MODELO BORROSO DE UN SISTEMA DE RECUPERACION DE INFORMACION*

Miguel Francisco Buitrago Botero
Ingeniero de Sistemas
Depto. Análisis y Programación
Empresas Públicas de Medellín
A.A. 4928 Medellín, Colombia.

William Jaramillo Jaramillo
Ingeniero de Sistemas
Depto. de Sistemas Delima
A.A. 4928 Medellín, Colombia.

Resumen

En la mayoría de los sistemas de recuperación de información el usuario debe expresar la pregunta en forma de expresiones booleanas. Aunque este tipo de sistemas trabajan, tienen una serie de desventajas que es necesario estudiar y tratar de remediar.

En la literatura conocida hay intentos para aplicar la teoría de conjuntos borrosos en la recuperación de información. En este trabajo explicamos las características de un sistema boolenao de recuperación de información y sus mayores desventajas; posteriormente y dado que el concepto de conjunto borroso es una generalización de la noción convencional de conjunto, proponemos un modelo borroso de un sistema de recuperación de información que soluciona algunas de las desventajas del booleano tradicional. El modelo incluye la forma como se deben expresar las preguntas al sistema y la función de recuperación que es de naturaleza borrosa.

* El presente trabajo es un extracto del proyecto de grado presentado por los autores a la universidad EAFIT.

1. Sistema booleano de recuperación de información

Un sistema booleano de Recuperación de Información (SBRI) es una quintupla [Radecki (R8)]:

$$I = (D, T, P, F, \sigma)$$

En donde D es un conjunto finito de documentos y T un conjunto finito de términos indices o descriptores. Los elementos de P son preguntas que un usuario potencial tiene para el sistema; la sintaxis de las preguntas la conforman los elementos de T conectados por los operadores lógicos Y, O, NO. El siguiente elemento de la quintupla es una relación en el producto cartesiano DxT. Formalmente:

$$F : DxT \longrightarrow \{0,1\}$$

donde F es una función que toma el valor de uno si el documento d trata el tema t y cero si no lo trata.

El último elemento, σ, es una función de la forma:

$$\sigma : DxP \longrightarrow \{0,1\}$$

que asigna para cada par ordenado (d,p); $d \in D$, $p \in P$ un valor en $\{0,1\}$ llamado relevancia formal* del documento para la pregunta p. La relevancia formal expresa el grado de similaridad entre un documento y una pregunta; en otras palabras, indica si el documento d satisface o no la pregunta p. Si $\sigma = 1$, el documento es relevante para la pregunta y el sistema lo recupera; si $\sigma = 0$, el documento no es relevante para la pregunta y por lo tanto no lo recupera.

* Existe diferencia entre relevancia y pertinencia [Radecki (R8)]. Relevancia es una medida de que tan adecuado es el contenido de un documento con respecto al contenido de la pregunta que el usuario formula al sistema. En un sentido amplio, pertinencia es una medida de que tan adecuado es el contenido de un documento con respecto a la necesidad de información de un usuario. El grado de relevancia de cada documento es asignado por un especialista en el área; mientras que el grado de pertinencia es asignado por el usuario. Un documento puede ser muy relevante pero poco pertinente porque el usuario ya esta familiarizado con su contenido. La relevancia formal denota el grado de similaridad, medido por una función de matching, entre la representación del documento y la pregunta formulada al sistema. En otras palabras, el grado de relevancia formal de un documento en la colección lo determina la función de matching que usa el sistema de recuperación de información.

Cuando la pregunta consta de sólo un elemento de T, el cálculo de para cada documento de D es inmediato; cuando la pregunta consta de varios elementos de T unidos por operadores lógicos, el cálculo se hace aplicando las reglas de la lógica matemática:

$$\sigma(d, t_1 \wedge t_2) = \text{Min}[\sigma(d,t_1), \sigma(d,t_2)]$$

$$\sigma(d, t_1 \vee t_2) = \text{Max}[\sigma(d,t_1), \sigma(d,t_2)]$$

$$\sigma(d, \neg t) = 1 - \sigma(d,t)$$

La forma de procesar la pregunta $p_1 = t_1 \wedge t_2$ es encontrar el conjunto de documentos que tratan el tema t_1 e interceptarlo con el conjunto de documentos que tratan el tema t_2. Aquellos documentos que pertenecen a la intersección (valor de relevancia formal igual a 1) se presentan como respuesta a la pregunta p_1. Hemos asociado a cada componente (término índice) de la pregunta un subconjunto de documentos de D y el sistema de recuperación efectua las bien conocidas operaciones de unión, intersección y complementación para corresponder al Y, O, NO de la expresion Booleana. Esta propiedad, que es una de las mayores ventajas de un SBRI se conoce con el nombre de separabilidad.

Despues de calcular la relevancia formal de todos los documentos en la coleccion para una pregunta p, el Sistema selecciona aquellos documentos cuyo valor de σ es igual a 1 y los entrega al usuario; este analiza la salida dada por el sistema y escoge algunos documentos o reformula la pregunta hasta satisfacer completamente su necesidad de información.

El sistema de Recuperación descrito recibe el nombre de Booleano porque el conjunto potencia de D con las operaciones unión ($\cup$) e intersección ($\cap$) satisface las condiciones necesarias para ser un Algebra de Boole, Whitesitt (W2).

2. Fallas de un SBRI

Aunque un SBRI es sencillo, fácil de entender y a nivel operativo el de uso más difundido, tiene ciertas desventajas, Salton (S1), que se resumen en:

2.1. Es difícil controlar la cantidad de documentos que dan respuesta a una pregunta. Dependiendo de la frecuencia de los descriptores en la colección de documentos y la combinación que se use para formular una pregunta, se pueden obtener muchos documentos que la respondan. El sistema no dispone de un mecanismo que permita restringir el número de documentos recuperados; y en caso que existiera, no se puede garantizar que los recuperados son los más importantes.

2.2. La salida que se obtiene en respuesta a una pregunta carece de cualquier orden de presunta importancia para el usuario.

Recordemos que la relevancia formal para un documento recuperado es igual a uno; es decir, todos son teoricamente de igual importancia.

2.3. No es posible asignar algun factor de relevancia de los descriptores con respecto a los documentos. Es posible que tanto el documento d_1 como el documento d_2 traten el tema t pero también es posible que el tratamiento que hace d_1 de t sea cuantitativa y/o cualitativamente diferente del que hace d_2; un SBRI no es capaz de reflejar esta diferencia. Un caso similar se presenta con los términos en las preguntas.

3. Sistema borroso de recuperación de información

Dado que el concepto de conjunto borroso es una generalización de la noción convencional de conjunto, Zadeh (Z1), la generalización de un SBRI a uno con características borrosas se puede hacer fácil y naturalmente, redefiniendo las funciones F y σ.

Las definiciones para D, P y T permanecen invariantes. Las únicas variaciones, aunque muy importantes, ocurren sobre los rangos de F y σ; en vez de definirlos en el conjunto $\langle 0,1 \rangle$ como haciamos en el booleano, lo haremos ahora en el intervalo $[0,1]$.

Un valor real de F en $[0,1]$ detalla la importancia del término t en la repesentación de un documento d. De igual manera, un valor de σ en $[0,1]$ indica la relevancia formal del documento d para la pregunta p. El problema consiste en encontrar un valor apropiado que represente con certeza el grado con que un documento d es relevante para una pregunta p.

Los valores de F parecen fáciles de calcular. En el ambiente de la Recuperación de Información es común escuchar expresiones como: "Este documento trata muy bien el tema", "Este documento es muy simple, te recomiendo mejor este libro", etc. En el primer caso es casi seguro que el valor de la función de pertenencia del documento al tema es alto; en el segundo, el adjetivo simple debe hacer referencia a un valor bajo en el valor de la función de pertenencia, mientras que el libro debe tener un valor un poco más alto.

La situación anterior nos hace pensar que, intuitivamente, tenemos claro el concepto de relevancia de un documento para con un término o viceversa; aunque debemos notar que se presenta un problema al definir el nivel de detalle que se debe alcanzar en el proceso de indización y por lo tanto en el calculo de F.

En un sistema borroso de recuperación, como en el caso booleano, se puede asociar a cada $t \in T$ un conjunto S, borroso en este caso, de documentos relacionados con dicho término; por lo tanto el sistema borroso también satisface la propiedad de separabilidad.

Disponemos ahora de un modelo mucho más flexible, con la garantía que tenemos las mismas propiedades de un SBRI tradicional; de hecho,

en los puntos extremos su comportamiento es análogo. La única excepción es que, en general; $S \cap S' \neq \emptyset$ y $S \cup S \neq D$, debido a la definición de conjunto borroso, pero que no tiene importancia para una pregunta p en el sentido general de la Recuperación de Información.

Indiscutiblemente esta generalización abre todo un mundo de posibilidades a los usuarios de un sistema de recuperación de información.

4. Pesos y umbrales

La generalización de un SBRI a un sistema boleano de recuperación con características borrosas inmediatamente proporciona la facilidad de entregar los documentos que más se acomoden a las necesidades reales del usuario. Otra ventaja es la de poder ordenar la salida por valor de relevancia formal. Esta última le sugiere por ejemplo a un usuario, leer primero aquellos documentos con mayor valor de relevancia formal, pues es donde se sospecha esta la mayor parte de lo que el necesita.

Hasta el momento no hemos hablado de la posibilidad de expresar preguntas del tipo αt, donde $t \in T$ y α es un número posiblemente en $[0,1]$. Existen en la literatura conocida dos interpretaciones del valor α; algunos autores lo llaman umbral y otros lo denominan peso.

Las ideas de peso y umbral nacen del hecho que es muy importante que el usuario pueda precisar la importancia relativa, dentro de una pregunta, de un término con respecto a los demás (peso), o quiera especificar ciertos niveles mínimos de información contenida en los documentos por recuperar (umbrales).

En el presente trabajo y por limitaciones espacio-temporales nos centraremos en el concepto de umbrales. Para mayor información sobre pesos sugerimos leer: Buitrago y Jaramillo (B4), Buell y Kraft (B7), Bookstein (B8).

5. Umbrales

Intuitivamente un umbral es un punto de corte, un tope, un "lugar" de transición. En recuperación de información es común escuchar expresiones del tipo: "Quiero un libro que contenga mucho de Bases de Datos", "El documento d trata poco el tema t", etc. Aquí, las palabras mucho y poco pueden interpretarse como umbrales. Un umbral es entonces un valor linguístico por encima o por debajo del cual sucede algo; en nuestro caso, la existencia o no existencia de documentos.

El concepto de umbral se basa en la noción intuitiva que $F(d,t) > \alpha$, es cualitativamente diferente del caso en el cual $F(d,t) < \alpha$; con α =

umbral. En la primera situación, σ debe reflejar el grado con que la función de pertenencia del documento d al tema t, [F(d,t)], sobrepasa el umbral α; en el segundo, el grado con que se acerca al umbral.

La manera más sencilla de usar umbrales en las preguntas consiste en definir un valor fijo α (umbral) para todos los términos de la pregunta; si un documento tiene una función de pertenencia para el término t mayor o igual que el umbral α ,[$F(d,t) > \alpha$], se le asigna un valor a σ de uno; Si la función de pertenencia es menor que α,σ toma el valor de cero. La salida es muy similar a la de un SBRI tradicional; todos los documentos recuperados tienen el mismo valor de relevancia formal ($\sigma = 1$).

Buell y Kraft (B9) hacen énfasis en que σ debe reflejar el mayor o menor grado con que el documento satisface el umbral asociado a un término t. Con base en este supuesto estudian las características que debería tener la función σ de manera análoga a como lo hace Bookstein (B8) para los pesos. De nuevo las limitaciones de espacio y tiempo nos impiden mostrar los detalles matemáticos de dichas características.

A continuación proponemos un modelo de un sistema de recuperación de información, que utiliza conceptos booleanos, borrosos y de umbrales. A más de resolver algunos problemas del SBRI, es sencillo e intuitivamente pragmatico.

6. Modelo

Como explicamos antes un sistema de recuperación de información es una quintupla de la forma (D,T,P,F,σ). En nuestro modelo, las definiciones para D, T, P y F son las mismas que las descritas en 1.; los elementos de P son, en notación BNF, del tipo:

$$p = (q \ op \ q) / q$$

$$q = at / \neg at$$

$$op = \vee / \wedge$$

$$\alpha \in [0,1]$$

Ejemplos de elementos validos de P son:

$$p_1 = at_1 \vee bt_2$$

$$p_2 = \neg at_1 \wedge bt_2$$

Un umbral α en preguntas del tipo at , indica el límite mínimo de información que debe contener un documento acerca del tema t para que sea recuperado por el sistema; un umbral α en preguntas del tipo $\neg at$, indica el límite máximo de información que debe contener un documento para que sea recuperado.

La definición anterior se basa en el hecho que si at indica el deseo de documentos que hablen de t con un valor de función de pertenencia mayor o igual que a, $\neg at$ debe indicar que no se desean documentos que hablen del término t con un valor mayor o igual que a; o sea que se desean aquellos que hablen de t con un valor menor que a.

En el caso del SBRI, si el término t ($t \in T$) aparece en una pregunta, este hace referencia a todos los documentos que hablen de el (función de pertenencia = 1); si el término t aparece negado, $\neg t$, este indica que no se desean documentos que hablen de él (función de pertenencia = 0). Si en nuestro modelo los umbrales se restrigen a 0 y 1, su comportamiento es análogo al Booleano. De lo anterior se infiere que el sistema booleano es un caso particular de un sistema borroso de recuperación de información.

Esta definición de umbrales permite al usuario expresar preguntas del tipo: "Quiero documentos que hablen mucho de t_1 y no mucho de t_2". En el lenguaje de consulta esta pregunta podría tener la siguiente sintaxis:

$$p = at_1 \wedge \neg bt_2 \text{ ; con } a \geq 0.9 \text{ y } b \leq 0.8$$

Como en el caso booleano, el sistema busca los documentos asociados al término t_1 que cumplen el nivel de umbral exigido (0.8) y los asociados a t_2 que satisfacen el umbral respectivo (0.8); luego efectua la intersección entre estos dos subconjuntos y la entrega al usuario en respuesta a la pregunta inicial, p.

El ultimo elemento de la quintupla, σ, lo definimos así:

$$\sigma = \begin{cases} F(d,t) & \text{si } F(d,t) \geq a \\ 0 & \text{si } F(d,t) < a \end{cases}$$

Para preguntas tipo at

$$\sigma = \begin{cases} 1 - F(d,t) & \text{si } F(d,t) < a \\ 0 & \text{si } F(d,t) \geq a \end{cases}$$

Para preguntas tipo $\neg at$

Para preguntas tipo at, si la pregunta esta conformada por sólo un término, asignamos como valor de relevancia formal el valor F(d,t) que tiene el documento; si la pregunta la conforman varios términos unidos por operadores lógicos, el valor de relevancia formal es el que resulta de aplicar las operaciones de máximo y mínimo. De manera similar se puede razonar para preguntas tipo $\neg at$. Un ejemplo ilustrará mejor esta situación.

Supongamos el siguiente conjunto de documentos y sus valores de función de pertenencia para el término t_1 y t_2

d_1:	$0.8\,t_1$	$0.62\,t_2$	d_4:	$0.9\,t_1$	$0.83\,t_2$
d_2:	$0.5\,t_1$	$0.5\,t_2$	d_5:	$0.6\,t_1$	$0.68\,t_2$
d_3:	$0.7\,t_1$	$0.8\,t_2$			

Si el usuario formula la siguiente pregunta:

$p_1 = 0.15\, t_1$

El sistema respondería:

$d_4 = 0.9$

$d_1 = 0.8$

Obviamente, el documento d_4 como el documento d_1 sobrepasan sus necesidades iniciales; en el primer caso en 0.15 y en el segundo en 0.05.

En el caso que la pregunta sea:

$p_2 = 0.7\, t_1 \wedge 0.7\, t_2$

La respuesta es:

$d_4 = 0.83$

$d_3 = 0.74$

Aquí hemos aplicado operación Min para calcular los valores 0.83 = [Min (0.9,0.83)] y 0.74= [Min (0.74,0.8)].

Veamos ahora en más detalle algunos aspectos interesantes de las respuestas a las preguntas p_1 y p_2.

La respuesta a la pregunta p_1 fueron los documentos d_4 y d_1; pero acaso el documento d_3 no parece ser importante para esta pregunta?. La respuesta es que sí parece ser importante, puesto que casi llega al umbral requerido. sólo le falto 0.01 para cumplirlo ($\alpha = 0.75$; $F(d_3, t_1) = 0.74$).

En el caso de la pregunta p_2 sucede algo similar con el documento d_5 sólo le falto 0.03 para cumplir con el umbral de t_1 y 0.02 para cumplir con el de t_2.

Para obviar la situación anterior, hemos definido una función $\mathcal{E}$, que incluye en los documentos por recuperar aquellos que estan cerca del umbral requerido. Dado que el concepto de cerca es subjetivo y no debe ser el mismo para umbrales altos que para umbrales bajos definimos a $\mathcal{E}$ de tal manera cumpla las siguientes propiedades:

- Para preguntas del tipo αt (positivas), fijamos a $\mathcal{E}$ alto para valores altos de α y bajo para valores bajos de α.

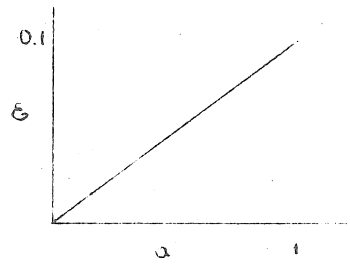
- Para preguntas del tipo $\neg \alpha t$ (negativas), $\mathcal{E}$ debe ser bajo para valores altos de α y alto para valores bajos de α.

La justificación es la siguiente: Sea $S(\alpha, t)$ el conjunto de documentos que satisfacen el umbral α para la pregunta αt; entonces, la cardinalidad de $S(\alpha, t)$ varia inversamente con α; esto es:

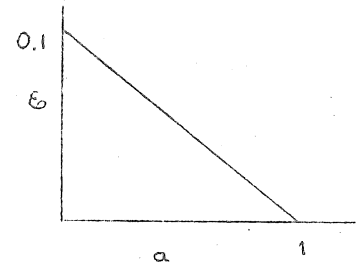
$$|S(a,t)| > |S(b,t)| \quad \text{si } a < b$$

lo cual indica que el número de documentos que satisfacen un umbral alto es menor que el número de documentos que cumplen un umbral bajo; es decir, es más fácil alcanzar un umbral de 0.2 que un umbral de 0.9. Este razonamiento nos lleva a pensar que $\frac{\partial \varepsilon}{\partial a} > 0$ (o sea: ε es una función creciente de a).

Una función que cumple con las características anteriores y que proponemos como punto de partida es:



para preguntas tipo at



para preguntas tipo ¬at

El máximo valor que toma la función epsilon es 0.1. Este valor corresponde a una apreciación subjetiva de los autores.

Un aspecto interesante de la definición de ε es que el caso en el cual en una pregunta aparece un término de la forma ¬at ($a = 0$) es diferente al caso en el cual el término no aparece en la pregunta. En el primero, el sistema recupera aquellos que estan cerca de 0; en el segundo, el sistema no toma ninguna acción al respecto. Por ejemplo,

$$p_1 = at \wedge \neg bt \; ; \; b = 0$$

$$p_2 = at$$

aunque a primera vista las preguntas son iguales, los conjuntos de documentos que las responden no lo son debido a la definición de ε.

Las ventajas que presenta el modelo antes explicado son:

- Conserva las propiedades del SBRI. El usuario puede formular preguntas de una forma más o menos natural por medio de expresiones conformadas por los elementos de T con el nivel de información que nececita (umbral) y los operadores lógicos Y, O, NO. Además, conserva la propiedad de separabilidad que permite descomponer una pregunta en sus términos individuales, seleccionar los subconjuntos correspondientes a cada término y

aplicar luego las operaciones de Max y Min para encontrar el conjunto por recuperar.

La propiedad de separabilidad también permite encontrar formas equivalentes, más fáciles de manipular, para una pregunta. Esto es especialmente útil al momento de procesar preguntas complejas o que involucran muchos operadores lógicos.

- Permite expresar la importancia de un documento para un término. Esta característica hace que el proceso de indización sea más preciso pues ya no es suficiente asignar términos a documentos sino que es necesario cuantificar el grado con que el documento trata el tema.

- De manera análoga, permite expresar, en las preguntas, el grado de información que debe contener un documento para que sea efectivamente recuperado. La consecuencia de esta propiedad es que la cardinalidad del conjunto de documentos recuperados es menor que la cantidad de documentos recuperados en respuesta a la misma pregunta, sin umbrales, por un SBRI.

- Permite ordenar la salida por valor de relevancia formal. Esta es una ventaja muy importante, puesto que el usuario dispone ahora de un orden recomendado de lectura que no proporcionaba el caso Booleano. Ante una cantidad de documentos recuperados, el usuario tiene más elementos para decidir cual debe leer primero.

- Por último, el modelo es general o sea que puede ser manejado como un SBRI tradicional.

Dentro de las desventajas tenemos:

- El hecho que el sistema permita una indización cualitativa además de ser una gran ventaja, puede llegar a convertirse en una desventaja, ya que realizar el proceso de indización con valores de función de pertenencia bajos se puede convertir en una tarea monumental.

- En el caso que se formule una pregunta del tipo τ_{at} con $\alpha = 0$ el sistema recupera aquellos documentos cerca de cero dependiendo del valor que tome μ (en nuestro caso aquellos documentos que tratan el tema t con un valor < 0.1). Si un usuario definitivamente no quiere estos documentos, esto, que inicialmente se planteó como una facilidad, se convierte en una desventaja.

  Una solución es pensar en expresar las preguntas como ternas del tipo: (t, α, β), conectadas por operadores lógicos. En la forma anterior de expresar la pregunta, β es un valor real en el intervalo $[0, 1]$, que opera como un multiplicador de μ. Un valor de β igual a uno no tiene ningún efecto

sobre ω . Un valor de γ^a en (0,1) disminuye el valor de hasta una cantidad igual a $\omega \times \gamma^a$. Un valor de γ^a igual a cero elimina el efecto de ω . De esta manera, un usuario que este interesado en documentos que traten el tema t_1 y no traten el tema t_2 podría expresar la pregunta así:

$$p = (t_1, a, 1) \wedge (\neg t_2, b, 0) \quad \text{con } b = 0$$

7. Conclusiones

El hecho de que la Teoría de Conjuntos clásica sea un subconjunto de la Teoría de Conjuntos Borrososos tiene implicaciones muy importantes en el proceso de conocimiento humano y no debe pasar desapercibido.

En este trabajo mostramos como se puede aplicar la teoría de conjuntos borrosos a la recuperación de información. El modelo descrito es sencillo y aparentemente de fácil implementación.

Es necesario que nuestros paises presten mayor atención a las nuevas metodologías y formas de trabajo, fomenten la creatividad y hagan uso de sus propios recursos.

Bibliografía y referencias

B1 Bookstein, A; Cooper, W. "A MODEL FOR INFORMACION RETRIEVAL SYSTEM". Library Quartely V.46, N.2, pp 153-167, 1976

B2 Bookstein, A. "THE PERILS OF MERGING BOOLEAN AND WEIGHTED RETRIEVAL SYSTEMS", ASIS, V.29, N.3, pp 136-157, 1978.

B3 Bookstein, A. "FUZZY REQUEST: AN APPROACH TO WEIGHTED BOOLEAN RETRIEVAL" ASIS V.31, pp 240-247, Julio 1980.

B4 Buitrago, Miguel F. y Jaramillo W. "CONJUNTOS BORROSOS Y RECUPERACION DE INFORMACION", Proyecto de grado, Universidad EAFIT, 1987

B5 Burgerss, R.S. "WHAT IS INFORMATION RETRIEVAL?", Sociedad de Bibliotecarios de Puerto Rico, Enero-Junio 1972

B6 Buell, Duncan A. "A GENERAL MODEL OF QUERY PROCESSING IN INFORMATION RETRIEVAL SYSTEMS", Information & Management, V.17, N.5, pp 249-262, 1981

B7 Buell, Duncan A. y Kraft, Donald H. "A MODEL FOR WEIGHTED RETRIEVAL SYSTEM", ASIS, V.32, N.1, pp 211-216, 1981

B8 Bookstein, A. "A COMPARISON OF TWO SYSTEMS OF WEIGHTED BOOLEAN RETRIEVAL" ASIS, Julio 1981.

B9 Buell, Duncan A. y Kraft Donald H. "THRESHOLD VALUES AND BOOLEAN RETRIEVAL SYSTEMS", Informacion & Management, V.17, pp 127-136, 1981

C1 Cleverdon, C. W., Mills, J. y KEEN, M., "FACTORS DETERMINING THE PERFORMANCE OF INDEXING SYSTEMS", V.1 -Desing, V.11-Test Result, AQSLIB Cranfield Project, Cranfield, 1966

C2 Craft, W.Bruce. "BOOLEAN QUERIES AND TERM DEPENDENCIES IN PROBABILISTIC RETRIEVAL MODELS", ASIS, V.37, N.2, pp 71-77, 1986

C3 Crouch, D. "A CLUSTERING ALGORITHM FOR LARGE AND DYNAMIC DOCUMENT COLLECTION", Ph.D. Tesis, Southern Methodist University, 1972

M3 Molina, J.F. "APORTES DE LA TEORIA DE SUBCONJUNTOS BORROSOS EN INVESTIGACION OPERACIONAL". Comunicación presentada en el X Congreso Colombiano de Calculo Electronico e Investigación Operacional, Bucaramanga, 1986

R2 Radecki, Tadeusz. "ON THE INCLUSIVENESS OF SYSTEMS FOR RETRIEVAL DOCUMENTS INDEXED BY UNWEIGHTED DESCRIPTORS", Information & Management, V.17, N.5 pp 227-237, 1981

R3 Radecki, Tadeusz. "FUZZY SET THEORICAL APPROACH TO DOCUMENT RETRIEVAL", Information & Management, V.15, pp 247-259, 1979

R4 Radecki, Tadeusz. "GENERALIZED BOOLEAN METHODS OF INFORMATION RETRIEVAL", Int. J. Man-Machine Studies, 1983, V.18, pp 407-439, 1983

R8 Radecki,Tadeusz. "A THEORICAL BACKGROUND FOR APPLYING FUZZY SET THEORY7 IN INFORMATION RETRIEVAL", Fuzzu sets and systems, V.2, pp 169-183, 1983

S1 Salton, Gerard, "EXTENDED BOOLEAN INFORMATION RETRIEVAL", Comunications of the ACM, V.26, N.12, pp 1022-1036, 1983

V1 Van Rijsbergen, C.J. y Sperk Jones, K. "A TEST FOR THE SEPARATION OF RELEVANT AND NON-RELEVANT DOCUMENTS IN EXPERIMENTAL RETRIEVAL COLLECTIONS", Journal of Documentation, V.29, pp 251-257, 1973

V2 Van Rijsbergen, C.J. "INFORMATION RETRIEVAL", Butterworths, 1979

W1 Waller, W.G. y Kraft, Donald H. "A MATHEMATICAL MODEL OF WEIGHTED BOOLEAN RETRIEVAL SYSTEM", Information & Management, V.15, pp 235-245, 1979

W2 Whitesitt, J.E. "ALGEBRA BOOLEANA Y SUS APLICACIONES", C.E.C.S.A, 1975

Z1 Zadeh, L. A. "THEORY OF FUZZY SETS", Memorandum No UCB/ERL M77-1, University of califirnia at Berkeley, Enero 4, 1977

Z2 Zadeh, L. A. "THE CONCEPT OF A LINGUISTIC VARIABLE AND ITS APPLICATION TO APPROXIMATE RESONING", Part I: Information Science, V.8, pp 301-357, 1977, Part III: Information Science, V.9, pp 43-80, 1976

Z3 Zadeh, L. A. "FUZZY SETS", Information and Control, V.8, pp 338-353, 1965