). La Participación total se representa por una doble línea uniendo la entidad correspondiente con el rombo de la interrrelación, mientras que la Participación parcial se representa con una línea simple (Connolly et al., 1998; Elmasri y Navathe, 2000).

En el modelo difuso que proponemos la Participación puede ser difusa utilizando (principalmente) un cuantificador difuso relativo. La representación se haría utilizando una *línea zigzag* indicando sobre esta línea el cuantificador difuso utilizado seguido, opcionalmente, de un umbral o grado de cumplimiento γ entre corchetes para dicho cuantificador, cuyo valor por defecto es 0.5. El umbral γ establece el requisito de que el cuantificador difuso asociado a esa Participación debe cumplirse forzosamente con grado mínimo γ. Por tanto, γ está asociado a cada cuantificador de cada restricción.

Cada vez que la base de datos se modifica, el DBMS evalúa el valor de Φ y se comprueba si se cumple la siguiente condición:

$$Q(\Phi) \geq \gamma \tag{4}$$

Si Q es una función creciente podemos cambiar la Ecuación (4), donde obtenemos que:

$$\Phi \geq Q^{-1}(\gamma) \tag{5}$$

Igualmente si Q es una función decreciente, obtenemos que:

$$\Phi \leq Q^{-1}(\gamma) \tag{6}$$

Definición 1: Sean E_1 y E_2 dos entidades interrelacionadas con R. Se define una **restricción de participación difusa** de E_1 en R, y se representa con un cuantificador Q sobre la línea que une E_1 con R, siendo esta línea sustituida por una línea *zigzag* (véase Figura 4a). También se admite la representación con una línea normal cruzada con un arco etiquetado con Q (véase Figura 4b). Está restricción establece que:

$$Q(\Phi) \geq \gamma \tag{7}$$

Donde γ es el umbral mínimo establecido para Q, y Φ es:

$$\Phi = \begin{cases} a & \text{si } Q \text{ es absoluto} \\ a/b & \text{si } Q \text{ es relativo} \end{cases} \tag{8}$$

Siendo a el número de instancias de E_1 relacionadas con E_2, y b siendo el número total de instancias de E_1.

*

Si Q tiene dos umbrales $[\gamma,\delta]$ entonces define un área de aviso, como se explicó en el apartado 2.2. Por tanto, siempre que existe un área de aviso habrá que generar un mensaje cuando la condición establecida por la restricción se cumpla y, a la vez, no se cumpla la misma restricción utilizando el umbral δ en vez del umbral γ.

O sea, la condición establecida por la restricción define el área permitida y como el área de aviso está dentro del área permitida, en el área de aviso debe forzosamente cumplirse esa condición. Como δ es más restrictivo (γ,δ) pudiera ser que la condición δ no se cumpla, y en ese caso es cuando se genera el aviso.

Para la definición 1, por ejemplo, el área de aviso está definida cuando se cumple la ecuación (7) y no se cumple que:

$$Q(\Phi) \geq \delta \tag{9}$$

o lo que es lo mismo, que se cumpla que:

$$Q(\Phi) \geq \gamma \ \wedge \ Q(\Phi) < \delta \tag{10}$$

Resumiendo, si la ecuación (10) se cumple se genera el aviso, si se cumple la ecuación (9) la operación en curso está totalmente permitida, y en cualquier otro caso se genera un error, que evite que se incumpla la restricción difusa impuesta.

Igualmente, como se trato en el apartado 2.2, si Q no tiene umbral γ se supone γ=0.5 y si Q tiene dos umbrales [γ,δ], el umbral δ establece el área de aviso ($Q(\Phi) < \delta$), dentro del área de permitida.

Ejemplo 2: Supongamos una entidad Empleado y otra Proyecto unidas a través de una interrelación Trabaja_en. La Participación de Empleado en esa interrelación puede ser representada a través de un cuantificador difuso relativo como "casi todos" (Figura 2), indicando que "Casi todos los empleados trabajan para algún proyecto". La Figura 4, representa estas dos entidades, la interrelación entre ellas y esa restricción difusa, utilizando las dos notaciones propuestas en la definición 1.

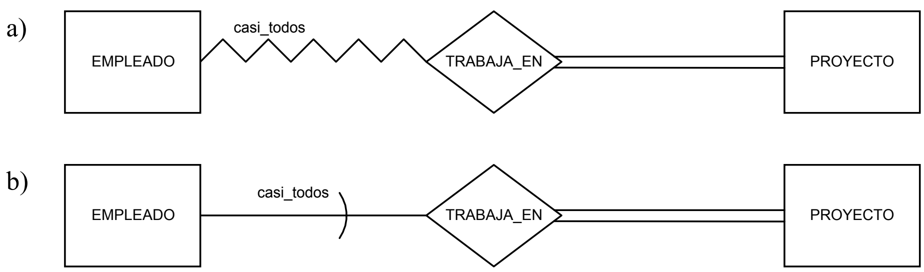

Figura 4: Representación de una restricción de Participación con el cuantificador difuso "casi_todos".

En la Figura 5 se muestra el mismo cuantificador con un grado de cumplimiento explícito γ = 0.2, "casi todos [0.2]", indicando el grado mínimo con el que este cuantificador debe cumplirse en la base de datos. O sea, si dividimos el número de empleados que trabajan para algún proyecto, entre el número de empleados total de la base de datos, el resultado Φ debe ser, de acuerdo con la Ecuación (2), un valor mayor o igual a $Q^{-1}(0.2) = 0.5$, ya que ese es el valor para el que el cuantificador "casi todos" alcanza el valor de cumplimiento de 0.2: $Q(0.5) = 0.2$. A partir de ese valor 0.5 este cuantificador obtiene un grado mayor a 0.2, que era la restricción impuesta por el umbral γ en el cuantificador inicial. En este ejemplo, el valor 0.5 obtenido por la expresión $Q^{-1}(0.2)$ indica la restricción de que en nuestra base de datos, un mínimo del 50% de los empleados deben trabajar para algún proyecto.

Figura 5: Restricción de Participación difusa usando "casi_todos" con grado de cumplimiento 0.2.

Resumiendo, un cuantificador puede aparecer de tres formas en un modelo de datos:

1. Sin umbral (utilizando su valor por defecto γ=0.5). Por ejemplo, aprox_2.
2. Con un umbral explícito γ representado como [γ]. Por ejemplo, aprox_8[0.25].
3. Con dos umbrales γ y δ, con $\gamma< \delta$, representado como [γ,δ]. Por ejemplo, aprox_3[0.25,0.75].

Observe que el cuantificador difuso expresado en este tipo de restricciones también puede ser absoluto, aunque debido al significado de una restricción de Participación este será, en general, relativo ya que el número de instancias de la entidad afectada puede variar mucho. En caso de que sea un cuantificador difuso absoluto, como por ejemplo "muchos" o "aproximadamente entre 100 y 200", este cuantificador restringirá la cantidad de instancias que se interrelacionan con la otra entidad. En nuestro ejemplo se restringirá la cantidad de empleados que tienen algún proyecto asignado para el que trabajar. Incluso, en algún modelo puede ser interesante establecer varios cuantificadores difusos como restricción de Participación difusa.

Una restricción de Participación difusa no es tan restrictiva como la restricción de Participación total, ni tan permisiva como la restricción de Participación parcial, por lo que las restricciones de Participación difusas extienden el modelo ER permitiendo una nueva expresividad que no se encuentra en el modelo tradicional

3.2 Restricción sobre el Tipo de Correspondencia Difusa

En modelado clásico, la restricción de Tipo de correspondencia respecto a la Cardinalidad expresa si la relación entre entidades es de "*uno a uno"* (1:1), de "*uno a muchos"* (1:N) o de "*muchos a muchos"* (N:M) (De Miguel et al., 1999), estas restricciones son llamadas razón de cardinalidad en (Elmasri y Navathe, 2000). El modelo que proponemos permite expresar el Tipo de correspondencia como un valor difuso utilizando (principalmente) un cuantificador difuso absoluto. La representación se haría escribiendo ambos cuantificadores justo bajo el rombo que representa la interrelación entre ambas entidades y separados por el carácter "**:**". El cuantificador situado a la izquierda del separador "**:**" corresponderá a la entidad situada a la izquierda y el cuantificador de la derecha para la otra entidad. Opcionalmente, puede ponerse uno o dos grados de cumplimiento entre corchetes [γ, δ] para cada cuantificador con el mismo significado y valor por defecto que el explicado anteriormente para las restricciones de Participación difusas.

Definición 2: Sean E_1 y E_2 dos entidades interrelacionadas con R como lo muestra la Figura 8, suponemos que e_i con i=1,2,...,b_1, son las instancias de E_1 y w_j con j=1,2,...,b_2, son las instancias de E_2. Una **restricción de tipo de correspondencia** consiste en dos cuantificadores difusos separados por el carácter ":", de la forma "Q_1:Q_2", que afecta respectivamente a E_1 y E_2, haciendo cumplir las siguientes dos restricciones. Se cumple que:

$$Q_1(\Phi_{1i}) \geq \gamma_1 \qquad \forall\ i = 1,2,...,b_1 \tag{11}$$

donde γ_1 es el umbral mínimo establecido por Q_1. Por otra parte, b_1 es el número total de instancias de E_1, y Φ_{1i} con i = 1,2,..., b_1 está definido como:

$$\Phi_{1i} = \begin{cases} a_i & \text{si } Q_1 \text{ es absoluto} \\ a_i/b_2 & \text{si } Q_1 \text{ es relativo} \end{cases} \tag{12}$$

siendo a_i el número de instancias de E_2 con las que se relaciona la instancia $e_i \in E_1$, y b_2 el número total de instancias de E_2. Se cumple que:

$$Q_2\ (\Phi_{2j}) \geq \gamma_2 \qquad \forall\ j = 1,2,..., b_2 \tag{13}$$

donde γ_2 es el umbral mínimo establecido para Q_2. Por otra parte, Φ_{2j} con j = 1,2,...,b_2, definido como:

$$\Phi_{2j} = \begin{cases} a_j & \text{si } Q_2 \text{ es absoluto} \\ a_j/b_1 & \text{si } Q_2 \text{ es relativo} \end{cases} \tag{14}$$

con a_j siendo el número de instancias de E_1 con las que se relaciona la instancia w_j de E_2.

*

Generalizando, el cuantificador Q_1 debe cumplirse para todos los valores de cuantificación Φ_{1i} con i=1,2,...,b_1, o, dicho de otra forma, los valores de Φ_{1i} deben estar en un área permitida de Q_1 según su umbral γ_1 establecido. También debe cumplirse Q_2 para todos los valores de cuantificación Φ_{2j} con j = 1,2,...,b_2, o dicho de otra forma, los valores Φ_{2j} deben estar en el área permitida de Q_2 según su umbral γ_2 establecido.

Ejemplo 3: Siguiendo con el ejemplo anterior, si suponemos que la entidad Empleado está a la izquierda de la interrelación Trabaja_en y la entidad Proyecto está a la derecha, una restricción de Tipo de correspondencia difusa podrá ser: Menos_de_aprox_ 3 [0.8] **:** aprox_8 [0.4]. Véase Figura 6.

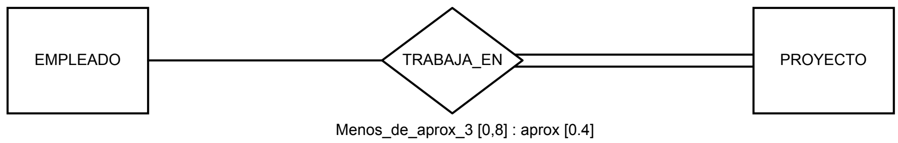

Figura 6: Representación gráfica de una restricción difusa en el Tipo de correspondencia.

Estas restricciones expresan la condición de que cada empleado trabajará en un máximo de aproximadamente 3 proyectos y cada proyecto tendrá aproximadamente 8 empleados, exigiendo que ambas restricciones se cumplan con los grados de verdad mínimos indicados entre corchetes, [0.8] y [0.4] respectivamente.

Observe que el cuantificador difuso de este tipo de restricciones también puede ser relativo, aunque debido al significado de una restricción de Tipo de correspondencia este será, en general, absoluto. En caso de que sea un cuantificador difuso relativo, este cuantificador indicará la cantidad de instancias de la otra entidad con las que se interrelaciona cada entidad con respecto al número total de instancias de la otra entidad que existen. Así, si en el Ejemplo 3 usamos el cuantificador “casi todos” a la izquierda esto significa que “cada empleado deberá trabajar en casi todos los proyectos existentes”. Incluso, en algún modelo puede ser interesante establecer varios cuantificadores difusos en algún o en ambos lados de una restricción de Tipo de correspondencia difusa.

Si una interrelación tiene grado mayor a dos, el cuantificador difuso de Tipo de correspondencia respecto a la Cardinalidad asociado a cada una de las entidades puede ponerse al lado del arco que une cada entidad con la interrelación. Si hubiera ya un cuantificador para la restricción de Participación difusa (línea *zigzag*), entonces, para evitar ambigüedad debemos poner delante del cuantificador de Tipo de correspondencia el texto “Card:”.

3.3 Restricción de Cardinalidad Difusa, con la Notación (min, max) Difusa

Las restricciones de Participación y Tipo de correspondencia *no difusa* de una entidad en una interrelación se puede expresar a través de esta restricción de Cardinalidad con la notación (min, max), donde min y max indican respectivamente, el número mínimo y máximo de instancias de la entidad que participan en la interrelación.

En modelado difuso, tanto min como max, pueden tomar valores que sean cuantificadores difusos de manera similar a la explicada anteriormente.

Definición 3: Sean E_1 y E_2 dos entidades interrelacionadas con R, denotaremos con e_i siendo i=1,2,...,b_1 las instancias de E_1, y con w_j siendo j=1,2,...,b_2 las instancias de E_2. Se define una **restricción usando la notación (min, max) difusa** de E_1 en R, y se representa escribiendo junto a la línea que une E_1 con R, dos cuantificadores (por lo general absolutos) entre paréntesis, (Q_{min}, Q_{max}), como lo muestra la Figura 7. Esta restricción establece que:

$$\lambda_{min} \leq \Phi_{min,i} \quad \wedge \quad \lambda_{max} \geq \Phi_{max,i} \quad \forall\, i = 1,2,...,b_1 \tag{15}$$

donde, b_1 el número total de instancias de E_1

$$\Phi_{min,i} = \begin{cases} a_i & \text{si } Q_{min} \text{ es absoluto} \\ a_i/b_2 & \text{si } Q_{min} \text{ es relativo} \end{cases}$$
$$\Phi_{max,i} = \begin{cases} a_i & \text{si } Q_{max} \text{ es absoluto} \\ a_i/b_2 & \text{si } Q_{max} \text{ es relativo} \end{cases} \tag{16}$$

En este caso a_i es el número de instancias de E_2 con las que se relaciona la instancia e_i de E_1, y b_2 es el número total de instancias de E_2. Además, tenemos que:

$$\lambda_{min} = \min\,\{\alpha : \alpha = (Q_{min})^{-1}(\gamma_{min})\} \tag{17}$$
$$\lambda_{max} = \max\{\beta : \beta = (Q_{max})^{-1}(\gamma_{max})\}$$

donde, γ_1 y γ_2 son los umbrales mínimos establecidos para Q_{min} y Q_{max} respectivamente.

*

Por lo tanto, el intervalo $[\lambda_{min},\lambda_{max}]$ es el área permitida. El área de aviso se define, si están demarcados dos umbrales δ_{min} y δ_{max} para Q_{min} y Q_{max} respectivamente, cuando se cumple la condición de la ecuación (16) y no se cumple que:

$$\lambda'_{min} \leq \Phi_{min,i} \quad \wedge \quad \lambda'_{max} \geq \Phi_{max,i} \quad \forall\, i = 1,2,...,b_2 \tag{18}$$

donde,
$$\lambda'_{min} = \min\,\{\alpha : \alpha = (Q_{min})^{-1}(\delta_{min})\} \tag{19}$$
$$\lambda'_{max} = \max\,\{\beta : \beta = (Q_{max})^{-1}(\delta_{max})\}$$

Así, el área de aviso es la unión de dos intervalos: $[\lambda_{min},\lambda'_{min}] \cup [\lambda'_{max},\lambda_{max}]$. Sabemos que: $\lambda_{min} < \lambda'_{min}$ y que $\lambda'_{max} < \lambda_{max}$ porque $\gamma_{min} < \delta_{min}$ y $\gamma_{max} < \delta_{max}$ respectivamente y los cuantificadores están definidos como funciones convexas.

Obsérvese que las ecuaciones (19) son iguales que las ecuaciones (17) cambiando γ_{min} y γ_{max} por δ_{min} y δ_{max} respectivamente. Además en la ecuación 18 se puede aplicar la ley de De Morga qudando:

$$\lambda'_{min} > \Phi_{min,i} \quad \vee \quad \lambda'_{max} < \Phi_{max,i} \quad \forall\, i = 1,2,...,b_2 \tag{20}$$

*

Consideraciones a tener en cuenta: En la elección de algunos cuantificadores a usar, en cualquiera de las restricciones anteriores, hay que considerar que el cuantificador difuso sea representativo, por lo que se hacen los siguientes alcances:

- ✓ Si $Q_{min}(0) \geq \gamma_{min}$ no usar este cuantificador $Q_{min}[\gamma_{min}]$, usar 0 en el lugar de Q_{min}: $[0, Q_{max}]$.
- ✓ Si $Q_{max}(\psi) \geq \gamma_{max}$, donde ψ es el valor máximo que puede tomar el dominio subyacente del cuantificador Q_{max}, entonces no debe usarse este cuantificador Q_{max}, sino que se debe usar la letra "N" en su lugar, denotando una relación de "muchos": $[Q_{min}, N]$.
- ✓ Además, si Q_{min} y Q_{max} son del mismo tipo (relativos o absolutos), implica que: $\Phi_i^{min} = \Phi_i^{max}$

En síntesis, se trata de no usar cuantificadores que especifiquen mal la restricción, como por ejemplo: [aprox_3_o_menos, casi_todos].

Ejemplo 4: En el contexto de los ejemplos anteriores pueden ponerse las siguientes restricciones utilizando la Cardinalidad difusa, las cuales están representadas en la Figura 7. En el lado de la entidad Empleado puede ponerse la restricción de Cardinalidad "(0, aprox_3 [0.25, 0.75])", indicando que un empleado puede no pertenecer a ningún proyecto (0 como mínimo) y como máximo puede pertenecer a aproximadamente 3 proyectos. Los dos valores tras el cuantificador indican que si éste se cumple con un grado mayor a 0.75 se permitirá la tarea asociada a la restricción normalmente, si se cumple con un grado entre 0.25 y 0.75 entonces se avisará al usuario pero se permitirá la tarea y si la restricción se cumple con un grado inferior a 0.25 querrá decir que se está incumpliendo la restricción con un grado no permitido y, por tanto, la tarea en curso no debe permitirse.

En el lado de la entidad Proyecto del mismo ejemplo, podemos encontrar la restricción (min, max) siguiente, "(aprox_2, aprox_8 [0.25])", indicando que un proyecto tiene que tener como mínimo aproximadamente 2 empleados (con grado 0.5 como mínimo). Si el cuantificador aprox_2 está definido como un triángulo como la Figura 1 con n=2 y margen=2, entonces el valor 0.5 se consigue con el valor mínimo 1, por lo que este cuantificador con el grado mínimo 0.5 garantiza la Participación total de la entidad Proyecto en la interrelación Trabaja_ en. Esta eventualidad es también indicada por la doble línea que une la entidad con la interrelación.

Por otra parte, el número de empleados en cada proyecto está restringido a un máximo de aproximadamente 8 (con grado mínimo 0.25), esto lo muestra la Figura 7.

Figura 7: Representación de dos restricciones difusas de Cardinalidad con notación (min, max).

Obsérvese que nos hemos limitado a representar sólo la restricción de Tipo de correspondencia difusa en este apartado, ya que la restricción de Participación difusa se refiere a todas las instancias de una entidad y no a cada una individualmente. Por eso, la restricción difusa de Participación puede establecerse simultáneamente a la notación (min, max) difusa. Eso mismo, en restricciones clásicas no tiene sentido, pues la notación (min, max) clásica expresa también la Participación clásica. Como ejemplo de esto último, mézclense los Ejemplos 2 y 4 (Figuras 4 y 7), prestando atención en las diferencias semánticas de cada tipo de restricción difusa. La representación gráfica resultante se muestra en la Figura 8.

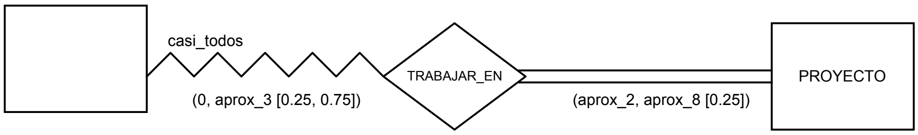

Figura 8: Gráfico imponiendo dos restricciones difusas sobre el mismo arco: Restricción de Participación difusa y Restricción de Cardinalidad Difusa con notación (min, max).

Es fácil deducir que para que ambas restricciones difusas no sean contradictorias el valor mínimo de la notación (min, max) debe ser cero. Un cuantificador difuso de Participación en cierta entidad restringe el número de instancias de esa entidad que se relacionan. Eso implica que pueden existir instancias que no se relacionen (excepto que el cuantificador sea el universal "todos", que no es propiamente difuso). Lo dicho implica que cada instancia puede relacionarse con un mínimo de cero instancias de la otra entidad de la relación. Si el cuantificar de Participación es el universal "todos", la Participación es total y el valor mínimo debe ser 1.

Así tenemos las siguientes notaciones (min, max):

a) (0, aprox_3 [0.25,0.75]) cada empleado trabaja como mínimo cero (o ningún) proyectos, máximo aprox_3[0.25,0.75].
b) (1, aprox_3[0.25,0.75]) cada empleados trabaja como mínimo en un proyecto, máximo aprox_3[0.25,0.75].
c) (casi_todos, aprox_3[0.25,0.75]) cada empleado trabaja con el mínimo valor del cuantificador difuso casi_todos en proyectos, y en un máximo aprox_3[0.25,0.75].

El cuantificador casi_todos para la participación (véase Figura 5) es: "casi todos los empleados están relacionados con proyectos", no siendo lo mismo que lo representado en c). Luego para la ecuación (3):

- ✓ Si min $Q^{-1}(\gamma) > \Phi$ participación es total, caso a)
- ✓ Si min $Q^{-1}(\gamma) \geq \Phi$ participación parcial se puede difuminar la participación difusa como el valor mínimo que representa el cuantificador difuso (en la Figura 2 es 0.4) y queda representado en la Figura 8.

De este modo, si la cardinalidad mínima es "cero" y la participación es parcial es necesario que el cuantificador de ésta quede expresada en el modelo (como lo muestra la Figura 8), pero si la cardinalidad mínima es 1 implica que la participación es total, por lo tanto "no" es necesario colocar la notación del cuantificador (véase Figura 7) a menos que sea una restricción acotada.

En el modelo ER clásico, la notación (min,max) sustituye a las otras dos notaciones para las restricciones de participación y de tipo de correspondencia, ya que si min=0 se trata de una participación parcial y si min>0 se trata de una participación total. Por otra parte, si max=1 la interrelación será 1:1 ó 1:N (en el lado del 1) y si max=N se tratará de una interrelación N:M ó 1:N (en el lado del N).

En cambio, en el modelo ER con restricciones difusas la notación (min,max) añade expresividad al modelo conceptual, pero sólo puede sustituir al tipo de correspondencia difusa si ésta es total, si es parcial necesita incluir en el modelo el cuantificador asociado a la restricción parcial como lo muestra la Figura 8. A continuación discutimos las restricciones tratadas en este trabajo con respecto a la notación (min, max).

3.4 Restricciones de Tipo de Correspondencia y Notación (min,max): Difusa

Con respecto a las restricciones de tipo de correspondencia (apartado 3.2) está claro que la semántica de ambas notaciones es diferente. Observe que en el Ejemplo 3, el cuantificador “aprox_8" indica que un Departamento debe tener aproximadamente 8 empleados, mientras que en el Ejemplo 4, el mismo cuantificador indica que un Departamento puede tener como máximo aproximadamente 8 empleados.

En general, una restricción de tipo de correspondencia difusa con cualquier tipo de cuantificador puede representarse mediante la notación (min,max) difusa haciendo que los dos valores (min,max) tomen el valor de ese tipo de correspondencia difusa. La restricción de tipo de correspondencia difusa expresada en el Ejemplo 3, puede expresarse en notación (min,max) difusa, haciendo que el valor de min sea igual al de max y ambos tomando el valor de “aprox_8".

La expresividad también es equivalente hacia el otro lado con una excepción: Si la notación (min,max) utiliza dos cuantificadores del mismo tipo (absolutos o relativos), entonces esa restricción puede expresarse mediante una restricción de tipo de correspondencia difusa mediante un cuantificador que englobe a ambos. Por ejemplo, la restricción del Ejemplo 4, puede expresarse con la notación de cardinalidad difusa utilizando en lugar de los cuantificadores “aprox_2" y “aprox_8" un cuantificador más amplio como “aprox_entre_2_y_8" (véase Figura 9). Si la notación (min,max) utiliza dos cuantificadores de distinto tipo (uno absoluto y otro relativo), entonces esa restricción no puede expresarse con un único cuantificador de tipo de correspondencia difusa. Esto es debido a que al ser cuantificadores de distinto tipo tienen distinto dominio y no pueden unirse en otro cuantificador que englobe a ambos.

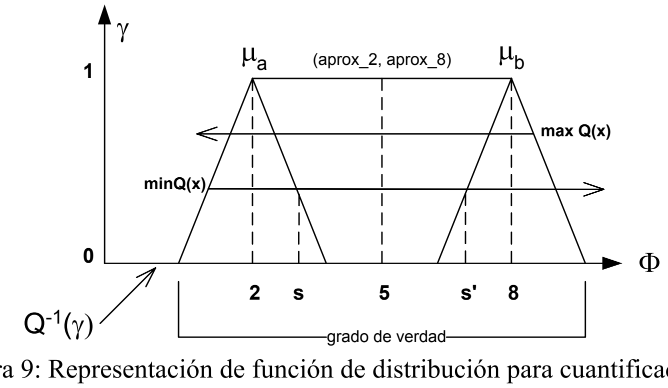

Figura 9: Representación de función de distribución para cuantificadores de cardinalidad (min, max).

Para la Figura 9, $\mu_{(a,b)}(x)$= sup min (μ_a (s), μ_b(s') para todo s ≤ x ≤ s'. La función de distribución entre a y b esta en línea más gruesa.

Debido a su significado, la notación (min,max) utilizará preferentemente cuantificadores absolutos aunque, también aquí son aceptados los relativos, al igual que ocurre en las restricciones de tipo de correspondencia difusa.

3.5 Restricción de Participación y Notación (min,max): Difusa

Por otra parte, no son excluyentes la notación (min,max) y una restricción de participación difusa. Mientras una restricción de participación difusa establece una condición sobre las instancias de una entidad de manera global, la notación (min,max) restringe la interrelación de cada instancia de forma individual con la otra entidad participante.

Por esto, las notaciones más útiles son la notación (min,max) difusa y la notación para las restricciones de participación difusa (con cuantificadores difusos relativos principalmente). La notación para las restricciones de cardinalidad difusa puede eliminarse para ser expresadas utilizando la notación (min,max) difusa.

4 Conclusiones y Trabajos Futuros

La lógica difusa permite acercar el funcionamiento de los sistemas de información al modo de trabajo de los seres humanos, pues las personas manejamos con gran frecuencia conceptos difusos (como "casi todos", "la mayoría", "aproximadamente 8"...) que incluyen cierta imprecisión y que los sistemas informáticos tradicionales no entienden y, por tanto, no pueden utilizar.

Este trabajo amplía los modelos de datos conceptuales a las restricciones difusas existentes en las interrelaciones, de tal forma, que estas características estén presentes en el modelo de datos y no dispersas en las aplicaciones. Hemos expuesto un formalismo para expresar de forma flexible las restricciones que pueden utilizarse en un modelo conceptual basado en el modelo ER, utilizando cuantificadores difusos. Las restricciones difusas expuestas tienen un novedoso significado y aportan gran expresividad al modelo conceptual, incluso fácil de entender por usuarios no técnicos, algo que es fundamental en el modelado conceptual.

Se ha mostrado también como la notación (min,max) puede sustituir las restricciones de tipo de correspondencia difusa y que una restricción de tipo de correspondencia difusa puede sustituir a la notación (min,max) sólo si ambos cuantificadores son del mismo tipo (absolutos o relativos). A pesar de esa equivalencia en la mayoría de los casos, creemos que es preferible usar la notación (min,max) por su mayor claridad.

En cuanto a los cuantificadores, cada uno son restricciones de interrelaciones diferentes, pudiendo ser representados cada uno de tres formas en un modelo de datos, dependiendo de los umbrales que se quiera representar.

Es importante destacar que ya existe un sistema de bases de datos difusas implementado para el SGBD Oracle® (Galindo et al., 1998; Galindo, 1999). Este sistema incorpora tanto la representación en formato relacional de diferentes tipos de datos difusos como la definición de un lenguaje de tratamiento de información difusa llamado FSQL (Fuzzy SQL). La instauración de los diferentes tipos de restricciones difusas consistiría en la creación de algunos disparadores (*triggers*) que aseguren el cumplimiento de las mismas, tal y como se ha explicado. Lógicamente, la eficiencia de un sistema de bases de datos difusas es menor que si no fueran difusas, pero hay que indicar que definiendo los datos difusos de forma lineal (triángulos, trapecios...) las operaciones a realizar son muy simples y no debería notarse apenas la diferencia, con un sistema medianamente potente.

Un interesante estudio para facilitar la tarea de utilizar cuantificadores difusos por parte de los diseñadores será clasificar los cuantificadores que pueden utilizarse en lenguaje natural y estudiar sus relaciones. Como se ha indicado anteriormente una misma restricción puede ser establecida por varios cuantificadores difusos y, en ese caso, la utilización de ciertos cuantificadores condiciona y limita la posibilidad de usar otros en la misma restricción.

Los siguientes pasos serán aplicar el modelo EER a un caso real (Urrutia et al. 2002) y generalizar la aplicación de la lógica difusa a metodologías de modelos basadas en objetos, a través de UML. Además, se requiere un esfuerzo extra para unificar las diferentes tendencias citadas en la introducción y otras tendencias que buscan modelar datos difusos, como el presentado en Urrutia y Piattini (2001) de transformación de datos imprecisos a etiquetas lingüísticas modeladas como atributos multivaluados derivados.

Referencias

- Chaudhry N., Moyne J., Rundensteiner E. (1994). A Design Methodology for Databases with uncertain Data. 7th International Working Conference on Scientific and Statistical Database Management, Charlottesville, VA, (www.mitexsolutions.com).
- Chen P. (1976). The Entity-Relationship Model Toward a Unied View of Data. *ACM Transactions on Database Systems* (TODS) 1, 1, pp. 9-36.
- Connolly T., Begg C., Strachon A. (1998). Data Bases System, a Practical aprroch to design, implementation and management. Second edition, Addison Wesley.
- De Miguel A., Piattini M., Marcos E. (1999). Diseño de bases de datos relacionales. Rama.
- Elmasri R., Navathe S. B. (2000). Fundamentals of Database Systems. Third Edition. Addison Wesley.
- Galindo J., Medina M., Pons O., Cubero J.C. (1998). A Server for Fuzzy SQL Queries, in Flexible Query Answering Systems. Eds. T. Andreasen, H. Christiansen and H.L. Larsen, Lecture Notes in Artificial Intelligence (LNAI) 1495, pp. 164-174. Ed. Springer.
- Galindo J. (1999), *Tratamiento de la Imprecisión en Bases de Datos Relacionales: Extensión del Modelo y Adaptación de los SGBD Actuales*. Ph. Doctoral Thesis, University of Granada (Spain). (www.lcc.uma.es).
- Galindo J., Medina J.M., Cubero J.C., García M.T. (2001), Relaxing the Universal Quantier of the Division in Fuzzy Relational Databases. *International Journal of Intelligent Systems*,16-6, pp. 713-742.
- Hammer M., McLeod D. (1981). Database Description with SDM: A Semantic Data Model. ACM *Transactions on Database Systems* (TODS) 6, 3.

- Kerre E.E., Chen G.Q. (1995). An Overview of Fuzzy Data Models. In “Studies in Fuzziness: Fuzziness in Database Management Systems”, pp. 23--41. Eds. P. Bosc and J. Kacprzyk. Physica-Verlag.
- Kerre E.E., Chen G. (2000). Fuzzy Data Modeling at a Conceptual Level: Extending ER/EER Concepts. In “Knowledge Management in Fuzzy Databases”, pp. 3-11. Eds. O. Pons, M.A. Vila and J. Kacprzyk. Ed. Physica-Verlag. ISBN: 3-7908-1255-2.
- Ma Z.M., Zhang W.J., Ma W.Y., Chen G.Q. (2001). Conceptual Design of Fuzzy Object-Oriented Databases Using Extended Entity-Relationship Model. International Journal of Intelligent Systems, 16(6), pp. 697--711.
- Marín N., Pons O., Vila M.A. (2000). Fuzzy Types: A New Concept of Type for Managing Vague Structures. *International Journal of Intelligent Systems*, 15, pp. 1061-1085.
- Medina J.M., Pons O., Vila M.A. (1994). GEFRED. A Generalized Model of Fuzzy Relational Databases. *Information Sciences*, 76(1-2), pp. 87-109.
- Pedrycz W., Gomide F. (1998). An Introduction to Fuzzy Sets: Analysis and Design. A Bradford Book. ISBN 0-262-16171-0. The MIT Press, Massachusetts.
- Petry F.E. (1996). Fuzzy Databases: Principles and Applications (with chapter contribution by Patrick Bosc). *International Series in Intelligent Technologies*. Ed. H.J. Zimmermann. Kluwer Academic Publ. (KAP).
- Urrutia A., Piattini M. (2001). Transformation of Imprecise Data to Linguistic Label for ER Models. 7th International Conference on Information System Analysis and Synthesis (ISAS2001), Orlando (USA).
- Urrutia A., Galindo J. (2002). Algunos Aspectos del Modelo Conceptual EER Difuso: Aplicación al Caso de una Agencia Inmobiliaria. XI Congreso Español sobre Tecnologías y Lógica Fuzzy (ESTYLF'2002), pp. 359-364. León (Spain).
- Varas M., Contreras R., Campos D. (1998). Constraints in Generalization Structures in Conceptual Database Schemes. Conferencia Internacional de la Sociedad Chilena de Ciencia de la Computación, SCCC'98 Antofagasta (Chile). (www.inf.udec.cl/~mvaras).
- Vert G., Morris A., Stock M. (2000). Jankowski P., Extending Entity-Relationship Modelling Notation to Manage Fuzzy Datasets. 8th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'2000, pp. 1131-1138 Madrid (Spain).
- Yager R.R. (1983). Quantified Propositions of a Linguistic Logic. *International Journal of Man-Machine* Studies, 19, pp. 195-227.
- Zadeh L.A. (1965). Fuzzy Sets. Information and Control, 8, pp. 338-353.
- Zadeh L.A. (1983). A Computational Approach to Fuzzy Quantifiers in Natural Languages. Computer Mathematics with Applications, 9, pp. 149-183.
- Zvieli, P. Chen (1986). ER Modeling and Fuzzy Databases. 2nd International Conference on Data Engineering, pp. 320-327.