

Sistemas gerenciadores de banco de dados relacionais *Fuzzy*: uma aplicação em recuperação de informação

Sarajane Marques Peres* e Clodis Boscarioli

Universidade Estadual do Oeste do Paraná, Rua Universitária, 1619, Jardim Universitário, Campus Cascavel, Colegiado de Informática, 85814-110, Cascavel, Paraná, Brasil. *Autor para correspondência. e-mail: smperes@dca.fee.unicamp.br

RESUMO. Este trabalho apresenta uma aplicação de Sistemas de Banco de Dados Relacionais *Fuzzy* na Recuperação de Informação Textual. O sistema implementado permite observar as vantagens obtidas na recuperação de documentos textuais quando a imprecisão e a incompletude da informação é considerada no processo de busca e tomada de decisão. Os testes realizados mostram, de forma simplificada, a eficiência alcançada através do processamento e a persistência da informação em sua forma natural, possibilitadas pelo uso integrado da Teoria de Conjuntos *Fuzzy* e da Tecnologia de Banco de Dados.

Palavras-chave: recuperação de informação, Sistemas Gerenciadores de Banco de Dados *Fuzzy*, Conjuntos *Fuzzy*.

ABSTRACT. Management systems of Fuzzy relational database: an application to information retrieval. This paper presents an application to Information Retrieval using the Management Systems of Fuzzy Relational Database. The system implementation provides the observation of advantages on textual document retrieval, when it regards to the imprecision and incompleteness information about the research process and decision making. In a simple way, the accomplished tests show the efficiency obtained through the processing and persistence of information natural form allowed by the integration between Fuzzy Sets Theory and Database Technology.

Key words: information retrieval, fuzzy database management systems, fuzzy sets.

Introdução

As informações imprecisas, comumente encontradas na modelagem do mundo e geralmente descritas em linguagem natural, são passíveis de serem melhor tratadas através de ferramentas baseadas na Teoria de Conjuntos Fuzzy do que através de ferramentas convencionais. Além disso, em muitos casos, é interessante que essas informações sejam persistentes e gerenciadas de forma adequada. Assim, surge a necessidade de viabilizar a persistência das informações em sua forma natural (imprecisa, vaga e/ou incompleta) e essa pode ser implementada através do uso da tecnologia de Sistemas Gerenciadores de Banco de Dados (SGBD's), porém com algumas modificações.

Para clarificar essa idéia, este estudo pretende mostrar a utilização da Teoria de Conjuntos *Fuzzy* em SGBD's relacionais¹, bem como o quanto essa

teoria pode adequar os processos envolvidos no tratamento da informação, enquanto suporte à tomada de decisão em diferentes áreas de conhecimento, ilustrando esse processo com uma implementação simplificada.

Uma visão sucinta sobre arquitetura de SGBD's Relacionais *Fuzzy*, Álgebra Relacional *Fuzzy* e FSQL (*Fuzzy Structured Query Language*) é mostrada na segunda seção, para estimular a discussão e a análise de fatores envolvidos na problemática da modelagem e do processo de recuperação de informação. Para ilustrar a utilização dessa tecnologia, dois sistemas para recuperação de informação textual foram implementados, descritos na terceira seção, baseados no modelo de recuperação propostos por Silva e Ribeiro-Neto (1998). O primeiro sistema trata as informações envolvidas na recuperação de uma forma clássica e o segundo as trata de forma *fuzzy*. Por fim, apresentam-se as conclusões e as referências bibliográficas utilizadas.

Banco de dados relacional *Fuzzy*

A imprecisão, a incerteza e a incompletude intrínsecas às informações que representam o

¹ Para esclarecimentos sobre a Teoria de Conjuntos *Fuzzy* e SGBD's Relacionais recomenda-se Codd (1970, 1972); Date (1977); Elsmari e Navathe (2000); Klir e Yuan (1995); Korth *et al.* (1999); Pedrycz e Gomide (1998).

mundo real estão também presentes nas informações contidas em uma base de dados que se propõe a modelá-lo. Isso evidencia a necessidade de desenvolver métodos que lidem com essas características dentro da tecnologia de Banco de Dados. Assim, Zemankova e Kandel (1985); Sheno e Melton (1989); Li e Liu (1990); Medina (1994); Braga (1998), entre outros pesquisadores, sugerem a aplicação da teoria de conjuntos *fuzzy* e lógica *fuzzy* em ferramentas de SGBD's. Para tal, existem na literatura, abordagens diferentes que procuram satisfazer esse preceito, adaptando o projeto de banco de dados e/ou a implementação do SGBD ao tratamento de informações com o uso da teoria de conjuntos *fuzzy* e lógica *fuzzy*.

SGBD's Relacionais *Fuzzy* fornecem uma abordagem de consultas *fuzzy* baseada em relações *fuzzy*, que é aplicável a situações reais, principalmente pelo fato de que essas consultas não requerem que a solução satisfaça completamente as condições da mesma, mas requerem que a solução as satisfaça com um certo grau de satisfação (Li e Liu, 1990). Esse tipo de sistema determina soluções aproximadas e não soluções ótimas.

Um Sistema de Banco de Dados pode ser chamado de Relacional *Fuzzy* se ele suporta um Sistema de Banco de Dados Relacional incluindo domínio *fuzzy*, chaves e regras de integridades *fuzzy* e uma linguagem de consultas que seja tão poderosa quanto a Álgebra Relacional *Fuzzy*. Um sistema que suporta somente um Banco de Dados Relacional *Fuzzy* mas tem uma linguagem de consulta menos poderosa que a Álgebra Relacional *Fuzzy* deve ser chamado de relacional-semi-*fuzzy* (Li e Liu, 1990).

Para Li e Liu (1990), um modelo Relacional *Fuzzy* é necessário para embasar a construção de SGBD's Relacionais *Fuzzy*, consistindo em dois componentes principais: a estrutura de dados *fuzzy* relacional e a álgebra relacional. Para Medina (1994), um SGBD Relacional *Fuzzy* deve satisfazer as seguintes prerrogativas: proporcionar mecanismos adequados para poder representar a informação *fuzzy* em todas as suas vertentes; apresentar uma estrutura adequada para o armazenamento do significado da informação *fuzzy*; implementar um conjunto mínimo de operadores para recuperar e tratar a informação atendendo a natureza *fuzzy* da mesma; atender os requisitos do modelo relacional, sempre que possível.

Aspectos de projeto

Na teoria convencional de Banco de Dados Relacionais, os atributos de uma relação devem ser atômicos. Para a abordagem proposta, essa restrição

deve ser relaxada já que se está tratando de atributos *fuzzy*, os quais são símbolos tomados de um conjunto finito $\{A_1, A_2, \dots, A_n\}$, podendo assumir mais de um valor, sem, no entanto, mudar seu significado.

Cada atributo A_i está associado a um domínio *fuzzy* $D_{f(A_i)}$ que é um conjunto de valores possíveis para esse atributo, ou seja, os valores dos atributos *fuzzy* dos elementos de um domínio *fuzzy* que consiste de valores lingüísticos *fuzzy* ou números *fuzzy* representados por meio de um conjunto *fuzzy*. Tal sistema permite-nos representar, precisamente, parcialmente ou “possibilisticamente”, valores conhecidos ou desconhecidos em uma estrutura comum.

Formalmente, considere A um atributo e D um universo de discurso de A (o conjunto de todos os valores possíveis que podem ser atribuídos para A). O valor disponível na base de dados é composto pelo valor do atributo A para um objeto x que deverá ser representado por uma distribuição $\Pi_{A(x)}$ em D . A função de distribuição de possibilidade $\Pi_{A(x)}$ associada ao atributo A de um objeto x é um mapeamento:

$$\Pi_{A(x)} : D \rightarrow [0,1] \quad (1)$$

em que D é um universo de discurso de $A(x)$ e $\Pi_{A(x)}(d)$ representa a possibilidade de que $A(x)$ assume o valor d em D .

A Tabela 1 ilustra o conceito sobre relaxamento de atomicidade de atributo. O atributo “grau *fuzzy*” tem o domínio alto, médio, baixo, cuja variável base é o atributo “grau” e, dependendo das funções de mapeamento utilizadas, o atributo “grau *fuzzy*” pode assumir os valores alto, médio ou baixo, com valores de pertinência diferentes (representados pelo atributo “pertinência”).

Note que, nas duas formas de representação, ferem-se regras básicas de SGBD's Relacionais: na primeira, onde se respeita a restrição de atomicidade de atributo, viola-se a restrição de chave (identificação de tuplas) e na segunda, a restrição de chave é respeitada, contudo não existe atomicidade de atributo². Entretanto, em ambas as formas, está-se representando a mesma informação.

As implementações dos SGBD's Relacionais clássicos são baseadas na lógica bivalorada. Uma tupla t satisfaz ou não a relação R . Já a lógica *fuzzy* permite que a tupla t pertença a uma relação R com um grau de pertinência μ se R é satisfeita com o

² Este aspecto é tratado através das operações de álgebra relacional, já que, para implementação do SGBD Relacional e também para sua interpretação, não é interessante que esta informação esteja expressa dessa forma.

valor de verdade μ por essa tupla, em que $0 \leq \mu \leq 1$.

Tabela 1. Exemplo de relação *fuzzy*: atributos com domínio *fuzzy*

Conceitos/Documentos				
@id_conceito	@id_documento	grau	grau_fuzzy	Pertinência
Administração	d1	0.7	alto	0.9
Administração	d1	0.7	médio	0.3
Administração	d2	0.5	médio	1.0
Administração	d2	0.5	baixo	0.3
ou				
Conceitos/Documentos				
@id_conceito	@id_documento	grau	grau_fuzzy	Pertinência
Administração	d1	0.7	alto, médio	0.9/0.3
Administração	d2	0.5	médio, baixo	1.0/0.3

Formalmente Li e Liu (1990), uma Relação *Fuzzy* com dados clássicos é definida em D_1, D_2, \dots, D_n e caracterizada por uma função de pertinência de n -variáveis:

$$\mu_R: D_1 \times D_2 \times \dots \times D_n \rightarrow [0,1] \quad (2)$$

em que $D_1 \times D_2 \times \dots \times D_n$ é o produto cartesiano de D_1, D_2, \dots, D_n .

A Tabela 2 mostra uma relação *fuzzy* que contém atributos com domínios clássicos, porém um dos atributos (“grau”) expressa o grau de satisfação da tupla com a relação que, nesse caso, expressa o quanto um documento³ cobre um determinado conceito.

Tabela 2. Relação *fuzzy* com dados clássicos

Conceitos/Documentos			
@id_conceito	@id_documento	grau	...
Administração	d1	0.7	...
Informática	d2	0.5	...

Note que os valores dos atributos são clássicos. Já uma Relação *Fuzzy* com dados *fuzzy* é formalmente definida como segue (Li e Liu, 1990).

Considere D_1, D_2, \dots, D_n n conjuntos universo, $P(D_i)$ coleções de todas as distribuições de possibilidade em D_i , $i = 1, 2, \dots, n$. Portanto uma relação *fuzzy* R de um esquema relacional (A_1, A_2, \dots, A_n) é definida por uma função de pertinência de n -variáveis:

$$\mu_R: D_f(A_1) \times D_f(A_2) \times \dots \times D_f(A_n) \rightarrow [0,1] \quad (3)$$

em que $D_f(A_i) = P(D_i) \cup \{\text{NULL}\}$ é chamado de domínio *fuzzy*, e D_i é um conjunto base.

Uma tupla *fuzzy* t de relações *fuzzy* R é um elemento (d_1, d_2, \dots, d_n) de $D_f(A_1) \times D_f(A_2) \times \dots \times D_f(A_n)$ com um grau de pertinência μ obtido por (03), e $d_i \in D_f(A_i)$.

Uma relação *fuzzy* pode ser mostrada como uma tabela bidimensional, como no caso de relações bivaloradas, adicionando uma coluna especial para expressar a função de pertinência. Contudo uma relação *fuzzy* é uma tabela tridimensional, já que é definida por uma função de pertinência (bidimensional) de um produto cartesiano no intervalo $[0,1]$.

Uma relação *fuzzy* com dados *fuzzy* pode ser observada na Tabela 3, onde “relevância” é um atributo com domínio composto por termos lingüísticos.

Tabela 3. Relação *fuzzy* com dados *fuzzy*

Consultas/Documentos			
@id_consulta	@id_documento	relevância	...
q1	d1	Média	...
q2	d1	Grande	...

Uma relação de identidade sobre um domínio D é usada (implicitamente) em SGBD’s relacionais convencionais e induz a classes de equivalência que afetam o resultado de operações e removem tuplas redundantes. A relação identidade é substituída em SGBD’s Relacionais *Fuzzy* pela relação de similaridade (declarada explicitamente), da qual a relação identidade é um caso especial.

Formalmente, segundo Li e Liu (1990), uma relação de similaridade $s(x,y)$ para um dado domínio D_i é um mapeamento de todos os pares de elementos no domínio dentro de um universo de discurso $[0,1]$:

$$s: D_i \times D_i \rightarrow [0,1]. \quad (4)$$

Uma relação de similaridade pode ser observada no modelo proposto, tomando-se parte⁴ da relação Conceitos/Termos, como mostrado na Tabela 4. A Tabela 5 mostra uma relação de similaridade referente ao domínio de um atributo *fuzzy*.

Tabela 4. Relação de similaridade (para casos especiais)

Conceitos/Termos			
@conceito	@termo	grau	...
Administração	administração	1.0	...
Administração	globalização	0.8	...

Tabela 5. Relação de similaridade para atributo de domínio *fuzzy*

Relevância	Insignificante	Pequena	Boa	Grande
Insignificante	1.0	0.7	0.3	0.0
Pequena	0.7	1.0	0.7	0.3
Média	0.3	0.7	1.0	0.7
Grande	0.0	0.3	0.7	1.0

³ Documento, conceito e cobertura são conceitos discutidos na seção que descreve o modelo de recuperação de informação textual proposto neste trabalho.

⁴ A relação de similaridade se manifesta quando um conceito é tomado como termo primitivo. Para um maior entendimento desse fenômeno, é necessário remeter à seção deste trabalho onde o modelo de recuperação de informações é explicado.

Um SGBD Relacional Fuzzy, conforme Li e Liu (1990), consiste de um banco de dados real e um banco de dados virtual. O banco de dados real contém as relações-base armazenadas explicitamente em memória secundária. Relações virtuais (ou visões) representam informação não obrigatória para ser explicitamente estocada, mas possível de ser derivada de banco de dados reais; sobre aquelas, as linguagens de consulta *fuzzy* (como a FSQL, descrita nas seções seguintes) podem atuar para responder às consultas dos usuários. Na Tabela 1, pode ser observada uma relação virtual que seria composta pela chave da relação exemplificada e pelos atributos destacados.

Algumas considerações sobre esses conceitos devem ser ressaltadas (Li e Liu, 1990): uma relação-base é uma relação que tem existência independente e deve ser representada por arquivos distintos; as informações visíveis pelo usuário podem ser originadas em uma relação-base ou em uma relação virtual; uma relação virtual é uma relação que não tem existência própria, mas é derivada de uma relação-base; qualquer tupla pertencente a uma relação virtual deve ter um grau de pertinência.

O grau de pertinência de uma tupla numa relação *fuzzy*, freqüentemente, não pode ser dado pelo usuário de forma direta, fácil e confiável, pois os usuários estão acostumados a expressar dados *fuzzy* em termos de valores lingüísticos oriundos da linguagem natural. Esses valores lingüísticos, no entanto, podem ser representados por meio de conjuntos *fuzzy*, os quais devem ser projetados a partir de técnicas específicas para tal, como as técnicas sugeridas em Pedrycz e Gomide (1998).

O processo de derivação de uma relação virtual deve envolver projeções da relação-base sobre certos atributos e certas restrições e/ou junções de duas ou mais relações-base. Em geral, toda relação-base pode ser representada em termos de relações estendidas, desde que todas as tuplas na relação-base tenham um grau de pertinência igual a 1 (um), significando que as tuplas são totalmente compatíveis com a relação. Contudo, com a existência de restrições *fuzzy* nos blocos de consultas *fuzzy* que constroem as relações virtuais, valores diferentes nos atributos das diferentes tuplas farão com que essas satisfaçam as condições com graus diferentes e, na relação virtual, elas deverão ter graus diferentes.

Linguagens de Consulta Relacional Fuzzy

Álgebra Relacional Fuzzy

A Álgebra Relacional *Fuzzy*, base para SQL *Fuzzy*, é uma coleção de operações sobre relações *fuzzy*, sendo que a Álgebra Relacional convencional

é um caso especial desta. Cada operação da Álgebra Relacional *Fuzzy* toma uma ou duas relações *fuzzy* e produz, como resultado, novas relações *fuzzy*. As operações básicas estão descritas na seqüência.

A união de duas relações *fuzzy* compatíveis R e S com funções de pertinência $\mu_R(t)$ e $\mu_S(t)$ respectivamente, denotada por $R \cup S$, e o conjunto de todas as tuplas t que satisfazem a relação (Li e Liu, 1990):

$$\mu_{R \cup S}(t) = \max \{ \mu_R(t), \mu_S(t) \}. \quad (8)$$

A interseção de duas relações *fuzzy* compatíveis R e S com funções de pertinência $\mu_R(t)$ e $\mu_S(t)$ respectivamente, denotada por $R \cap S$, e o conjunto de todas as tuplas t que satisfazem a relação (Li e Liu, 1990):

$$\mu_{R \cap S}(t) = \min \{ \mu_R(t), \mu_S(t) \}. \quad (9)$$

A diferença de duas relações *fuzzy* compatíveis R e S com funções de pertinência $\mu_R(t)$ e $\mu_S(t)$ respectivamente, denotada por $R - S$, e o conjunto de todas as tuplas t que satisfazem a relação (Li e Liu, 1990):

$$\mu_{R - S}(t) = \min \{ \mu_R(t), (1 - \mu_S(t)) \}. \quad (10)$$

O produto cartesiano de n relações *fuzzy* R_1, R_2, \dots, R_n com funções de pertinência $\mu_{R_i}(t_i)$ ($i = 1, 2, \dots, n$) respectivamente, denotada por $R_1 \times R_2 \times \dots \times R_n$, é o conjunto de todas as tuplas t com grau de pertinência $\mu_{R_1 \times R_2 \times \dots \times R_n}(t)$ definido por (Li e Liu, 1990):

$$\mu_{R_1 \times R_2 \times \dots \times R_n}(t) = \min \{ \mu_{R_1}(t_1), \mu_{R_2}(t_2), \dots, \mu_{R_n}(t_n) \} \quad (11)$$

em que $t = (t_1, t_2, \dots, t_n)$.

A projeção de um conjunto de atributos A_1, A_2, \dots, A_n a partir de uma relação *fuzzy* $R = (D_i(A_i) \times D_i(A_2) \times \dots \times D_i(A_n))$ é o conjunto de todas as tuplas t formadas pelos referidos atributos e com grau de pertinência definido por (baseado em Braga, 1998):

$$\mu_R(d_{i_1}, d_{i_2}, \dots, d_{i_k}) = \max_{d_j \in D_i(A_j), j \in \{i_1, i_2, \dots, i_k\}} \{ \mu_R(d_{i_1}, d_{i_2}, \dots, d_{i_k}) \}. \quad (12)$$

A seleção de tuplas de uma relação *fuzzy* R a partir de um predicado σ envolvendo um conjunto de atributos A_1, A_2, \dots, A_n e um conjunto de elementos de domínio de cada atributo $D_i(A_1), D_i(A_2), \dots, D_i(A_n)$, e um conjunto de operadores *fuzzy* é o conjunto de todas as tuplas t que satisfazem ao referido predicado com algum grau de pertinência, o qual é determinado através do cálculo apropriado para os operadores *fuzzy* envolvidos (lógicos, aritméticos ou especiais), conforme Braga (1998).

Fuzzy Structured Query Language (FSQL)

Fuzzy Structured Query Language (FSQL) é uma extensão da SQL padrão, apresentada em (Li e Liu, 1990), através da qual dados fuzzy podem ser

manipulados apropriadamente e SQL. Essa linguagem tem a capacidade de modelar o tipo de consulta *fuzzy* que humanos empregam em ambiente *fuzzy* e retornam uma resposta *fuzzy* (ou não *fuzzy*) que satisfaz os requisitos intrínsecos à consulta.

FSQL preserva o estilo e a estrutura de SQL, porém, em FSQL, valores lingüísticos *fuzzy* podem ser definidos por meio de conjunto *fuzzy* e operadores *fuzzy* são definidos. Assim, com SQL, a função da FSQL é suportar definição, manipulação e controle de dados (*fuzzy* e não *fuzzy*) em SGBD's relacionais *fuzzy*.

A operação básica do FSQL é o mapeamento, o qual implica transformar os valores de um banco de dados em requisitos do usuário (Li e Liu, 1990). É sintaticamente representada pelo bloco de consulta *fuzzy*:

```
SELECT < lista de atributos >
FROM < lista de relações >
WHERE < condição >
WITH < grau >
```

O resultado é uma relação composta pelos valores de atributos selecionados de tuplas das relações especificadas, opcionalmente sujeitas a condições da cláusula WHERE com um dado grau.

SELECT e FROM definem a operação de projeção e WHERE é uma expressão lógica *fuzzy*. WHERE e WITH determinam quais tuplas das relações se qualificam para a operação de projeção, isto é, somente os valores dos atributos daquelas tuplas - para os quais a condição (expressão *fuzzy*) é verdade com um dado grau de satisfação - aparecerão no resultado. WHERE e WITH contêm especificações de seleção *fuzzy* e junção *fuzzy* e um caso especial, que representa o caso convencional, é quando o grau tem valor 1 (um).

A consulta (14) tem como resposta a Tabela 6. Observe que a primeira tupla da relação resposta foi recuperada em virtude da característica *fuzzy* da modelagem do SGBD relacional *fuzzy*. No caso clássico, se o atributo "grau_fuzzy" estivesse presente em sua forma clássica, a primeira tupla não existiria⁵ e somente a segunda seria recuperada.

```
SELECT *
FROM Conceitos/Documentos
WHERE grau_fuzzy = médio AND id_conceito =
"Administração"
WITH = 0.3
```

Uma consulta *fuzzy* pode retornar mais tuplas que uma consulta clássica, já que o modelo de valores lingüísticos representa a fundamental imprecisão da percepção humana e também geram informações adicionais na recuperação dos dados, porque a característica de distribuição de possibilidade de um único valor lingüístico se refere a uma grande quantidade de informação. Com FSQL, alguém pode refazer consultas em um banco de dados, tomando vantagem de valores lingüísticos, pois essa abordagem permite que os usuários se expressem de uma forma mais próxima à linguagem natural.

Tabela 6. Relação resultante da consulta 14

Conceitos/Documentos				
@id_conceito	@id_documento	grau	grau_fuzzy	Pertinência
Administração	d1	0.7	médio	0.3
Administração	d2	0.5	médio	1.0

O bloco de consultas *fuzzy*, como um todo, representa uma composição das operações de projeção, de seleção e de junção da álgebra relacional *fuzzy*. A consulta (15) trata-se de uma projeção com o diferencial de que tuplas redundantes serão eliminadas:

```
SELECT DISTINCT < lista de atributos >
FROM < relação >
```

O conceito de redundância é, nessa abordagem, algo *fuzzy* e pode ser tratado de diferentes formas. A maneira mais simples é tratá-lo de forma convencional, porém considerando que a coluna de pertinência da tupla seja tratada como um atributo da relação.

A operação de junção é executada através da agregação de atributos na cláusula SELECT de mais do que uma relação, dispostas na cláusula FROM, incluindo um critério na cláusula WHERE.

```
SELECT atribAa, atribAb, atribBb
FROM A, B
WHERE atribAa = atribBb
```

FSQL permite ainda a construção de operações de União, de Interseção e de Diferença, que combinam o resultado de blocos de consultas *fuzzy* independentes, exigindo, para tal, que eles tenham resultados compatíveis (mesmo número de atributos e atributos de domínio compatíveis). A especificação das ações de inserção, de remoção e de alteração de tuplas também é permitida.

Arquiteturas

Segundo Medina (1994), existem diferentes níveis *fuzzy* que podem ser cobertos por um SGBD Relacional *Fuzzy*: a obtenção da informação *fuzzy* a

⁵ Não existiria porque o atributo grau_fuzzy seria valorado com "alto", visto que o grau de pertinência a esse conjunto é maior (veja Tabela 1). A consulta 14 não retornaria a primeira tupla porque esta não satisfaria a cláusula WHERE.

partir de dados armazenados de forma precisa; a representação e a recuperação da informação *fuzzy*; a representação e a recuperação da informação *fuzzy* e tratamento da mesma (no sentido de inferir novas informações). Para Bosc e Pivert (1997a, 1997b), *apud* Braga (1998), existem três arquiteturas para a utilização da teoria *fuzzy* e da lógica *fuzzy* na obtenção de informação *fuzzy*.

Entretanto, apesar da definição dessas arquiteturas, percebe-se que os limites de cada uma delas é extensível, sobrepondo-se em alguns pontos, e a junção de aspectos de diferentes arquiteturas é incentivada na literatura para suportar uma modelagem adequada a diferentes problemas.

A primeira arquitetura é abordada por Harmon (1986), Bosc (1988) e Kacprzyk (1986) *apud* Medina (1994), na qual o mecanismo de consulta é implementado utilizando-se extensões de linguagens existentes, como a SQL, recuperando a informação contida na base de dados sob um aspecto *fuzzy*. As tuplas contidas na base devem conter um atributo que expresse o grau de pertinência dessas à relação e podem conter atributos cujos domínios envolvam valores expressos em termos lingüísticos.

A consulta é dividida em duas partes, em que a primeira é convencional e a segunda envolve aspectos *fuzzy* no intuito de ordenar as tuplas recuperadas seguindo o seu grau de pertinência à relação. Essas duas partes são conectadas por um operador conjuntivo (AND). Um exemplo ilustrado pela consulta (18) (considerando a Tabela 1) mostra como uma consulta é processada nessa arquitetura.

```
SELECT id_documento
FROM Conceitos/Documentos
WHERE grau_fuzzy = alto AND grau >= 0.7
```

 (18)

O processamento da consulta (18) é executado de forma clássica e os resultados devem ser ordenados (também de forma clássica como a execução de um ORDER BY) pelo atributo “grau”. Essa ordenação constitui o aspecto *fuzzy* desta arquitetura já que o referido atributo expressa o quão a tupla é pertinente à relação.

A segunda arquitetura faz uso das relações de similaridade para tratar os atributos *fuzzy* contidos nas relações. Para cada atributo *fuzzy*, uma relação de similaridade pode ser criada, proporcionando uma maior flexibilidade na recuperação da informação.

Essa arquitetura é abordada por Vague, Ichikawa e Hirakawa (1986) e por Ares e Motro (1988) *apud* Braga (1998) e Zemankova e Kandell (1985), e faz o processamento de informações *fuzzy* com a utilização de relações de similaridade. O exemplo (considerando a Tabela 3 e a Tabela 5) ilustrado pela

consulta (19) mostra como uma consulta é processada nessa arquitetura.

```
SELECT id_documento
FROM Consultas/Documentos
WHERE id_consulta = q1 AND relevância = grande
WITH grau_similaridade = 0.6
```

 (19)

O processamento da consulta (19) é realizado de forma clássica para o atributo clássico “id_consulta” na cláusula WHERE e de forma *fuzzy* para o atributo *fuzzy* “relevância”, considerando a relação de similaridade a ele relacionada. Como resultado dessa consulta, obter-se-á a primeira tupla da relação Consultas/Documentos, já que o valor *médio* do atributo “relevância” é considerado *grande* com o grau de similaridade 0.7, segundo a relação de similaridade. Como a cláusula WITH determina o limite de 0.6 para o grau de relevância, a tupla satisfaz a consulta.

A terceira arquitetura, abordada por Li e Liu (1990) e Medina (1994), procura inserir aspectos *fuzzy* na representação e no processamento das informações. Essa arquitetura faz uso de relações virtuais e implementa, no processador de consultas, algoritmos especializados que utilizam a teoria *fuzzy* e a lógica *fuzzy* para tratar as operações das linguagens de consultas.

Esse processamento envolve diretamente consultas precisas e imprecisas, utilizando conjunções e disjunções entre ambos os tipos de consultas. Os atributos dos predicados envolvem variáveis lingüísticas, termos nebulosos, modificadores e quantificadores.

Para a implementação dessa arquitetura, é necessário definir uma gramática completa para o FSQL (baseada na Álgebra Relacional *Fuzzy* e no Cálculo Relacional *Fuzzy*), assim como construir um modelo de mapeamento com as funções que determinam os conjuntos *fuzzy* para o domínio *fuzzy* dos atributos. Além disso, a operações aritméticas e agregadas devem ser implementadas segundo a teoria de conjuntos *fuzzy*. A partir dessa implementação, todo o processador de consulta do SGBD Relacional *Fuzzy* trata as informações sob os preceitos da Teoria de Conjuntos *Fuzzy* e Lógica *Fuzzy*, fazendo das informações clássicas um caso especial de informações *fuzzy*.

Modelo para recuperação de informações textuais

Segundo Yates e Ribeiro-Neto (1999), recuperação de informação trabalha com representação, armazenamento, organização e acesso à informação. A representação e a organização dos itens de informação deveriam fornecer ao usuário

um acesso fácil àquilo que realmente o interessa. Entretanto, expressar a necessidade do usuário não é uma tarefa trivial.

Ainda segundo os mesmos autores, recuperação de dados consiste em determinar quais objetos de uma coleção (ex. Documentos) contêm as palavras chave da consulta do usuário, enquanto que recuperação de informação objetiva recuperar informações sobre um assunto, satisfazendo a uma consulta.

A recuperação de informação geralmente trabalha com textos em linguagem natural que não se encontram bem estruturados e possuem semântica ambígua; então, para satisfazer a uma consulta do usuário, o sistema precisa saber interpretá-la sintática e semanticamente, de forma a ser capaz de atribuir uma relevância a um objeto recuperado. Yates e Ribeiro-Neto (1999) afirmam que a noção de relevância é o centro da recuperação de informação.

O sistema⁶ proposto usa um procedimento de análise sintática baseado em satisfações percentuais de cobertura de um documento a uma consulta, e a análise semântica é trabalhada pelo modelo *fuzzy*.

O sistema é composto pela junção de diferentes aspectos das arquiteturas brevemente discutidas e pretende ilustrar a potencialidade da tecnologia de SGBD's Relacionais *Fuzzy* e implementa um modelo melhorado de recuperação de informação textual, baseado em Silva e Ribeiro-Neto (1998)⁷ e Macali (1998)⁸.

Entretanto, não é objetivo deste trabalho a implementação completa desse sistema, já que não se possui, no momento, uma plataforma que suporte tal implementação. As seções que seguem descrevem a proposta bem como os resultados esperados, além de mostrar os resultados obtidos da implementação parcial.

Objetivo do sistema

O objetivo principal desse sistema é recuperar documentos textuais, a partir de uma base de dados textuais, que sejam do interesse do usuário. O interesse do usuário é expresso através de consultas formuladas ao sistema e através da especificação de alguns limites de aceitação, ou seja, da especificação

de um valor para o grau mínimo de satisfação requerida.

A consulta a ser submetida ao sistema deve ser composta por palavras (termos primitivos) conectadas pelos operadores lógicos AND e/ou OR. O conectivo AND expressa a exigência da pertinência das palavras a ele relacionadas, no documento recuperado. O conectivo OR expressa o desejo de que as palavras a ele conectadas apareçam no texto, mas não expressa qualquer tipo de exigência.

Os limites de aceitação que o usuário deve parametrizar são referentes ao nível de rigorosidade que deve ser empregado na recuperação de um documento. Essa rigorosidade está diretamente ligada aos seguintes itens: o quanto o documento deve estar inserido dentro de um determinado contexto (conceito⁹); o quanto a consulta deve estar inserida num determinado contexto (conceito); o quanto o documento deve satisfazer à consulta (contendo ou não os termos da consulta); o quanto os fatores acima estão relacionados (resultando na relevância do documento à consulta sob um conceito específico).

Descrição do modelo

Considerando um espaço amostral claramente definido, um conjunto de conceitos, documentos e consultas são indexados por conceitos primitivos (termos). Cada termo está relacionado aos conceitos de acordo com um grau de pertinência, o qual expressa o relacionamento entre o termo e o conceito.

A Tabela 7 ilustra o relacionamento entre os termos primitivos t_i e os conceitos C_i , onde quanto maior a ligação entre o termo e o conceito, maior será o grau de pertinência deste termo ao conceito especificado. Assim, de acordo com os dados fornecidos na tabela, pode-se observar, por exemplo, que o termo t_3 possui melhor identificação com o conceito C_1 do que com o conceito C_2 .

As indexações de documentos e de consultas, d_i e q_i respectivamente, são representadas pela existência ou não existência de um termo primitivo no documento ou nas consultas em questão. As Tabelas 8 e 8 representam esses fatos.

Um documento e/ou uma consulta são associados a um conceito. De acordo com o grau de cobertura dos mesmos a esses conceitos e, a partir disso, buscar-se-á identificar os documentos que cobrem os mesmos conceitos que as consultas cobrem.

⁶ Este sistema foi implementado sobre um SGBD relacional clássico, incluindo em seu modelo entidade-relacionamento, atributos que modelam características *fuzzy* e proporcionam a execução das idéias aqui descritas. Para uma implementação totalmente *fuzzy* (terceira arquitetura descrita em 2.3), seria necessária a implementação de algoritmos especializados para compor o *kernel* de um SGBD relacional *fuzzy*.

⁷ Modelo de recuperação de informação textual baseado em Redes Bayesianas.

⁸ Modelo de recuperação de informação baseado em Silva e Ribeiro-Neto (1998), com introdução de recursos de teoria de conjuntos *fuzzy* na indexação dos conceitos.

⁹ Para esta aplicação, um conceito se refere a uma área do conhecimento.

Tabela 7. Relacionamento entre termos e conceitos

Conceitos/Termos	t1	t2	t3	...	tn
C1	0,2	0,7	0,9	...	0,1
C2	0	0,2	0,1	...	0,8
C3	1	0,8	0,3	...	0,5
...
Cm	0,1	0,5	0,7	...	0,3

Tabela 8. Relacionamento entre termos e documentos

Termos/Documentos	d1	d2	...	dn
t1	1	1	...	0
t2	0	1	...	0
t3	1	0	...	0
...
Tm	0	1	...	1

Tabela 9. Relacionamento entre termos e consultas

Termos/Consultas	c1	c2	...	cn
t1	1	0	...	0
t2	0	1	...	0
t3	0	1	...	0
...
Tm	0	0	...	1

A cobertura de um documento a um conceito é obtida através da soma normalizada dos graus de pertinência de cada termo existente no documento, em relação a esse conceito. Assim, a cobertura do documento d_j ao conceito C_i é dada por:

$$\text{Cob}(d_j, C_i) = \sum_j \mu(t_j, C_i) / \text{Card}(t), \text{ em que } j = 1, 2, \dots, m \text{ e } t_j \in d_j. \quad (20)$$

A cobertura de uma consulta a um conceito é obtida através da aplicação dos operadores Max e Min, executando as operações de união e de interseção respectivamente, obedecendo à semântica especificada na consulta, onde o conectivo lógico OU representa união e o conectivo lógico E representa a interseção.

Se a consulta envolve apenas um termo, a cobertura dessa ao conceito é igual ao grau de relacionamento desse termo a esse conceito. Se a consulta envolve dois ou mais termos, tem-se:

$$\text{Cob}(q_j, C_i, \text{OU}) = \max \{ \mu(t_j, C_i) \}, \text{ em que } j = 1, 2, \dots, m \text{ e } t_j \in q_j. \quad (21)$$

ou

$$\text{Cob}(q_j, C_i, \text{E}) = \min \{ \mu(t_j, C_i) \}, \text{ em que } j = 1, 2, \dots, m \text{ e } t_j \in q_j. \quad (22)$$

A cobertura do documento à consulta é obtida da constatação de que o documento possui, como índices, os termos que indexam as consultas. Quanto maior for o número de termos em comum, no caso de consultas envolvendo o conectivo lógico OU, maior é a área de cobertura do documento sobre a consulta. Se a consulta envolve apenas um termo ou o conectivo lógico E, a cobertura do documento à consulta exige a existência de todos os termos da consulta como índices de documentos.

Entretanto, um mesmo conceito pode assumir diversas ramificações, e o fato de um documento cobrir o mesmo conceito que uma consulta não significa que ele deva ser recuperado, já que existe a

possibilidade de o documento e a consulta cobrirem ramificações diferentes do mesmo conceito.

Por exemplo, se a consulta possui o termo “Álgebra Linear”, estará cobrindo o conceito “Matemática”; entretanto, documentos que cobrem este conceito mas que tratam do assunto “Trigonometria” não são interessantes para a recuperação. Assim, é necessário definir, também, a cobertura dos documentos candidatos à recuperação em relação à consulta.

A relevância do documento para uma consulta é um atributo que permite estabelecer um *ranking* de recuperação e pode ser mapeado para valores lingüísticos a partir de uma função de mapeamento sobre o grau de pertinência do documento em relação àquela consulta. Esse grau de pertinência é utilizado para instanciar o atributo “relevância” e estabelece o grau do relacionamento de um documento em relação a uma consulta, sendo obtido por uma média ponderada dos graus de cobertura da consulta e do documento ao conceito e do grau de cobertura do documento à consulta.

Para modelagem do sistema clássico, foram realizadas duas modificações básicas: o relacionamento dos termos aos conceitos foram mapeados para 0 e 1, sendo que para os graus de relacionamento maiores do que 0,5 estabeleceu-se grau 1 e para os demais, grau 0; a relevância do documento se restringiu à cobertura que esse documento apresentava para a consulta relacionada.

Testes e análises

Os testes foram realizados sobre uma base de dados instanciada empiricamente, contendo a seguinte configuração: 10 documentos textuais; 4 conceitos; 24 termos. O número de consultas executadas totalizou 31 execuções, resultando em 77 recuperações.

As consultas foram divididas em três blocos: consultas envolvendo como termos os conceitos e as combinações entre eles; consultas envolvendo termos primitivos e combinações entre eles; consultas envolvendo combinações entre termos e conceitos.

Cada bloco de consultas foi executado no sistema clássico e no sistema *fuzzy*, sendo que, neste último, duas configurações de parâmetros foram executadas (com grau 0.0 para todos os itens parametrizáveis e com limiar 0.5 para os três primeiros itens - cobertura do documento ao conceito, cobertura da consulta ao conceito e cobertura do documento à consulta - e 0.0 para o último - relevância do documento à consulta).

O resultado que os sistemas apresentam estão na forma de relevância que os documentos recuperados

apresentam para a consulta relacionada, sob a ótica de um determinado conceito. Os testes ficaram divididos como mostra a Tabela 10.

Os resultados obtidos nos três primeiros testes mostraram que o sistema *fuzzy* tem uma capacidade maior de generalização, recuperando uma variedade maior de documentos do que o sistema clássico. Contudo, as relevâncias atribuídas para cada documento recuperado foram mais restritas no sistema *fuzzy*, o que leva a acreditar que o modelo *fuzzy* realiza uma análise mais criteriosa sobre a relação que um documento tem com uma consulta.

Tabela 10. Testes realizados

Número do teste	Sistema	Parâmetros	Termos da consulta
Teste 1	Clássico	----	Conceitos
Teste 2	<i>Fuzzy</i>	0.0 0.0 0.0 0.0	Conceitos
Teste 3	<i>Fuzzy</i>	0.5 0.5 0.5 0.0	Conceitos
Teste 4	Clássico	----	Termos
Teste 5	<i>Fuzzy</i>	0.0 0.0 0.0 0.0	Termos
Teste 6	<i>Fuzzy</i>	0.5 0.5 0.5 0.0	Termos
Teste 7	Clássico	----	Termos e Conceitos
Teste 8	<i>Fuzzy</i>	0.0 0.0 0.0 0.0	Termos e Conceitos
Teste 9	<i>Fuzzy</i>	0.5 0.5 0.5 0.0	Termos e Conceitos

Tabela 9. Resultados parciais (Consulta: administração AND contabilidade OR informática)

Teste 1		
Documento	Conceito	Relevância
Relações Comerciais e Globalização	administração	0.75.
Relações Comerciais e Globalização	contabilidade	0.75.
Relações Comerciais e Globalização	engenharia agrícola	0.453125.
Relações Comerciais e Globalização	informática	0.71875.
Estudo do Imobilizado Tangível Depreciável	administração	0.75.
Estudo do Imobilizado Tangível Depreciável	contabilidade	0.75.
Estudo do Imobilizado Tangível Depreciável	engenharia agrícola	0.45833334
Estudo do Imobilizado Tangível Depreciável	informática	0.7291666
Projeto de Irrigação por Gotejamento para Culturas de Pepino	administração	0.6041666
Projeto de Irrigação por Gotejamento para Culturas de Pepino	contabilidade	0.5833334
Projeto de Irrigação por Gotejamento para Culturas de Pepino	engenharia agrícola	0.375
Projeto de Irrigação por Gotejamento para Culturas de Pepino	informática	0.5833334
Análise de Programas de Inf. aplicados ao estudo de funções	administração	0.625.
Análise de Programas de Inf. aplicados ao estudo de funções	contabilidade	0.625.
Análise de Programas de Inf. aplicados ao estudo de funções	engenharia agrícola	0.375
Análise de Programas de Inf. aplicados ao estudo de funções	informática	0.625
Educação a Distância como alternativa ao Processo Educacional	administração	0.59375
Educação a Distância como alternativa ao Processo Educacional	contabilidade	0.5625
Educação a Distância como alternativa ao Processo Educacional	engenharia agrícola	0.3125.
Educação a Distância como alternativa ao Processo Educacional	informática	0.625
Teste 2		
Documento	Conceito	Relevância
Relações Comerciais e Globalização	administração	0.521875
Relações Comerciais e Globalização	contabilidade	0.5078125
Relações Comerciais e Globalização	engenharia agrícola	0.3140625
Relações Comerciais e Globalização	informática	0.6859375
Estudo do Imobilizado Tangível Depreciável	administração	0.5291667
Estudo do Imobilizado Tangível Depreciável	contabilidade	0.5333333
Estudo do Imobilizado Tangível Depreciável	engenharia agrícola	0.3125
Estudo do Imobilizado Tangível Depreciável	informática	0.6641666
Projeto de Irrigação por Gotejamento para Culturas de Pepino	administração	0.352082333
Projeto de Irrigação por Gotejamento para Culturas de Pepino	contabilidade	0.3458333
Projeto de Irrigação por Gotejamento para Culturas de Pepino	engenharia agrícola	0.22708333
Projeto de Irrigação por Gotejamento para Culturas de Pepino	informática	0.5604167
Análise de Programas de Inf. aplicados ao estudo de funções	administração	0.36875
Análise de Programas de Inf. aplicados ao estudo de funções	contabilidade	0.3625
Análise de Programas de Inf. aplicados ao estudo de funções	engenharia agrícola	0.225
Análise de Programas de Inf. aplicados ao estudo de funções	informática	0.603125
Sistemas de Informação inteligentes para conservação de solo	administração	0.375
Sistemas de Informação inteligentes para conservação de solo	contabilidade	0.3125.
Sistemas de Informação inteligentes para conservação de solo	engenharia agrícola	0.23750001.
Sistemas de Informação inteligentes para conservação de solo	informática	0.55625.
Educação a Distância como alternativa ao Processo Educacional	administração	0.359375
Educação a Distância como alternativa ao Processo Educacional	contabilidade	0.33125
Educação a Distância como alternativa ao Processo Educacional	engenharia agrícola	0.18125
Educação a Distância como alternativa ao Processo Educacional	informática	0.575
Teste 3		
Documento	Conceito	Relevância
Relações Comerciais e Globalização	administração	0.521875
Relações Comerciais e Globalização	contabilidade	0.5078125
Estudo do Imobilizado Tangível Depreciável	administração	0.5291667.
Estudo do Imobilizado Tangível Depreciável	contabilidade	0.5333333.

Tabela 10. Resultados parciais (Consulta: programa AND processo OR processamento)

Teste 4			
Documento	Conceito	Relevância	
Relações Comerciais e Globalização	administração	0.625	
Relações Comerciais e Globalização	contabilidade	0.625	
Relações Comerciais e Globalização	engenharia agrícola	0.578125	
Relações Comerciais e Globalização	informática	0.59375	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	administração	0.7291666	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	contabilidade	0.7083334	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	engenharia agrícola	0.75	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	informática	0.7083334	
Data Warehouse: uma Proposta de Aplicação	administração	0.625	
Data Warehouse: uma Proposta de Aplicação	contabilidade	0.6	
Data Warehouse: uma Proposta de Aplicação	engenharia agrícola	0.6	
Data Warehouse: uma Proposta de Aplicação	informática	0.625	
Teste 5			
Documento	Conceito	Relevância	
Relações Comerciais e Globalização	administração	0.396875	
Relações Comerciais e Globalização	contabilidade	0.3828125	
Relações Comerciais e Globalização	engenharia agrícola	0.3390625	
Relações Comerciais e Globalização	informática	0.5234375	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	administração	0.47708333	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	contabilidade	0.4708333	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	engenharia agrícola	0.5020833	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	informática	0.6479167	
Data Warehouse: uma Proposta de Aplicação	administração	0.41	
Data Warehouse: uma Proposta de Aplicação	contabilidade	0.3675	
Data Warehouse: uma Proposta de Aplicação	engenharia agrícola	0.3425	
Data Warehouse: uma Proposta de Aplicação	informática	0.54249996	
Teste 6			
Documento	Conceito	Relevância	
Projeto de Irrigação por Gotejamento para Culturas de Pepino	engenharia agrícola	0.5020833	

Tabela 11. Resultados parciais (Consulta: análise OR custo AND administração)

Teste 7			
Documento	Conceito	Relevância	
Estudo do Imobilizado Tangível Depreciable	administração	0.75	
Estudo do Imobilizado Tangível Depreciable	contabilidade	0.75	
Estudo do Imobilizado Tangível Depreciable	engenharia agrícola	0.583334	
Estudo do Imobilizado Tangível Depreciable	informática	0.7291666	
Data Warehouse: uma Proposta de Aplicação	administração	0.75	
Data Warehouse: uma Proposta de Aplicação	contabilidade	0.725	
Data Warehouse: uma Proposta de Aplicação	engenharia agrícola	0.6	
Data Warehouse: uma Proposta de Aplicação	informática	0.75	
Teste 8			
Documento	Conceito	Relevância	
Estudo do Imobilizado Tangível Depreciable	administração	0.6416665	
Estudo do Imobilizado Tangível Depreciable	contabilidade	0.5333333	
Estudo do Imobilizado Tangível Depreciable	engenharia agrícola	0.275	
Estudo do Imobilizado Tangível Depreciable	informática	0.39791667	
Data Warehouse: uma Proposta de Aplicação	administração	0.64750004	
Data Warehouse: uma Proposta de Aplicação	contabilidade	0.4925	
Data Warehouse: uma Proposta de Aplicação	engenharia agrícola	0.28	
Data Warehouse: uma Proposta de Aplicação	informática	0.4425	
Teste 9			
Documento	Conceito	Relevância	
Estudo do Imobilizado Tangível Depreciable	administração	0.6416665	
Estudo do Imobilizado Tangível Depreciable	contabilidade	0.5333333	
Data Warehouse: uma Proposta de Aplicação	administração	0.64750004	

Note ainda que a maior relevância de um documento para uma determinada consulta em relação aos conceitos mudou do sistema clássico para o sistema *fuzzy* e também mudou em relação às duas execuções no sistema *fuzzy*. Parte dos resultados pode ser observada na Tabela 11.

No teste 3, em particular, percebe-se um resultado mais refinado, mostrando que, com o estabelecimento de limiares de satisfação, os documentos recuperados sofrem um processo de filtragem mais rigoroso, resultando em um conjunto de documentos que parecem mais adequados à consulta realizada, dentro de conceitos específicos.

Nos testes 4, 5 e 6, observou-se que o número de documentos recuperados, em geral, não variou do sistema clássico para o sistema *fuzzy*; entretanto, a relevância no sistema *fuzzy* foi bem mais criteriosa, como mostra a Tabela 10.

Essa característica de maior rigorosidade se deve ao fato de que, na análise dos testes 1, 2 e 3, os termos tinham um grau de relacionamento igual a 1 com os conceitos. Nos testes 4, 5 e 6, o grau de relacionamento entre os termos e os conceitos é mais fraco (veja Tabela 4). Observe que, para o teste 6, os resultados são bem mais refinados.

Para os testes 7, 8 e 9, como pode ser observado na Tabela 11, os resultados obtidos têm um comportamento semelhante aos anteriores.

Conclusão

Com o desenvolvimento deste estudo, pôde-se abrir um espaço para discussão da viabilidade da aplicação da Teoria de Conjuntos *Fuzzy* e Lógica *Fuzzy* no contexto do armazenamento persistente e do processamento de informações *fuzzy*. A implementação apresentada pretendeu ilustrar a tecnologia proposta, envolvendo aspectos das diferentes arquiteturas existentes para SGBD's Relacionais *Fuzzy*.

Com o modelo de recuperação de informação proposto e com as extensões previstas no Modelo Entidade Relacionamento para o referido sistema, foi possível vislumbrar as oportunidades que essa tecnologia oferece. No entanto, para validar o modelo de forma robusta, seria importante testá-lo em bases textuais maiores, com maior diversidade de termos, conceitos e documentos, e realizar um estudo estatístico para melhor fundamentar as análises.

Ainda assim, os resultados obtidos mostraram-se estimuladores ao aprofundamento e ao aperfeiçoamento do modelo proposto e, principalmente, da tecnologia de SGBD's Relacionais *Fuzzy*.

Referências

- BRAGA, A.L. *Ferramentas de manipulação nebulosa de dados com aplicação em sistemas e Informação Geográfica*. 1998. Dissertação (Mestrado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1998.
- CODD, E.F. Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*, New York, v. 13, n.6, p.377-387, 1970.
- CODD, E.F. Relational completeness of data base sublanguages. *Data Base Systems, Courant Computer Science Symposia Series*, New York, n.6, p.65-98, 1972.
- DATE C. J. *An introduction to database systems*. 2.ed. Reading, Addison-Wesley, 1977.
- ELSMARI, R.A.; NAVATHE, S.B. *Fundamentals of Database Systems*. 2.ed. Redwood City: The Benjamin/Cummings Publishing Company, INC., 2000.
- LI, D.; LIU, D. *A Fuzzy PROLOG Database System*. New York: John Wiley & Sons INC., 1990.
- KLIR, G.J.; YUAN, B. *Fuzzy Sets and Fuzzy Logic*. New Jersey: Prentice-Hall, 1995.
- KORTH, H. et al. *Sistemas de banco de dados*. 3. ed. São Paulo: Makron Books, 1999.
- MACALI, L. *Recuperação de informação com utilização de lógica difusa*. 1998. Monografia (Graduação) – Universidade Estadual do Oeste do Paraná, Cascavel, 1998.
- MEDINA, J.M. *Bases de datos relacionales difusas. Modelo teórico y aspectos de su implementación*. 1994. Tese (Doutorado) - Universidade de Granada, Espanha, 1994.
- PEDRYCZ, W.; GOMIDE, F. *An introduction to fuzzy sets: analysis and design*. Cambridge: MIT Press, 1998.
- SHENOI, S.; MELTON, A. Proximity relations in the fuzzy relational database model. *Fuzzy Sets and Systems*, Amsterdam, v.31, n. 3, p.285-296, 1989.
- SILVA, I.R.; RIBEIRO-NETO, B. Avaliação de desempenho de um modelo bayesiano para recuperação de informação em bibliotecas digitais. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 13., 1999, Maringá. *Anais...* Maringá: UEM, 1998. p.285-299.
- YATES, R.B.; RIBEIRO-NETO, B. *Modern information retrieval*. Addison Wesley, 1999.
- ZEMANKOVA, M.; KANDEL, A. Implementing imprecision in information systems. *Information Sciences*, New York, v.37, n.1-3, p.107-141, 1985.

Received on April 17, 2002.

Accepted on July 29, 2002.