

## UN SRI LINGÜÍSTICO DIFUSO BASADO EN EL OPERADOR LOWA

- Autor:** E. Herrera-Viedma, J. Domínguez  
Dpto. de Ciencias de la Computación e I. A., Universidad de Granada, Granada, España  
viedma@decsai.ugr.es; jdoming@fedro.ugr.es
- Abstract:** A fuzzy linguistic information retrieval system (IRS) defined using an ordinal fuzzy linguistic approach is proposed. The system accepts linguistic weighted Boolean queries. Then, the queries are combinations of index terms which are connected by the logical operators AND, OR and NOT; and terms are weighted by means of ordinal linguistic values according a threshold semantic. This semantic is used to express qualitative restrictions on the documents retrieved for a given term, i.e., restrictions on the index term weights that appear in the document representations. The Boolean connectives AND and OR are modelled by means of the linguistic aggregation operator, the LOWA operator, which introduces a soft computing in the evaluation subsystem of linguistic IRS. The advantage of this linguistic IRS with respect to others is that the use of the LOWA incorporates more flexibility in the retrieval process<sup>1</sup>.
- Key words:** Information retrieval; linguistic modelling; linguistic weighted queries; linguistic aggregation.
- Resumen** Se propone un sistema de recuperación de información lingüístico difuso diseñado usando una aproximación lingüística difusa ordinal. Este sistema acepta consultas Booleanas ponderadas lingüísticas. Entonces, las consultas son combinaciones de índices conectados mediante los operadores lógicos Booleanos AND, OR y NOT, y los índices son ponderados mediante etiquetas lingüísticas ordinales expresando restricciones semánticas umbrales o límite. Estas restricciones umbrales son restricciones cualitativas requeridas a los documentos que se recuperan para cada índice, es decir, son condiciones que deben satisfacer los pesos de los índices en las representaciones de los documentos. Los operadores Booleanos AND y OR se modelan mediante el operador de agregación de información lingüística LOWA el cual realiza una computación flexible en la evaluación del sistema de recuperación de información lingüístico propuesto. Su ventaja con respecto a otros sistemas es que la aplicación del operador LOWA incorpora mucha más flexibilidad en el proceso de recuperación de información.
- Palabras Clave:** Recuperación de información; modelado lingüístico; consultas ponderadas lingüísticas; agregación de información lingüística.

---

<sup>1</sup> Esta investigación es financiada por la CICYT y por el proyecto PB-981305

## 1. Introducción

La Recuperación de la Información (RI) es un área de trabajo situada entre las Ciencias de la Información y las Ciencias de la Computación, dedicada al almacenamiento y recuperación de documentos. Un Sistema de RI (SRI) tiene tres componentes:

1. *Una Base de Datos*: para almacenar los documentos y la representación de sus contenidos informativos cuyos componentes son los términos índices y sus pesos. Estos pesos indican lo que un documento representa los conceptos inducidos por los términos índice.
2. *El Sistema de Consulta*: tiene un lenguaje de consultas que permite a los usuarios especificar sus necesidades de información mediante la formulación de consultas.
3. *Un Sistema de Evaluación*: evalúa los documentos para cada consulta mediante una función de emparejamiento y asigna un Valor de Estado de Recuperación (RSV) a cada documento.

Los SRI Booleanos constituyen la base de la mayoría de los SRI comerciales [4, 11]. Las consultas Booleanas son combinaciones de términos índices conectados mediante los operadores lógicos AND, OR y NOT. Normalmente, estos operadores se modelan usando la unión de conjuntos, la intersección de conjuntos y el complemento de un conjunto, respectivamente. Los SRI Booleanos presentan algunos inconvenientes [11]: en las consultas no se puede ponderar lo significativo que un término puede ser; la evaluación de los operadores lógicos es muy estricta; y es difícil clasificar los documentos recuperados por juicios de relevancia. Algunos SRI difusos [2, 3, 5] superan estas limitaciones usando valores numéricos en las consultas y asignando un RSV numérico a cada documento recuperado. Sin embargo, parece más normal describir los contenidos de los documentos que se buscan usando etiquetas lingüísticas para ponderar lo significativo que cada término índice es en la consulta, etiquetas como “importante” o “muy importante”, en lugar de valores numéricos. Por otro lado, las evaluaciones de los operadores AND y OR son realizadas mediante la *t-norma* Min y la *t-conorma* Max, y éstos presentan el inconveniente de que operan muy restrictivamente y muy inclusivamente, respectivamente.

Aquí, presentamos un SRI lingüístico difuso que soporta consultas Booleanas lingüísticas ponderadas y que soluciona el inconveniente de los operadores. Se define una aproximación lingüística difusa ordinal como en [8]. Los términos de las consultas pueden ser ponderados mediante etiquetas lingüísticas interpretadas como *límites significativos* [8]. Estos pesos lingüísticos son interpretados como límites a satisfacer en el emparejamiento entre documentos y consultas. Los operadores Booleanos AND y OR se modelan mediante el operador de agregación de información lingüística LOWA [9]. Él es un operador “or-and” que permite flexibilizar el proceso de evaluación de consultas. Los

documentos recuperados se clasifican en clases de relevancia identificadas por etiquetas lingüísticas.

El artículo se estructura así. El enfoque lingüístico difuso y el operador LOWA se presentan en la Sección 2. El SRI se define en la Sección 3. Un ejemplo se muestra en la Sección 4. Finalmente, la Sección 5 presenta algunas conclusiones.

## 2. El Enfoque Lingüístico Difuso Ordinal

Esta técnica es una técnica aproximada diseñada para modelar los aspectos cualitativos en los problemas [8]. En un enfoque lingüístico difuso ordinal asumimos un conjunto de etiquetas finito y  $S = \{s_i, i \in H = 0, K, T\}$  totalmente ordenado en el sentido normal y con un cardinal impar (7 o 9 etiquetas) [1]. La etiqueta central representa una valoración de "aproximadamente 0.5" y el resto se sitúan simétricamente alrededor de ella [1]. La semántica de las etiquetas se define en base a la estructura ordenada que forman el conjunto de etiquetas considerando que el par  $(s_i, s_{T-i})$  es igualmente informativo. Cada etiqueta tiene asociado un número difuso sobre  $[0,1]$ , que se define mediante una función de pertenencia trapezoidal representada por una 4-tupla  $(a_i, b_i, \alpha_i, \beta_i)$  (los primeros dos parámetros indican el intervalo en el que el valor de pertenencia es 1; el tercero y cuarto indican la amplitud a la izquierda y a

$$1. - \text{Orden} : s_i \geq s_j \quad \text{si} \quad i \geq j.$$

$$2. - \text{Negación} : \text{Neg}(s_i) = s_j, \text{ con } j = T - i.$$

$$3. - \text{Máximo} : \text{MAX}(s_i, s_j) = s_i \text{ si } s_i \geq s_j.$$

$$4. - \text{Mínimo} : \text{MIN}(s_i, s_j) = s_i \text{ si } s_i \leq s_j.$$

la derecha de la distribución). Además, requerimos las siguientes propiedades:

Por ejemplo, podemos usar el siguiente conjunto de nueve etiquetas con su semántica asociada para evaluar las variables lingüísticas del SRI [1]:  $\{T = \text{Total} = (1, 1, 0, 0), EH = \text{Extremely\_High} = (.98, .99, .05, .01), VH = \text{Very\_High} = (.78, .92, .06, .05), H = \text{High} = (.63, .80, .05, .06), M = \text{Medium} = (.41, .58, .09, .07), L = \text{Low} = (.22, .36, .05, .06), VL = \text{Very\_Low} = (.1, .18, .06, .05), EL = \text{Extremely\_Low} = (.01, .02, .01, .05), N = \text{None} = (0, 0, 0, 0)\}$ .

### 2.1 El Operador LOWA

El operador LOWA es un operador de agregación de valores lingüísticos ordenados basado en la computación simbólica [9]. El opera por computación directa sobre las etiquetas considerando solamente su orden y no su semántica asociada.

**Definición 1.** Sea  $A = \{a_1, \dots, a_m\}$  un conjunto de etiquetas para agregar, entonces el operador LOWA,  $\Phi$ , se define como  $\Phi(a_1, \dots, a_m) = W \cdot B^T = C^m \{w_k, b_k, k = 1, \dots, m\} = w_1 \ominus b_1 \oplus (1 - w_1) \ominus C^{m-1} \{\beta_h, b_h, h = 2, \dots, m\}$ , donde  $W$

$= [w_1, \dots, w_m]$ , es un vector de pesos, tal que,  $w_i \in [0, 1]$  y  $\sum_i w_i = 1$ .  $\beta_h = w_h / (\sum_{k=1}^m w_k)$ ,  $h = 2, \dots, m$ , y  $B = \{ b_1, \dots, b_m \}$  es un vector ordenado asociado a  $A$ , tal que,  $B = \sigma(A) = \{ a_{\sigma(1)}, \dots, a_{\sigma(m)} \}$ , donde,  $a_{\sigma(j)} \leq a_{\sigma(i)} \forall i \leq j$ ,  $\sigma$  es una permutación sobre  $A$ .  $C^m$  es el operador de combinación convexa de  $m$  etiquetas y si  $m=2$ , entonces  $C^2\{ w_i, b_i, i = 1, 2 \} = w_1 \ominus s_j \oplus (1 - w_1) \ominus s_i = sk$ , tal que  $k = \min \{ T, i + \text{round}(w_1 \cdot (j - i)) \}$ ,  $s_j, s_i \in S$ , ( $j \geq i$ ), siendo "round" el operador de redondeo, y  $b_1 = s_j$ ,  $b_2 = s_i$ . Si  $w_j = 1$  y  $w_i = 0$  con  $i \neq j \forall i$ , entonces  $C^m\{ w_i, b_i, i = 1, \dots, m \} = b_j$ .

El operador LOWA es un operador "or-and" [9]. Por tanto, él opera entre el operador MIN y MAX. En nuestro SRI usamos el LOWA para evaluar las consultas Booleanas. Para clasificar los operadores OWA en función de cómo operan entre el *and* y el *or*, Yager [14] definió una medida de *orness*, asociada al vector  $W$ ,

$$\text{orness}(W) = \frac{1}{m-1} \sum_{k=1}^m (m-k)w_k.$$

Fijado un  $W$ , cuanto más próximo esté el comportamiento del operador OWA a un *or* más cerca estará su medida *orness* a 1; mientras cuanto más próximo esté a un *and*, más cerca estará a 0. Normalmente, un operador OWA con muchos pesos no nulos en las primeras posiciones se comporta como un *or* ( $\text{orness} \geq 0.5$ ), mientras que cuando están en las últimas posiciones se comporta como un *and*.

### 3. El SRI Lingüístico

En esta sección, presentamos el SRI lingüístico que maneja consultas Booleanas lingüísticas ponderadas y modela los operadores Booleanos de manera flexible.

#### 3.1 La Base de Datos

La base de datos almacena los documentos  $D = \{d_1, \dots, d_m\}$  y sus respectivas representaciones  $R = \{R_1, \dots, R_{d_m}\}$ . Los documentos se representan mediante términos índices  $T = \{t_1, \dots, t_j\}$ . Cada término tiene asociado un peso [13] que describe el contenido de los documentos.  $F : D \times T \rightarrow [0,1]$  es una función de indexación numérica que asigna a cada documento  $d_j$  y a cada término índice  $t_i$  un peso numérico entre 0 y 1.  $F(d_j, t_i)$  es un peso que representa en que grado es significativo  $t_i$  en  $d_j$ .  $F(d_j, t_i) = 0$  implica que el documento  $d_j$  no es representado por  $t_i$ , y  $F(d_j, t_i) = 1$  implica que el documento está perfectamente representado por  $t_i$ .  $F(d_j, t_i) \in (0,1)$  representa los diferentes grados significativos intermedios. La calidad de la recuperación depende mucho del criterio seguido para calcular  $F$ . Pueden usarse diferentes métodos para obtener  $F$  [3, 12]. Aquí no estudiaremos este aspecto, asumimos cualquiera de los existentes.

#### 3.2 El Sistema de Consultas Lingüístico

Presentamos un sistema de consultas basado en un lenguaje de consultas Booleano ponderado. Entonces, cada consulta se expresa como una combinación

de términos índices ponderado, conectados mediante los operadores Booleanos AND ( $\wedge$ ), OR ( $\vee$ ), y NOT ( $\neg$ ), y ponderados mediante etiquetas lingüísticas. Como en [4], usamos la variable lingüística *Importancia* para modelar los pesos lingüísticos, pero definida mediante el enfoque lingüístico difuso como hicimos en [8]. Por tanto, consideramos un conjunto de etiquetas ordenadas S para expresar los pesos lingüísticos.

Entonces, una consulta es cualquier expresión Booleana legítima cuyos componentes atómicos (átomos) son 2-tuplas  $\langle t_i, c_i \rangle$  pertenecientes al conjunto,  $T \times S$ ;  $t_i \in T$ ,  $c_i$  es un valor de la variable lingüística *Importancia*, asociado a una *semántica de límite*. Por tanto, el conjunto Q de consultas legítimas queda definido mediante estas reglas sintácticas:

1.  $\forall q = \langle t_i, c_i \rangle \in T \times S \rightarrow q \in Q$ .
2.  $\forall q, p \in Q \rightarrow q \wedge p \in Q$ .
3.  $\forall q, p \in Q \rightarrow q \vee p \in Q$ .
4.  $\forall q \in Q \rightarrow \neg(q) \in Q$ .
5. Todas las consultas legítimas  $q \in Q$  sólo son aquellas obtenidas aplicando las anteriores reglas.

### 3.2.1 La Semánticas de los Pesos

Asumimos que los pesos de los términos de las consultas se interpretan según una *semántica de límite simétrica* [8]. De acuerdo a esta semántica los pesos son requerimientos que deben satisfacer para cada término en el proceso de emparejamiento entre consultas y documentos. Con esta semántica un usuario puede buscar documentos exigiendo una presencia mínimamente aceptable de un término en sus representaciones [10] o documentos con una ausencia máximamente aceptable de un término. Entonces, cuando un usuario busca documentos en los que los conceptos representados por un término  $t_i$  tengan un valor de *High Importance*, él aceptaría cualquier documento con un valor F mayor que *High*; por el contrario, cuando busca documentos en los que los conceptos representados por un término  $t_i$  tengan un valor de *Low Importance*, él aceptaría cualquier documento con un valor F menor que *Low*. En una consulta  $\langle t_i, c_i \rangle$ , los pesos que implican presencia de un término en la representación de los documentos, esto es  $c_i \geq s_{T/2}$  (e.g. *High, Very High*), deben ser tratados de forma diferente a los que implican ausencia, esto es  $c_i < s_{T/2}$  (e.g. *Low, Very Low*). Entonces, si  $c_i \geq s_{T/2}$  la consulta  $\langle t_i, w_i \rangle$  es sinónima de la consulta  $\langle t_i, \text{al menos } c_i \rangle$ , lo que significa que los documentos deseados son aquellos en los que F tome valores tan altos como sea posible; y si  $c_i < s_{T/2}$  es sinónima de la consulta  $\langle t_i, \text{como mucho } c_i \rangle$ , lo que significa que los documentos deseados son aquellos en los que F tome valores tan bajos como sea posible. Por ejemplo, un usuario que realiza esta consulta  $q = ((t_5, V H) \wedge (t_6, L)) \vee (t_7, H)$  está expresando su interés en un conjunto de documentos formado, por un lado, con un número de componentes tratando al menos en un muy alto grado con los conceptos representados por  $t_5$  y con un número de componentes tratando como mucho en un grado bajo con los

conceptos representados por  $t_6$ , y por otro lado, con un número de componentes tratando al menos con un grado alto con los conceptos representados por  $t_7$ .

### 3. 3 El Sistema de Evaluación Lingüístico

Este sistema se encarga de evaluar los documentos en términos de su relevancia frente a una consulta. Normalmente, los sistemas de evaluación de consultas Booleanas actúan mediante un *proceso constructivo de abajo arriba* basándose en la *propiedad de separabilidad* ( uno de los más importantes criterios de la *lista de deseos*) [7,13]. Esta propiedad establece que la evaluación de un átomo de una consulta debe ser completamente independiente de las evaluaciones de los átomos restantes. Si esta propiedad es verdad, entonces el sistema de evaluación primero evalúa los átomos de la consulta, luego las combinaciones Booleanas de los átomos, y de este modo trabajando de abajo arriba el proceso se repite hasta que la consulta completa se termina evaluando [3].

De acuerdo a la interpretación de la *semántica de límite simétrica* es obvio que no existe ninguna relación entre los átomos de una consulta, y por tanto, la evaluación de un átomo es totalmente independiente de la evaluación del resto. Entonces, nuestro sistema de evaluación evalúa una consulta Booleana ponderada lingüística en dos pasos:

- 1.- Se establece la relevancia de los documentos de acuerdo solamente a los átomos de la consulta aplicando la interpretación de la *semántica de límite simétrica*. Para cada documento se obtiene un RSV parcial calculado con respecto a cada átomo.
- 2.- Se establece la relevancia de los documentos de acuerdo a las combinaciones Booleanas de átomos, y el proceso se repite trabajando de abajo arriba hasta que se obtiene la relevancia de los documentos con respecto a la consulta completa. Para cada documento se obtiene un RSV total calculado con respecto a toda la consulta.

Normalmente, en los SRI Booleanos difusos los conectivos AND y OR se modelan mediante la t-norma Min y la t-conorma Max. Como es bien conocido, el problema de estos operadores de evaluación es que son muy restrictivos y muy inclusivos, respectivamente. Para evitar este problema, proponemos sustituir estos operadores en la evaluación de los conectivos lógicos por el operador LOWA, que actúa entre ambos, e incorpora el concepto de computación flexible en el sistema de evaluación.

#### 4. 3. 1 La Función de Emparejamiento Lingüística $E^*$

El sistema de evaluación lingüística se modela mediante una función de emparejamiento lingüística  $E^*: Q \times D \rightarrow S$ , que, para una consulta  $q \in Q$  asigna a cada documento  $d_j \in D$  valor lingüístico ordinal  $RSV_j = E^*(q, d_j) \in S$ .  $E^*$  se define recursivamente mediante las siguientes reglas:

1.-  $E^*(q, d_j) = g^1(Label(F(d_j, t_i)), c_i), \forall q = \langle t_i, c_i \rangle, \forall j$ , donde *Label* es una función que asigna una etiqueta de  $S$  a un valor numérico  $r \in [0, 1]$ , y es definida como  $Label(r) = \text{Sup}_q \{ \{s_q \in S : \mu_{s_q}(r) = \text{Sup}_v \mu_{s_v}(r)\} \}$ ;  $g^1$  es una función de emparejamiento lingüística que interpreta la semántica de límite simétrica como se muestra en la siguiente tabla [8]:

$g^1(s_a, s_b) =$	
$s_0$	Si $s_b \geq s_{T/2} \wedge s_a = s_0$
$s_{i1}$	Si $s_b \geq s_{T/2} \wedge s_0 < s_a < s_b$
$s_{i2}$	Si $s_b \geq s_{T/2} \wedge s_b \leq s_a < s_T$
$s_T$	Si $s_b \geq s_{T/2} \wedge s_a = s_T$
$s_T$	Si $s_b \geq s_{T/2} \wedge s_a = s_0$
$Neg(s_{i1})$	Si $s_b \geq s_{T/2} \wedge s_0 < s_a < s_b$
$Neg(s_{i2})$	Si $s_b \geq s_{T/2} \wedge s_b < s_a < s_T$
$s_0$	Si $s_b \geq s_{T/2} \wedge s_a = s_T$

$$i_1 = \text{Max}\{ 0, \text{round}(b - ((b-a)/K))\}$$

$$K \in \{ 2, 3, 4, \dots, b \}.$$

$$i_2 = \text{Min}\{ T, \text{round}(b + ((b-a)/K))\}$$

**Nota 1:**  $g^1$  es una función de emparejamiento diferente a las clásicas funciones asociadas a la semántica de límite las cuales son monótonas crecientes. Ella es simétrica respecto a los valor límite central ( $s_{T/2}$ ), lo que significa que  $g^1$  es creciente en *Label*( $F$ ) para los valores límites mayores que  $s_{T/2}$  y decreciente en *Label*( $F$ ) para los valores límite que son menores que  $s_{T/2}$ .  $K$  es un parámetro de sensibilidad definido para controlar la importancia de la proximidad o distancia entre *Label*( $F$ ) y  $s_b$  en el resultado final. Cuanto mayor es el valor de  $K$ , más pequeña es la importancia de la distancia.  $K = 1$  significa que la semántica de límite simétrica no es aplicada.

2.- *Las consultas negadas*,  $\neg q$ . Consideramos que el sistema de evaluación sólo puede encontrarse con átomos ponderados negados,  $\langle \neg t_i, c_i \rangle$ . Esto puede conseguirse aplicando las leyes de De Morgan's laws a cualquier consulta. Esto simplifica la definición del NOT. Cuando hay que evaluar un átomo ponderado negado hay dos posibilidades, se puede negar el término ponderado o ponderar el término negado. Como se hizo en [6], optamos por la última posibilidad. Por tanto,

la evaluación de un documento  $d_j$  para un átomo ponderado negado  $\langle \neg t_i, c_i \rangle$  se define desde la negación del valor  $\text{Label}(F(d_j, t_i))$ . Es decir,  $E^*(q, d_j) = g^1(\text{Neg}(\text{Label}(F(d_j, t_i))), c_i), \forall q = \langle t_i, c_i \rangle, \forall j$ .

3.-  $E^*(\bigwedge_k^{M \geq 2} q_k, d_j) = \Phi(E^*(q_1, d_j), \dots, E^*(q_M, d_j)), \forall j$ , usando un vector de ponderación del operador LOWA,  $W$ , tal que  $\text{orness}(W) < 0.5$ .

4.-  $E^*(\bigvee_k^{M \geq 2} q_k, d_j) = \Phi(E^*(q_1, d_j), \dots, E^*(q_M, d_j)), \forall j$ , usando un vector de ponderación del operador LOWA,  $W$ , tal que  $\text{orness}(W) \geq 0.5$ .

**Nota 2:** Los RSV son valores lingüísticos tomados de la variable lingüística "*Importance*", pero con una diferente interpretación a la dada en los pesos de las consultas. En este caso, esta variable lingüística se usa para representar el concepto de relevancia como en [4,8]. Por ejemplo, un documento  $d_j$  con un valor  $RSV_j = \text{High}$  significa que el documento satisface una consulta con un grado de relevancia alto.

#### 4. Ejemplo

Supongamos una base de datos pequeña que tiene el siguiente conjunto de términos índices  $T = \{t_1, \dots, t_{10}\}$  y un conjunto de siete documentos  $D = \{d_1, \dots, d_7\}$ . Los documentos son indexados mediante una función de indexación  $F$  que asigna los siguientes pesos a cada término en los documentos:  $R_{d1} = 0.7/t_5 + 0.4/t_6 + 1/t_7$ ;  $R_{d2} = 1/t_4 + 0.6/t_5 + 0.8/t_6 + 0.9/t_7$ ;  $R_{d3} = 0.5/t_2 + 1/t_3 + 0.8/t_4$ ;  $R_{d4} = 0.9/t_4 + 0.6/t_6 + 1/t_7$ ;  $R_{d5} = 0.7/t_3 + 1/t_4 + 0.4/t_5 + 0.8/t_9 + 0.6/t_{10}$ ;  $R_{d6} = 1/t_5 + 0.99/t_6 + 0.8/t_7$ ;  $R_{d7} = 0.8/t_5 + 0.02/t_6 + 0.8/t_7 + 0.9/t_8$ . Consideremos el conjunto de etiquetas lingüísticas dado en la Sección 2 para expresar los valores de la variable lingüística *Importance*. Supongamos que el SRI recibe la siguiente consulta,  $q = ((t_5, VH) \wedge (t_6, L)) \vee (t_7, H)$ .

Primero, obtenemos la representación de los documentos expresada en forma lingüística usando la función de traslación *Label*:  $R_{d1} = H/t_5 + M/t_6 + T/t_7$ ;  $R_{d2} = T/t_4 + M/t_5 + H/t_6 + VH/t_7$ ;  $R_{d3} = M/t_2 + T/t_3 + H/t_4$ ;  $R_{d4} = VH/t_4 + VL/t_6 + T/t_7$ ;  $R_{d5} = H/t_3 + T/t_4 + M/t_5 + H/t_9 + M/t_{10}$ ;  $R_{d6} = T/t_5 + EH/t_6 + H/t_7$ ;  $R_{d7} = H/t_5 + EL/t_6 + H/t_7 + VH/t_8$ . Supongamos que usamos el parámetro de sensibilidad  $K=2$ , que da una gran importancia a la distancia entre los valores lingüísticos en  $g^1$ . Entonces, las evaluaciones de los átomos de acuerdo a la semántica de límite simétrica expresada por  $g^1$  son:

1.  $E^*(\langle t_5, VH \rangle, d_1) = VH$ ,  $E^*(\langle t_5, VH \rangle, d_2) = H$ ,  $E^*(\langle t_5, VH \rangle, d_5) = H$ ,  $E^*(\langle t_5, VH \rangle, d_5) = H$ ,  $E^*(\langle t_5, VH \rangle, d_6) = T$ ,  $E^*(\langle t_5, VH \rangle, d_7) = VH$ .
2.  $E^*(\langle t_6, L \rangle, d_1) = M$ ,  $E^*(\langle t_6, L \rangle, d_2) = M$ ,  $E^*(\langle t_6, L \rangle, d_4) = VH$ ,  $E^*(\langle t_6, L \rangle, d_6) = L$ ,  $E^*(\langle t_6, L \rangle, d_7) = VH$ .
3.  $E^*(\langle t_7, H \rangle, d_1) = T$ ,  $E^*(\langle t_7, H \rangle, d_2) = VH$ ,  $E^*(\langle t_7, H \rangle, d_4) = T$ ,  $E^*(\langle t_7, H \rangle, d_6) = H$ ,  $E^*(\langle t_7, H \rangle, d_7) = H$ .

La evaluación de la sub-expresión  $(t_5, VH) \wedge (t_6, L)$  usando  $\Phi$  con  $W = [0.3, 0.7]$  es:



1.  $E^*(\langle t_5, VH \rangle \wedge \langle t_6, L \rangle, d_1) = \Phi(VH, M) = H$ ; 2.  $E^*(\langle t_5, VH \rangle \wedge \langle t_6, L \rangle, d_2) = \Phi(H, M) = M$ ;
3.  $E^*(\langle t_5, VH \rangle \wedge \langle t_6, L \rangle, d_4) = \Phi(N, VH) = VL$ ; 4.  $E^*(\langle t_5, VH \rangle \wedge \langle t_6, L \rangle, d_5) = \Phi(H, N) = VL$ ;
5.  $E^*(\langle t_5, VH \rangle \wedge \langle t_6, L \rangle, d_6) = \Phi(H, L) = M$ ; 6.  $E^*(\langle t_5, VH \rangle \wedge \langle t_6, L \rangle, d_7) = \Phi(H, VH) = H$ .

Finalmente, la consulta completa se evalúa usando el operador  $\Phi$  con  $W=[0.7, 0.3]$ :

1.  $E^*(q, d_1) = \Phi(H, T) = EH$ ; 2.  $E^*(q, d_2) = \Phi(M, VH) = H$ ; 3.  $E^*(q, d_4) = \Phi(VL, T) = VH$ ;
4.  $E^*(q, d_5) = \Phi(VL, N) = EL$ ; 5.  $E^*(q, d_6) = \Phi(M, H) = H$ ; 6.  $E^*(q, d_7) = \Phi(H, H) = H$ .

Debería observarse que usando los clásicos operadores Max y Min sería:  $[RSV_1=T, RSV_2=VH, RSV_4=T, RSV_5=N, RSV_6=H, RSV_7=H]$ , y por tanto, el resultado alcanzado con el operador LOWA es menos estricto.

## 5. Conclusiones

En este artículo, hemos presentado un SRI lingüístico definido usando un enfoque lingüístico difuso ordinal que acepta consultas Booleanas ponderadas lingüísticas. Este SRI aprovecha la propiedad de ser un operador or-and del operador LOWA, usado para agregar la información lingüística ordinal. El operador LOWA se aplica para modelar la evaluación de los conectivos lógicos AND y OR en el proceso de evaluación de las consultas. De este modo, el concepto de computación flexible (soft) se introduce en la recuperación del SRI y se consigue eliminar la evaluación demasiado estricta de los clásicos operadores que modelan los conectivos lógicos (t-norma Min y t-conorma Max).

## Referencias

- 1 Bonissone, P.P. & Decker, K.S. (1986). Selecting Uncertainty Calculing and Granularity: An Experiment in Trading-off Precision and Complexity. Kanal, L.H. & Lemmer, J.F. (Eds.), *Uncertainty in Artificial Intelligence*. North-Holland, 217-247.
- 2 Bookstein, A. (1980). Fuzzy Request: and Approach to Weighted Boolean Searches. *Journal of the American Society for Information Science*, 31, 240-247.
- 3 Bordogna, G., Carrara, C. & Pasi, G. (1991). Query Term Weights as Constraints in Fuzzy Information Retrieval. *Information Processing & Management*, 27, 15-26.
- 4 Bordogna, G. & Pasi, G. (1993). A Fuzzy Linguistic Approach Generalising Boolean Information Retrieval: A Model and Its Evaluation. *Journal of the American Society for Information Science*, 44, 70-82.
- 5 Buell, D. & Kraft, D.H. (1981). Threshold Values and Boolean Retrieval Systems. *Information Processing & Management*, 17, 127-136.
- 6 Buell D. & Kraft D.H. (1981). A Model for a Weighted Retrieval System. *Journal of the American Society for Information Science*, 32, 211-216.

- 7 Cater, C.S. & Kraft, D.H. (1989). A Generalization and Clarification of the Waller-Kraft Wish List. *Information Processing & Management*, 25, 15-25.
- 8 Herrera-Viedma, E. (2001). Modelling the Retrieval Process for an Information Retrieval System Using an Ordinal Fuzzy Linguistic Approach. *Journal of the American Society for Information Science*. To appear.
- 9 Herrera, F., Herrera-Viedma, E., & Verdegay J.L. (1996). Direct Approach Processes in Group Decision Making Using Linguistic OWA Operators. *Fuzzy Sets and Systems*, 79, 175-190.
- 10 Kraft, D.H., Bordogna, G., & Pasi, G. (1994). An Extended Fuzzy Linguistic Approach to Generalize Boolean Information Retrieval. *Information Sciences*, 2, 119-134.
- 11 Korfhage, R. R. (1997). *Information Storage and Retrieval*. New York: Wiley Computer Publishing.
- 12 Salton, G. & Buckley, C. (1988). Term Weighting Approaches in Automatic Text Retrieval. *Information Processing & Management*, 24, 513-523.
- 13 Waller, W.G. & Kraft, D.H. (1979). A Mathematical Model of a Weighted Boolean Retrieval System. *Information Processing & Management*, 15, 235-245.
- 14 Yager, R.R. (1988). On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making. *IEEE Transactions on Systems, Man, and Cybernetics*, 18,183-190.