

PS-660

DATA RETRIEVAL USING FUZZY LOGIC

Marcos Tarquinio Simao de Mello (Centro Universitario da Cidade Cidade do Rio de Janeiro, Escola de Ciências Exatas e Tecnologia, NUPAC, Brasil) - mtsmello@gmail.com
Gustavo Gomes (Centro Universitario da Cidade Cidade do Rio de Janeiro, Escola de Ciências Exatas e Tecnologia, NUPAC, Brasil) - gustavo.gomesc@gmail.com

Colaborador:

Ronaldo Ribeiro Goldschmidt (Centro Universitario da Cidade Cidade do Rio de Janeiro, Escola de Ciências Exatas e Tecnologia, NUPAC, Brasil) - ronaldo_goldschmidt@yahoo.com.br

It is common in several everyday situations the approximate way of the human reasoning to use uncertainty and imprecision. We many times are not available all the necessary information to the assumption of decision. The computation offers us powerful resources for information recovery starting from databases. The language SQL is itself a technology example that allows the formulation of complex queries for relational databases. In general the execution of such queries presupposes the specification needs much defined parameters. And the answers presented should attend rigorously to the specified parameters. However, not always we wish to obtain exact values in our queries. Many times approximate results are enough to attend to the demand. The language SQL does not have resources for obtainment of approximate answers. On the other hand, the Fuzzy Logic is a paradigm of the Computational Intelligence offer ways to work with uncertainty and imprecision. Ahead of this scenario, it shows very oportune the availability of a computational tool based on Fuzzy Logic, that assists people in the query process involving flexibility in the information recovery. That article presents three methods that make possible the information recovery: query from fuzzy predicates; query with example values; and query with the resources of the two previous methods integrated in a single method. All the methods are based on resources offered by the Fuzzy Logic.

Keywords: fuzzy logic, retrieval database, imprecision

MÉTODOS DE RECUPERAÇÃO DE INFORMAÇÃO UTILIZANDO LÓGICA NEBULOSA

É comum em diversas situações cotidianas o modo aproximado de o raciocínio humano utilizar incerteza e imprecisão. Muitas vezes não se encontram disponíveis todas as informações necessárias à tomada de decisão. A computação nos oferece recursos poderosos para recuperação de informação a partir de bancos de dados. A própria linguagem SQL é um exemplo de tecnologia que permite a formulação de consultas complexas a bancos de dados relacionais. Em geral a execução de tais consultas pressupõe a especificação precisa de parâmetros bem definidos. E as respostas apresentadas devem atender rigorosamente aos parâmetros especificados. No entanto, nem sempre desejamos obter valores exatos em nossas consultas. Muitas vezes resultados aproximados são suficientes para atender à demanda. A linguagem SQL não dispõe de recursos para obtenção de respostas aproximadas. Por outro lado, a Lógica Nebulosa é um paradigma da Inteligência Computacional fornece meios para lidar com incerteza e imprecisão. Diante deste cenário, mostra-se bastante oportuna a disponibilidade de uma ferramenta computacional baseada em Lógica Nebulosa, que auxilie as pessoas no processo de consulta envolvendo flexibilidade na recuperação da informação. Esse artigo apresenta três métodos que viabilizam a recuperação de informação: consulta a partir de predicados nebulosos; consulta com valores de exemplo; e consulta com os recursos dos dois métodos anteriores integrados em um único método. Todos os métodos são baseados nos recursos oferecidos pela Lógica Nebulosa.

1 Introdução

Como o volume de dados armazenados nas diversas bases de dados corporativas está passando por um crescimento exponencial, torna-se cada vez mais indispensável poder extrair informações que contenham algum tipo de conhecimento para essas organizações. Muitas vezes para se obter informações a partir de uma base de dados é necessário lidar com algumas informações imprecisas. Por exemplo, podemos querer identificar bons pagadores em uma base de dados de uma financeira. O conceito de cliente bom pagador é impreciso e subjetivo. Um especialista poderia definir tal conceito como, por exemplo, alguém que recebesse um salário alto e tivesse poucas despesas que são também conceitos subjetivos.

Os métodos tradicionais para recuperação de informações em repositórios de dados usam bancos relacionais geralmente incapazes de lidar com informações imprecisas e subjetivas. A Lógica Nebulosa, paradigma da Inteligência Computacional voltado à modelagem do modo aproximado do raciocínio humano e a sua incorporação em sistemas de apoio à decisão, mostra-se como uma alternativa com potencial para tratar estas situações. Diante deste cenário, o presente artigo tem como objetivo apresentar um ambiente baseado em Lógica Nebulosa para recuperação nebulosa de informações (RNI). Este ambiente está inserido em uma ferramenta de Mineração de Dados denominada JavaBramining (JB), desenvolvida no Núcleo de Projetos e Pesquisa em Aplicações Computacionais do Centro Universitário da Cidade do Rio de Janeiro, e possui dois métodos:

- Um para formulação e execução de consultas nebulosas, onde podem ser definidos termos lingüísticos para um ou mais atributos de um banco de dados.
- O outro para recuperação de dados a partir de exemplos, onde podem ser apresentados um ou mais casos como situações de referência. O método deverá recuperar situações similares aos exemplos apresentados utilizando recursos da Lógica Nebulosa.

O presente artigo encontra-se estruturado em mais seis seções. Na seção dois encontram-se apresentados conceitos básicos sobre Lógica Nebulosa para a melhor compreensão dos métodos propostos. Trabalhos relacionados são comentados na seção três. A seção quatro descreve os métodos propostos, enquanto a seção cinco relata os experimentos realizados e os resultados obtidos. Considerações finais e perspectivas de trabalhos futuros estão na seção seis.

2 Lógica Nebulosa

A lógica nebulosa, também conhecida como lógica difusa ou lógica fuzzy (do inglês fuzzy logic), é uma teoria matemática que permite modelar o modo aproximado do raciocínio, imitando assim a habilidade humana de tomada de decisão em ambientes de incerteza e imprecisão [1].

No mundo real, usamos palavras como *muito* e *pouco*, *grande* e *pequeno*, *frequentemente* e *raramente* entre outras palavras para descrever e qualificar algumas situações decorrentes do nosso dia-a-dia. Estas situações não são nitidamente definidas e não podem ser precisamente descritas, já que os seres humanos associam diferentes descrições para essas palavras. Segundo [2], utilizando a lógica nebulosa é possível a manipulação de parâmetros numéricos e de informações de linguagem.

O conceito mais importante da lógica nebulosa é a função de pertinência. Ela representa numericamente o grau de certeza em que determinado elemento pertença a um conjunto específico. A função de pertinência mapeia os elementos do universo de discurso entre os valores 0 e 1, isso quer dizer que quanto mais próximo de 1 a função de pertinência mapeia um dado elemento, mais próximo do conceito de um determinado conjunto esse elemento estará.

Matematicamente falando, pode-se denominar um elemento contido em um universo de discurso U de x . Este mesmo elemento pertence a um conjunto nebuloso A . Então, este conjunto nebuloso pode ser definido como:

$$A = \{(\mu_A(x), x) \mid \mu_A(x) \in [0,1]\}, \text{ onde } \mu_A(x) \text{ é a função de pertinência.}$$

Funções de pertinência podem assumir diversos formatos, dependendo do problema que se deseja modelar. Alguns exemplos de tipos de funções de pertinência que se pode citar são: *triangular*, *trapezoidal*, *sigmoidal* e *linear*.

2.1 Operações com conjuntos nebulosos.

A teoria de conjuntos nebulosos e a lógica nebulosa permitem a realização de operações sobre os conjuntos nebulosos. Sendo A e B dois conjuntos nebulosos expressos em U , podem ser definidas as funções de pertinência a seguir, para cada operação [1].

Interseção: A função de pertinência do operador nebuloso de interseção, denominado T-NORM, pode ser definida de diversas maneiras. Abaixo estão dois dos exemplos mais usuais:

$$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\} \quad \text{mínimo}$$

$$\mu_{A \cap B}(x) = \mu_A(x) * \mu_B(x) \quad \text{produto}$$

Assim como na lógica clássica, a operação de interseção corresponde ao operador lógico de conjunção “E” [1].

União: A função de pertinência do operador nebuloso de união, denominado T-CONORM (ou S-NORM), pode ser definida, por exemplo, como:

$$\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\} \quad \text{Máximo}$$

$$\mu_{A \cup B}(x) = \min x\{1, \mu_A(x) + \mu_B(x)\} \quad \text{Soma Limitada}$$

Assim como na lógica clássica, a operação de união corresponde ao operador lógico de disjunção “OU” [1].

Complemento: A função do operador nebuloso de complemento pode ser definida por:

$$\neg\mu_A(x) = 1 - \mu_A(x)$$

A operação de complemento corresponde ao operador lógico de negação “NÃO” [1].

Na figura 6, está ilustrado graficamente o resultado das operações acima citadas.

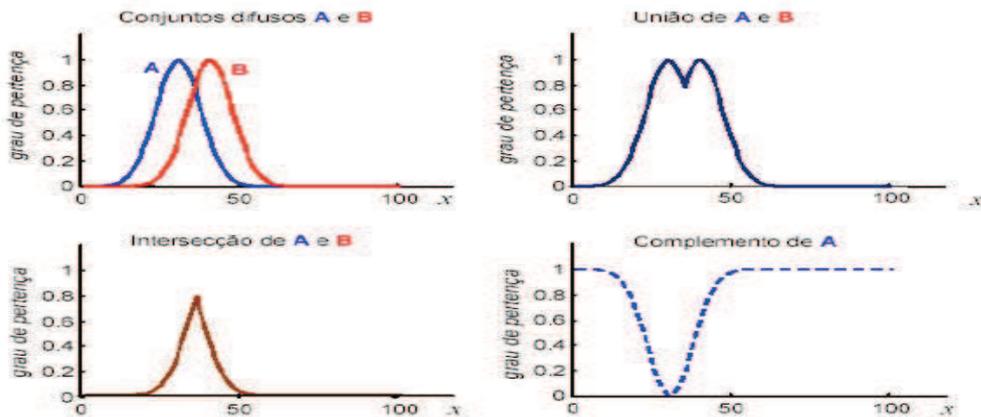


Figura 1: Operações nebulosas realizadas entre os conjuntos A e B [2].

3 Trabalhos Relacionados

3.1.1 Consultas Nebulosas no Fuzzy Query

Uma abordagem relacionada à recuperação nebulosa de informação que pode ser citada para consultas nebulosas é a aplicação Fuzzy Query desenvolvida pela empresa *Sonalysts*, que usa recursos do paradigma da lógica nebulosa para recuperar informação principalmente em bancos de dados relacionais. Com a aplicação, é possível conectar-se a diversas fontes de dados (bancos de dados, arquivos texto com delimitações, planilhas, entre outras fontes) e manipular os dados a serem obtidos usando recursos nebulosos.

Na aplicação, as seguintes opções estão disponíveis:

- Seleção de um atributo e definição de seus respectivos conjuntos nebulosos;
- Construção de uma expressão constituída por termos lingüísticos, modificadores e valores numéricos que determina como será a recuperação de informação de acordo com os conjuntos nebulosos definidos;
- Visualização do efeito de termos lingüísticos como aproximadamente, a cerca de, muito, entre outros termos sobre um conjunto nebuloso.

Na figura abaixo temos um exemplo de criação de um conjunto nebuloso no Fuzzy Query para o atributo média com formato sigmoidal crescente.

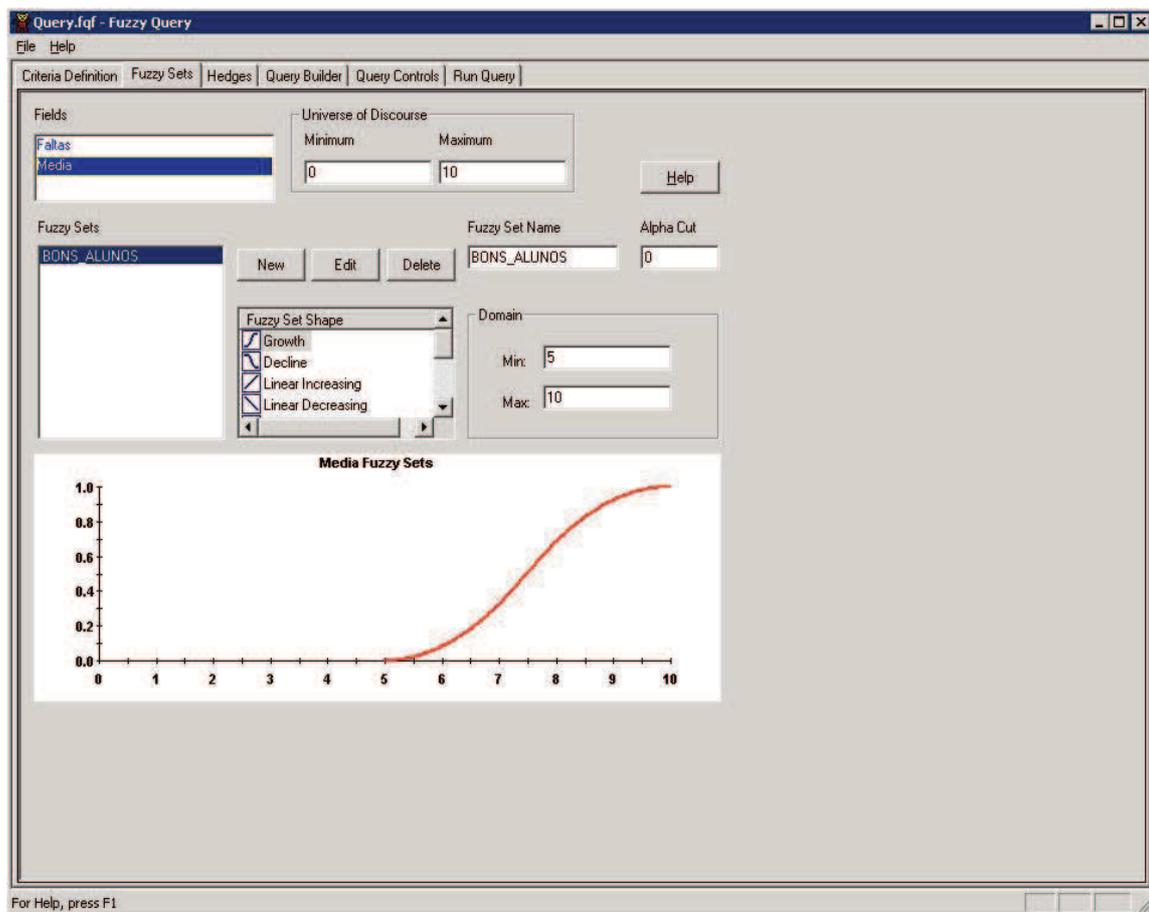


Figura 2: Exemplo de criação de um conjunto nebuloso no Fuzzy Query

O Fuzzy Query carece da funcionalidade de consulta a partir de exemplos, permitindo apenas consultas com a definição dos conjuntos nebulosos, esse aplicativo também não permite o uso de diversos tipos de operadores nebulosos. Outros pontos que deixam o Fuzzy Query em desvantagem em relação à proposta dos métodos a serem implementados por esse projeto são: a ferramenta não é multiplataforma, restringindo o usuário a uso apenas em sistemas operacionais Microsoft Windows; o acesso à base de dados que pode ser feito somente via ODBC; e não ser um projeto aberto e reutilizável.

As vantagens que o Fuzzy Query tem quando comparado com a abordagem de RNI proposta por esse trabalho são: implementação de uma solução híbrida de consulta, onde se pode colocar tanto termos nebulosos quanto termos *crisp*; e apresentação dos conjuntos nebulosos de um atributo em um único gráfico.

3.1.2 Tipos de dados nebulosos no GOA++

Outra abordagem, adotada em bancos de dados orientados a objeto (SGBDOO), é a extensão da API GOA++ de gerência de objetos armazenáveis para inclusão de um tipo de dados *fuzzy* associado a um tipo de dados numérico.

Segundo MAURO [3], o GOA++ é uma API gerenciadora de objetos armazenáveis compatível com o modelo ODMG 2.0 e com a linguagem de definição e consulta ODL e OQL, respectivamente. Sua estrutura interna é flexível e adaptável, o que faz da adição de novos tipos de dados um processo mais fácil e menos custoso.

O tipo de dados *fuzzy* adicionado ao GOA++ é formado por um atributo com o valor base, um atributo com o valor nebuloso associado ao valor base e uma lista de termos lingüísticos (cada termo é uma representação de um conjunto nebuloso para esse atributo) associados a tabelas de correlação entre os valores base e valores *fuzzy*. Com base nessa tabela de correlação, é usada, no processo de “fuzzyficação”, uma função de pertinência de extrapolação para obtenção dos valores não identificados por ela [4]. A abaixo mostra o modelo simplificado das classes relacionadas à definição do tipo de dados *fuzzy*.

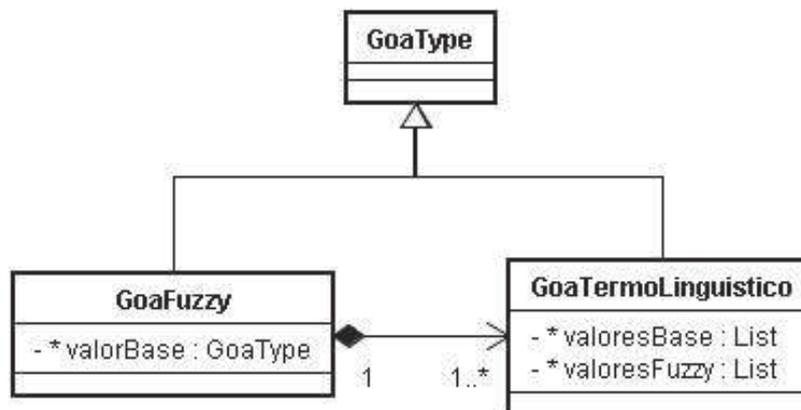


Figura 3: Modelo para definição do tipo de dados fuzzy no GOA++ (linguagem C++)

Comparando o GOA++ com a proposta do trabalho em questão, os seguintes pontos foram identificados como desvantagens: ser uma extensão aplicável apenas a bancos de dados orientados a objeto (SGBDOO); dificuldade para implantar, necessitando de um profissional especializado nesse tipo de banco de dados.

Já a vantagem do GOA++ em relação à proposta abordada está no melhor desempenho na execução das consultas nebulosas, por elas serem realizadas diretamente no banco de dados.

3.1.3 Fuzzy Lookup no SQL Integration Service

O SQL Server Integration Service (SSIS) da Microsoft oferece um processo chamado de Fuzzy Lookup para executar, no nosso caso, uma consulta a partir de exemplos no formato *string*, usando técnicas de lógica nebulosa em um ambiente gráfico simples [5].

O processo de transformação é dividido em quatro etapas:

- Seleção da fonte de dados com registros que serão usados no processo nebuloso (fonte para comparação);
- Seleção da(s) tabela(s) e do(s) campo(s) que serão mostrados na saída dos resultados (fonte para avaliação), associação da fonte de comparação ao(s) campo(s) que será(ão) usado(s) no processo nebuloso, e criação das condições para a saída dos resultados com base nos valores nebulosos que serão obtidos (QBE);
- Processamento nebuloso para obtenção dos resultados com base nas condições definidas no passo anterior;
- Apresentação dos resultados junto com campos adicionais de similaridade e confiança associados a cada registro. Os resultados podem ser exportados para outras fontes de dados ou apenas visualizados.

No processamento nebuloso, primeiramente é realizada a comparação tradicional para verificar se os dados comparados são iguais (comparação *crisp* da fonte de comparação com a fonte de avaliação). Nos casos em que os registros não possuem nenhum registro na fonte de comparação que seja igual a ele, é aplicada a técnica de similaridade entre os dados entre as mesmas fontes.

A técnica de similaridade consiste em dividir os valores (*strings*) do(s) campo(s) escolhido(s) na fonte de comparação em diversas combinações oriundas da original (*tokens*). As combinações obtidas são submetidas ao processo nebuloso, tendo como referência o campo da fonte de avaliação que se deseja encontrar a similaridade com algum valor do(s) campo(s) da fonte de comparação. No término do processo nebuloso, são gerados valores de similaridade e confiança para cada registro.

Nos resultados após o processamento, são acrescentadas as colunas *_Similarity* (similaridade) e *_Confidence* (confiança) para indicar o grau de similaridade encontrado. A coluna de similaridade indica o quanto um registro da fonte de avaliação está próximo de um dos registros da fonte de comparação. O valor 1.0 indica exatidão na comparação. A coluna de confiança indica a probabilidade de que cada possível combinação dos registros da fonte de comparação seja uma combinação da fonte de avaliação.

Quando os atributos de similaridade e confiança têm valores altos, significa que os registros associados a eles estão mais próximos da exatidão.

A abordagem do Fuzzy Lookup tem como desvantagem em relação ao trabalho abordado o fato de não oferecer qualquer tipo de consulta com definição de conjuntos nebulosos (consulta nebulosa). Porém, a vantagem do Fuzzy Lookup está na capacidade de reconhecer uma *string* que seja similar a outra.

4 Abordagem Proposta

A proposta do projeto de Recuperação Nebulosa de Informação tem como objetivo principal modelar, desenvolver e testar dois métodos (funcionalidades) de recuperação nebulosa de informação (RNI), especificamente para atributos com valores numéricos.

Para se chegar ao modo aproximado humano é necessário ter conhecimento dos seguintes elementos:

- Termos lingüísticos: um termo lingüístico é composto por um termo com significados variáveis segundo o ponto de vista de cada individuo, como os termos muito, pouco, alto e quente, seguido de um termo constante como preço, idade e altura ao qual o termo variável faz referencia;
- Operações de interseção e união entre termos lingüísticos: trata-se das operações de interseção (E) e união (OU) da lógica nebulosa que podem ser aplicadas entre dois elementos. O elemento pode ser tanto um termo lingüístico quanto uma composição de termos lingüísticos e operações;
- Expressão nebulosa: uma expressão nebulosa trata-se da representação de um conceito do raciocínio humano por meio de termos lingüísticos e operações da lógica nebulosa entre os termos; e
- Conceito: trata-se de um ou mais termos lingüísticos que expressam uma informação imprecisa.

Após ter sido apresentado os elementos que compõem a recuperação e o seu entendimento ter ficado mais claro, será apresentado a seguir a descrição dos métodos propostos para a recuperação de informação resumidamente:

1. Consulta nebulosa (CN): o método permite definir uma expressão nebulosa formada por termos lingüísticos e operações de interseção e união com intuito de obter resultados que sejam próximos a um conceito do raciocínio humano. Cada termo lingüístico é constituído de um atributo como sendo o termo constante e um conjunto nebuloso associado ao atributo como sendo o termo variável;
2. Consulta a partir de exemplo (CPE): o método permite definir uma expressão nebulosa formada por termos lingüísticos e operações de interseção e união com intuito de obter resultados que sejam próximos a um conceito do raciocínio humano. Cada termo lingüístico é constituído de um atributo como sendo o termo constante e do termo em torno de como sendo o termo variável. O termo “em torno de” tem como objetivo obter resultados que estejam em torno de um valor de exemplo, o conjunto em torno é definido por um número nebuloso triangular onde a base desse triangulo é definida por um percentual de restrição definido pelo o usuário, e o valor de exemplo o ponto de pertinência máxima (vértice superior do triângulo); e
3. Consulta híbrida: este método permite recuperar informações utilizando os recursos disponíveis na consulta nebulosa (CN) e na consulta a partir de exemplo (CPE), ou seja, une os dois métodos citados acima em um único método.

Os métodos de RNI foram disponibilizados no *suite* de mineração de dados JavaBramining implementado por CASTANEDA [6] com exceção do método de consulta híbrida que esta disponível sem interface gráfica.

O intuito dessa extensão foi aproveitar alguns dos recursos de acesso e importação de diversas fontes de dados já disponibilizados pelo JB.

Na Figura 4, ilustra um modelo esquemático do funcionamento geral do protótipo em questão:

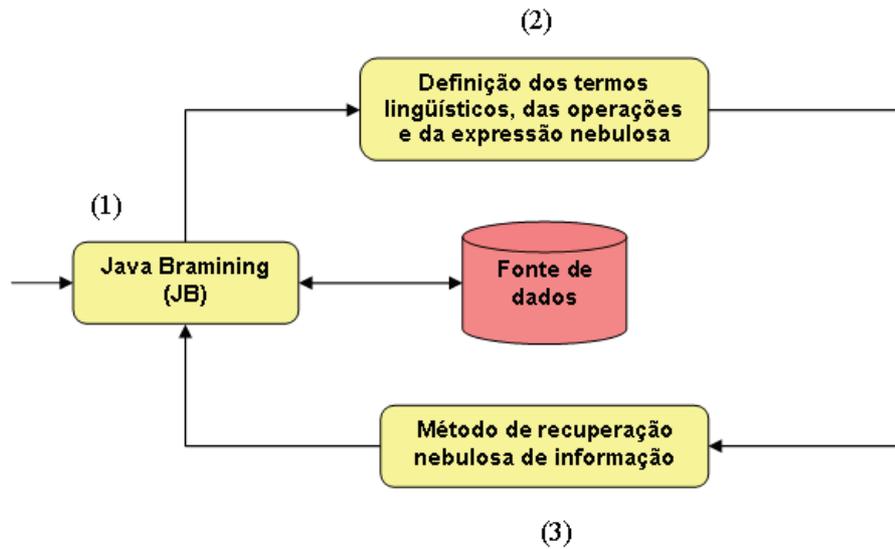


Figura 4: Modelo esquemático do protótipo RNI

O processo inicia-se com o acesso ao *suite* de mineração de dados JavaBramining (1).

No *suite*, uma fonte de dados deverá ser selecionada para se trabalhar e, além de vários métodos de mineração já existentes, serão disponibilizados dois métodos de recuperação nebulosa para serem selecionados. Quando um dos métodos de recuperação nebulosa for selecionado, será iniciada a fase de configuração da recuperação nebulosa com as informações específicas para cada um dos tipos de recuperação oferecidos (2).

Após serem devidamente configuradas as informações, o método deverá ser executado, e deverá ser iniciado o processamento da recuperação nebulosa (3).

Ao ser concluído o processamento, o método retornará os dados com os resultados obtidos e o ciclo de definição e execução dos métodos de recuperação nebulosa poderá ser feito novamente, se necessário.

Um ponto importante a ser considerado é que os dados a serem usados pelo processo de recuperação devem ser previamente extraídos de fontes de dados quaisquer, e agrupados em uma única tabela (conjunto de registros), seja ela normalizada ou não.

A implementação dos métodos foi feita de tal forma que possibilita a reutilização e extensão desses métodos sem que estes tenham qualquer ligação com o JavaBramining. Dessa forma, foi criada a biblioteca *Inexact* (incerta ou que torna valores exatos mais próximos do mundo real). A biblioteca com a implementação dos métodos de RNI está disponível no *SourceForge.net* com o seguinte título: *Inexact - Information Fuzzy Retrieval*.

5 Protótipo, experimento e resultados

5.1 Protótipo

Como já informado em outras seções desse artigo, os dois métodos de RNI propostos foram disponibilizados para o usuário final por intermédio da ferramenta JavaBramining como métodos de pré-processamento com exceção da consulta híbrida.

Para cada método de RNI, foram criadas interfaces (telas) específicas para que o usuário possa preencher as informações necessárias ao método e, posteriormente, executá-lo.

Nas figuras 5 e 6 são ilustradas as interfaces que o método de consulta nebulosa (CN) oferece para definição do processo de recuperação. Na primeira figura (aba *Fuzzy Sets*), é ilustrada a interface de definição dos atributos e seus respectivos conjuntos nebulosos que serão utilizados na construção da expressão nebulosa. O botão *View Fuzzy Set* possibilita que, quando um conjunto nebuloso for selecionado, a sua representação gráfica possa ser visualizada.

Na segunda figura (aba *Query Builder*) é apresentada a interface de construção da expressão nebulosa com a lista dos conjuntos nebulosos associados aos atributos e operadores nebulosos, como também a pertinência mínima e tipo de operadores nebulosos usados na recuperação. O botão *Execute* deve ser acionado assim que toda a configuração nas duas abas do método tenha sido devidamente preenchida.

Figura 5: Tela do método CN (aba *Fuzzy Sets*)

Information Fuzzy Retrieval - Fuzzy Query

File Help

Fuzzy Sets Query Builder Query Results

Predicates and fuzzy operators to be used in the fuzzy expression build

AND
 OR

Fuzzy expression:

Minimum pertinence: Processing type:

Figura 6: Tela do método CN (aba *Query Builder*)

Information Fuzzy Retrieval - Query By Example

File Help

Example Attributes Query Builder Query Results

Attributes: Example value: Percentage:

attribute	example value	percentage
petallength	5	20
sepalwidth	6	20

Figura 7: Tela do método CPE (aba *Example Attributes*)

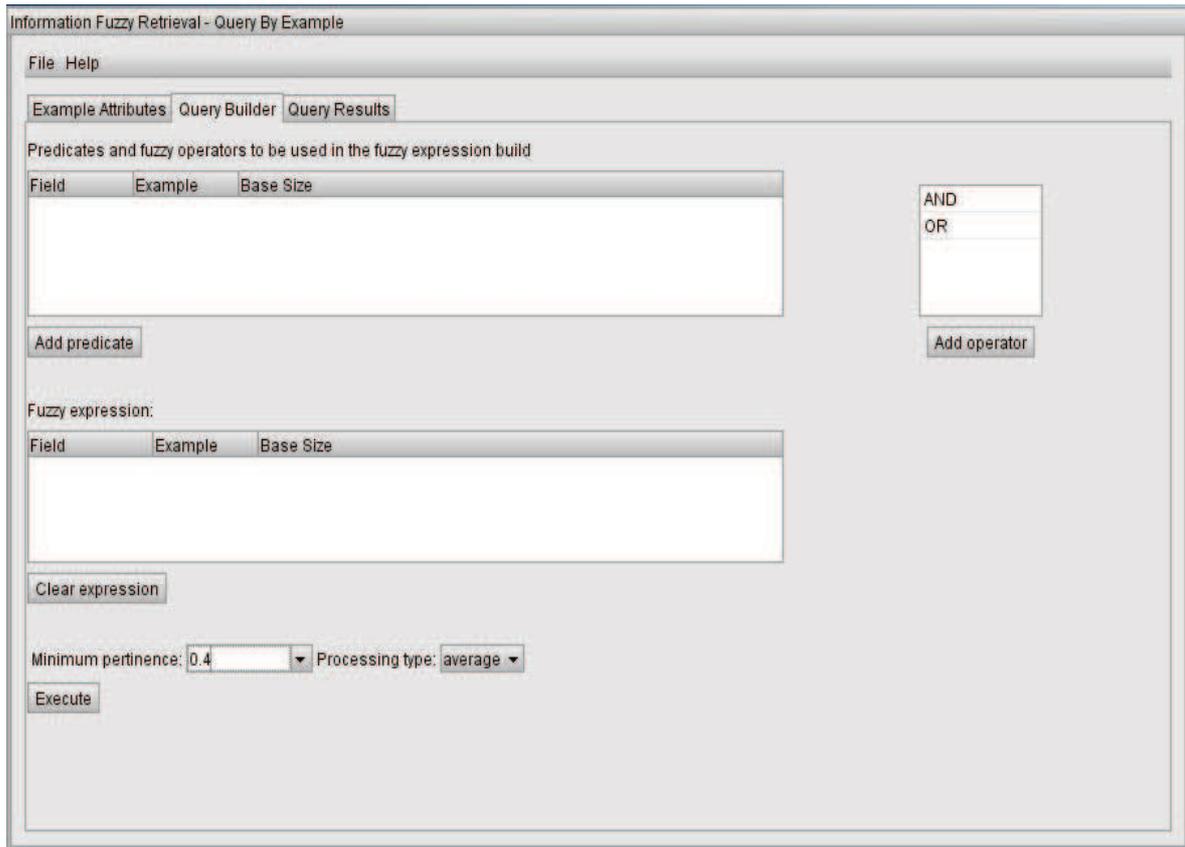


Figura 8: Tela do método CPE (aba *Query Builder*)

Nas figuras 7 e 8 são ilustradas as interfaces que o método de consulta a partir de exemplo (CPE), oferece para definição do processo de recuperação. Na primeira figura (aba *Example Attributes*) é ilustrada a interface de definição dos atributos de exemplo, onde serão informados os valores de exemplo e percentual de similaridade para cada atributo a ser usado na recuperação. O botão *View Example* possibilita que, quando um atributo de exemplo seja selecionado, a sua representação gráfica possa ser visualizada.

Na segunda figura (aba *Query Builder*) é apresentada a interface de construção da expressão nebulosa com a lista dos atributos de exemplo e operadores nebulosos, como também a pertinência mínima e tipo de operadores nebulosos (AND, OR) a serem usados na recuperação, bem como o tipo de processamento para cada operador nebuloso (Zadeh, Média ou Produto) O botão *Execute* deve ser acionado assim que toda a configuração nas duas abas do método tenha sido devidamente preenchida.

JB_retrieval	class	sepalength	sepalwidth	petalwidth	petallength
0.6	Iris-virginica	7.9	3.8	2	6.4
0.6	Iris-virginica	7.7	3.8	2.2	6.7
0.575	Iris-virginica	7.2	3.6	2.5	6.1
0.525	Iris-virginica	7.2	3.2	1.8	6
0.525	Iris-versicolor	7	3.2	1.4	4.7
0.5125000000000000	Iris-virginica	6.9	3.2	2.3	5.7
0.5125000000000000	Iris-setosa	5.7	4.4	0.4	1.5
0.5	Iris-virginica	6.7	3.3	2.5	5.7
0.5	Iris-virginica	6.8	3.2	2.3	5.9
0.5	Iris-virginica	6.9	3.1	2.3	5.1
0.5	Iris-virginica	6.9	3.1	2.1	5.4
0.5	Iris-virginica	7.7	3	2.3	6.1
0.5	Iris-virginica	7.4	2.8	1.9	6.1
0.5	Iris-virginica	7.2	3	1.6	5.8
0.5	Iris-virginica	6.7	3.3	2.1	5.7
0.5	Iris-virginica	7.7	2.8	2	6.7
0.5	Iris-virginica	7.7	2.6	2.3	6.9
0.5	Iris-virginica	7.3	2.9	1.8	6.3
0.5	Iris-virginica	7.6	3	2.1	6.6
0.5	Iris-virginica	7.1	3	2.1	5.9
0.5	Iris-versicolor	6.9	3.1	1.5	4.9
0.4750000000000000	Iris-virginica	6.7	3.1	2.4	5.6
0.4750000000000000	Iris-versicolor	6.7	3.1	1.5	4.7
0.4750000000000000	Iris-versicolor	6.7	3.1	1.4	4.4
0.475	Iris-virginica	6.8	3	2.1	5.5

Figura 9: Tela de resultados dos métodos de RNI

Na Figura 9 é ilustrada a interface utilizada para exibição dos resultados obtidos após a execução de um método de RNI. Ambos os métodos têm essa interface (aba *Query Results*) para exibir os resultados em forma de tabela. O campo adicional *JB_retrieval* contém o valor nebuloso obtido após a avaliação da expressão nebulosa, e determina o quanto o registro está próximo do conceito representado pela expressão (valores tendendo a 1), e o quanto o registro não se enquadra dentro do conceito representado pela expressão (valores tendendo a 0).

Além do botão de execução dos métodos (*Execute*) ilustrado nas figuras acima, cada método tem disponível duas opções de tratamento da configuração: gravação das informações preenchidas na interface (opção *File* → *Save configuration* do menu superior); e carga da configuração que tenha sido gravada em outro momento (opção *File* → *Load configuration* do menu superior).

5.2 Experimento Loja Virtual

O paradigma da lógica nebulosa pode ser aplicado em diversos cenários do dia-a-dia de empresa para realizar a recuperação de informação, facilitando-as a obterem resultados mais próximos do modo como o ser humano raciocina do que os métodos tradicionais.

Neste artigo iremos exemplificar um cenário hipotético onde uma loja de informática deseja fazer uma análise de alguns produtos que vêm apresentando um alto índice de reclamação por parte dos clientes em relação à entrega. Para isso, um especialista determinou que, primeiramente, deve-se fazer uma análise da logística desses produtos, ou seja, seu controle de estoque.

Dessa forma, os seguintes atributos relacionados aos produtos em estoque serão analisados:

- *Quantidade em estoque*: quantidade em número de itens que um produto tem disponível em estoque;
- *Frequência de saída*: frequência de saída de um determinado produto em um intervalo de 6 meses; e
- *Tempo médio de entrega*: tempo médio de entrega (em dias corridos) de um produto por parte do fornecedor.

Para fazer a análise citada anteriormente, foi necessário que um especialista definisse a expressão nebulosa adequada para que se chegasse ao conceito desejado para a análise.

Depois de uma análise feita pelo especialista, chegou-se a seguinte expressão em linguagem natural:

Quantidade em Estoque Baixo
***OU* Quantidade em Estoque Médio *E* Tempo de Entrega Demorada**
***OU* Frequência de Saída em torno de 50**

Nesse caso, foi utilizada uma consulta híbrida onde foram definidos os conjuntos nebulosos Estoque Baixo, Estoque Médio, Entrega Demorada e um exemplo para o atributo Frequência de Saída.

A consulta híbrida é executada construindo-se os conjuntos nebulosos e dando os exemplos para cada atributo com sua restrição, como mostrado no capítulo anterior, existe será usado a aba “Fuzzy Sets” e também a aba “Example Attributes”, ou seja teremos a mistura dos conceitos de consulta nebulosa e consulta a partir de exemplo com restrições, na aba “Query Builder” e possível construir a expressão nebulosa híbrida com os dois conceitos. A interface gráfica para a consulta híbrida ainda não foi implementada portanto na Figura 10 temos o arquivo xml referente a consulta em questão.

```

1 <?xml version="1.0" ?>
2 <!DOCTYPE inexact-config SYSTEM "src/dtd/inexact-hybridquery.dtd">
3 <inexact-config>
4   <attributes>
5     <fuzzy-attribute name="quantidade">
6       <discourse-universe minimum="0" maximum="40" />
7       <fuzzy-sets>
8         <fuzzy-set name="quantidade.baixa" type="sigmoidal-decreasing">
9           <points>
10            <point>0</point>
11            <point>8.5</point>
12            <point>17</point>
13          </points>
14        </fuzzy-set>
15        <fuzzy-set name="quantidade.media" type="sigmoidal-increasing">
16          <points>
17            <point>8</point>
18            <point>13.5</point>
19            <point>35</point>
20          </points>
21        </fuzzy-set>
22      </fuzzy-sets>
23    </fuzzy-attribute>
24    <fuzzy-attribute name="entrega">
25      <discourse-universe minimum="0" maximum="30" />
26      <fuzzy-sets>
27        <fuzzy-set name="entrega.demorada" type="sigmoidal-increasing">
28          <points>
29            <point>4</point>
30            <point>8</point>
31            <point>20</point>
32          </points>
33        </fuzzy-set>
34      </fuzzy-sets>
35    </fuzzy-attribute>
36    <example-attribute name="frequencia" value="50" percentage="30" />
37  </attributes>
38  <retrieval name="pertinencia" minimum="0.0" processing="average" ordering="desc">
39    <fuzzy-predicate attribute="quantidade" set="quantidade.baixa" />
40    <operator type="or" />
41    <fuzzy-predicate attribute="quantidade" set="quantidade.media" />
42    <operator type="and" />
43    <fuzzy-predicate attribute="entrega" set="entrega.demorada" />
44    <operator type="or" />
45    <example-predicate attribute="frequencia" />
46  </retrieval>
47 </inexact-config>

```

Figura 10: Arquivo XML com a consulta híbrida

Tendo a expressão definida, é necessário modelar os conjuntos nebulosos que a compõe. Abaixo segue a tabela com os atributos e os conjuntos nebulosos que determinam a análise desejada. Cada conjunto nebuloso identificado na primeira coluna da tabela está representado graficamente conforme o nome da figura citado ao lado do nome do conjunto.

Atributo e conjunto nebuloso	Formato do conjunto	Enumeração dos pontos do conjunto
Estoque Baixo (Figura 11, linha vermelha)	Sigmoidal decrescente	{0, 8.5, 17}
Estoque Médio (Figura 11, linha verde)	Sigmoidal crescente	{8, 21.5, 35}
Entrega Demorada (Figura 12)	Sigmoidal crescente	{4, 12, 20}
Frequência em torno de 50 (Figura 13)	Triangular (conjunto em torno de 50, com restrição de 30%)	{35, 50, 65}

Tabela 1: Conjuntos nebulosos para expressão definida pelo o especialista.

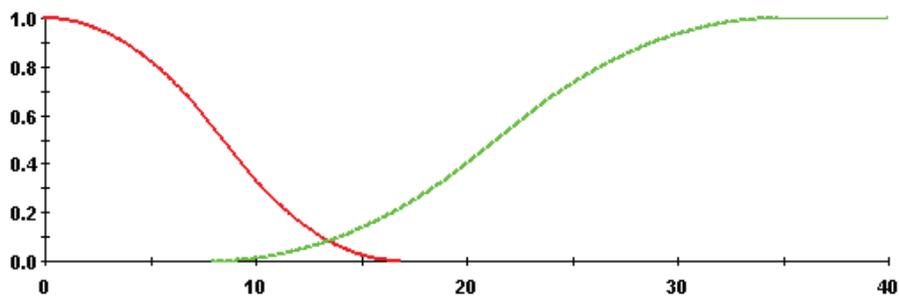


Figura 11: Conjuntos nebulosos Estoque Baixo e Estoque Médio

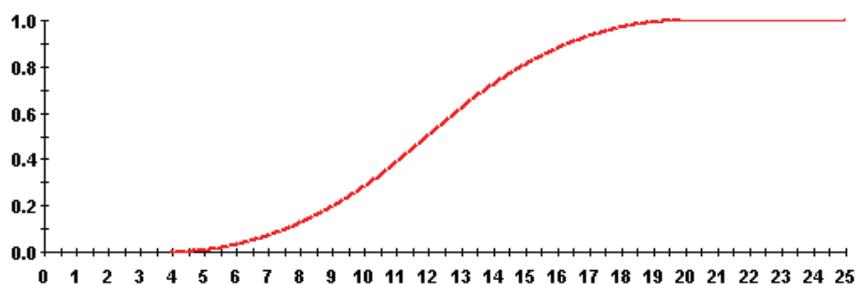


Figura 12: Conjunto Nebuloso Entrega Demorada

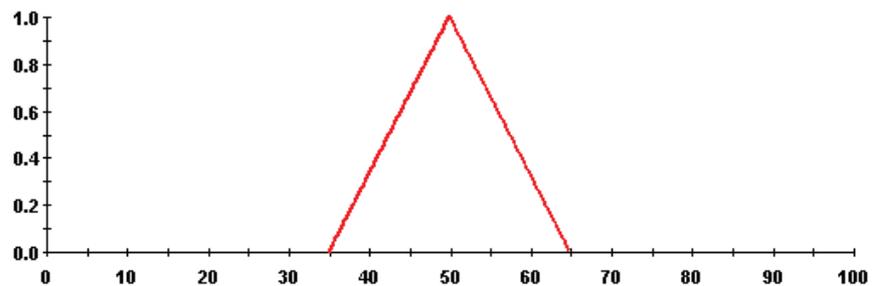


Figura 13: Conjunto frequência em torno de 50, com restrição de 30%

A tabela abaixo representa os produtos que serão analisados pela consulta híbrida.

Nome do Produto	Quantidade em estoque	Frequência de saída	Tempo médio de entrega
Memória 1GB	0	80	17
Processador Athlon 2Ghz	2	58	15
Processador Intel 2.4Ghz	1	55	15
Placa Mae Asus AK-230	5	42	10
Placa Mae elitegroup p223	15	22	9
Nvidia GeForce 7500 512m	7	44	8
Ati Radeon 8900 256m	30	12	6
HD SATA 250GB	22	18	7
HD IDE 180GB	16	20	6
Monitor LCD 15"	12	32	10
Monitor LCD 17"	9	47	8
Monitor LCD 19"	37	12	5
Pendrive 2GB	12	25	12
Pendrive 4GB	9	49	13
Mp3 Player 2GB	5	22	11
Placa Captura ATI	18	19	7
Placa PCTV Creative	33	57	3
Mémoria 2GB	9	28	10
Teclado Sem Fio	21	48	6
Mouse Sem fio	10	34	7

Tabela 2: Lista de Produtos a serem analisados pela expressão nebulosa.

Após a definição de todos os conjuntos nebulosos e a elaboração da expressão nebulosa, os dados apresentados na tabela 2 foram submetidos ao processo da consulta híbrida implementado na biblioteca *Inexact*.

As informações são submetidas em forma de tabela (HashMap) para a biblioteca *Inexact* onde cada tupla é “fuzzificada” de acordo com o conjunto nebuloso associado ao atributo, após a “fuzzificação”, são aplicados os operadores nebulosos (AND e OR) de acordo com o tipo de processamento de cada operador (Zadeh , Media ou Produto) então é avaliada a pertinência e retornado o resultado para a aba “Query Results”.

No caso da consulta híbrida o mesmo procedimento é feito, com o diferencial que a consulta a partir de exemplo é considerada como um conjunto nebuloso triangular onde o ponto de pertinência máxima é o próprio exemplo dado para o atributo escolhido, e a base do triângulo é calculada levando em consideração a restrição fornecida pelo usuário, ou seja , no caso em questão o exemplo de “50” é o ponto de pertinência máxima e a base do triângulo será calculada a partir de 30% de 50 ou seja $50 + 15 = 65$ e $50 - 15 = 35$ (Figura 13). Quando é feita a consulta híbrida, usamos as mesmas operações de AND e OR que usamos na consulta nebulosa e temos o retorno do resultado para a aba “Query Results”.

Os resultados da consulta estão apresentados na tabela abaixo, onde a primeira coluna apresenta a pertinência de cada tupla da amostra de dados submetida ao método, ou seja, o quão próximo ela está do conceito definido pela consulta híbrida (expressão nebulosa). As tuplas estão em ordem decrescente de pertinência para facilitar a visualização.

Pertinência	Nome do Produto	Quantidade em estoque	Frequência de saída	Tempo médio de entrega
0.774	Pendrive 4GB	9	4	13
0.711	Processador Intel 2.4Ghz	1	55	15
0.639	Processador Athlon 2Ghz	2	58	15
0.634	Teclado Sem Fio	21	48	6
0.603	Monitor LCD 17"	9	47	8
0.532	Memória 1GB	0	80	17
0.494	Nvidia GeForce 7500 512m	7	44	8
0.465	Placa PCTV Creative	33	57	3
0.439	Placa Mae Asus AK-230	5	42	10
0.306	Mp3 Player 2GB	5	22	11
0.224	"Monitor LCD 19"	37	12	5
0.217	Ati Radeon 8900 256m	30	12	6
0.209	Pendrive 2GB	12	25	12
0.171	Mémoria 2GB	9	28	10
0.142	HD SATA 250GB	22	18	7
0.116	Monitor LCD 15"	12	32	10
0.100	Mouse Sem fio	10	34	7
0.069	Placa Captura ATI	18	19	7
0.013	Placa Mae elitegroup p223	15	22	9
0.013	HD IDE 180GB	16	20	6

Tabela 3: Resultado da consulta nebulosa.

Com os resultados em mãos, podemos perceber que os produtos com grau de pertinência mais próximo do valor 1 são aqueles produtos que devem ter a maior prioridade na reposição do estoque e maior atenção em relação à entrega, pois são os registros que mais se aproximaram do conceito definido pela consulta.

Outro ponto interessante que pode ser visualizado nos resultados foi que os produtos com uma quantidade baixa de itens em estoque não foram necessariamente os produtos com maior pertinência. Isso aconteceu porque na consulta foram considerados outros atributos tão importantes quanto à quantidade em estoque.

6 Considerações finais

O objetivo da elaboração desse artigo foi de apresentar opções de recuperação de informação que lidem com a incerteza e a imprecisão, termos estes que expressam justamente o modo como o ser humano muitas vezes raciocina.

O paradigma da lógica nebulosa contido no segmento da inteligência computacional tornou possível que tais termos se apliquem em ambientes computacionais, no intuito de efetuar a recuperação de informação. Esse mesmo paradigma pode ser utilizado em diversas áreas, tais como a administrativa, a científica e a geomática.

Foram apresentados então três métodos para a recuperação de informação usando a lógica nebulosa: consulta nebulosa, consulta a partir de exemplo e consulta híbrida. O primeiro permite realizar consultas por meio da construção de uma expressão nebulosa formada por termos lingüísticos e operadores nebulosos. O segundo permite realizar consultas a partir de uma tupla de exemplo para que sejam obtidos os registros semelhantes

a ela. Já o terceiro permite realizar consultas usando os recursos oferecidos tanto pela consulta nebulosa quanto pela consulta a partir de exemplo.

Para demonstrar como a lógica nebulosa pode ser útil na obtenção de informações mais próximas do que o ser humano raciocina, foi realizado um experimento com dados hipotéticos para se analisar o comportamento logístico dos produtos em estoque de uma loja de produtos de informática.

Os resultados da consulta mostraram-se satisfatórios ao que se desejava encontrar com a expressão nebulosa definida pelo especialista, tornando estes resultados mais próximos de um conceito definido pelo raciocínio humano do que o uso da lógica clássica.

7 Referências

- [1] GOLDSCHMIDT, Ronaldo Ribeiro; OMAR, Paulo. **Uma Introdução à Lógica Nebulosa**. Centro Universitário da Cidade do Rio de Janeiro, Rio de Janeiro, 2006.
- [2] ENIO, S.; DANTAS, M. **Uma abordagem de monitoração de tráfego de rede utilizando lógica difusa**. Artigo apresentado a SBC, 2005.
- [3] MAURO, R. C. *et al.* “GOA++: Tecnologia, Implementação e Extensões aos Serviços de Gerência de Objetos”. In: **Anais do XII Simpósio Brasileiro de Banco de Dados**. Fortaleza, Ceara, 1997. p. 272-286.
- [4] BOULLOSA, J. R. F.; CRUZ, F. C. A.; XEXEO, G. Incerteza em Bancos de Dados: Tipo de Dados Nebulosos no GOA++. **XIV Simpósio Brasileiro de Bancos de Dados**. SBBD99. Florianópolis, SC, 1999.
- [5] NATHAN, Jay. **Fuzzy Lookups and Groupings Provide Powerful Data Cleansing Capabilities**. Disponível em <<http://msdn.microsoft.com/msdnmag/issues/05/09/SQLServer2005/>>. Visualizado em 15/04/2007. MSDN Magazine, 2005.
- [6] CASTANEDA, Rafael. J.Bramining: **Um ambiente para descoberta de Conhecimento em Bases de Dados**. Centro Universitário da Cidade do Rio de Janeiro, Rio de Janeiro, 2005.