

UNIVERSIDADE ESTADUAL PAULISTA  
“Júlio de Mesquita Filho”  
Pós-Graduação em Ciência da Computação

Eduardo Roque Nóbrega de Assis

Uma Abordagem Nebulosa de Raciocínio Baseado em Casos para  
Recomendação de Sistemas Marítimos de Produção de Petróleo

UNESP  
São José do Rio Preto - 2012

Eduardo Roque Nóbrega de Assis

## Uma Abordagem Nebulosa de Raciocínio Baseado em Casos para Recomendação de Sistemas Marítimos de Produção de Petróleo

Dissertação de mestrado apresentada para a obtenção do título de Mestre em Ciência da Computação, área de Concentração em Sistemas Inteligentes, junto ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Orientador: Prof. Dr. Ivan Rizzo Guilherme

Co-Orientador: Prof. Dr. João Fernando Marar

UNESP  
São José do Rio Preto – 2012

Assis, Eduardo Roque Nóbrega de.

Uma abordagem nebulosa de raciocínio baseado em casos para recomendação de sistemas marítimos de produção de petróleo / Eduardo Roque Nóbrega de Assis. - São José do Rio Preto : [s.n.], 2012.

94 f. : il. ; 30 cm.

Orientador: Ivan Rizzo Guilherme

Coorientador: João Fernando Marar

Dissertação (mestrado) - Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas

1. Computação. 2. Inteligência artificial. 3. Agentes inteligentes (Software). 4. Raciocínio baseado em casos. 5. Lógica difusa. 6. Sistemas de produção de petróleo. I. Guilherme, Ivan Rizzo. II. Marar, João Fernando. III. Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas. IV. Título.

CDU – 004.891

Eduardo Roque Nóbrega de Assis

## Uma Abordagem Nebulosa de Raciocínio Baseado em Casos para Recomendação de Sistemas Marítimos de Produção de Petróleo

Dissertação de mestrado apresentada para a obtenção do título de Mestre em Ciência da Computação, área de Concentração em Sistemas Inteligentes, junto ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

### BANCA EXAMINADORA

Prof. Dr. Ivan Rizzo Guilherme  
UNESP – Rio Claro  
Orientador

Prof. Dr. Eraldo Pereira Marinho  
UNESP – Rio Claro

Profa. Dra. Sandra Aparecida Sandri  
INPE – São José dos Campos

UNESP  
São José do Rio Preto, 03 de Fevereiro de 2012

*Dedico este trabalho ao Gordão e à Sô.*

## **AGRADECIMENTOS**

Primeiramente agradeço ao amigo, orientador e Professor Ivan Rizzo Guilherme pela oportunidade de realizar esse trabalho e outras concedidas.

Agradeço a todos os amigos e familiares que me deram suporte, apoio e me incentivaram durante a realização deste trabalho.

Particularmente agradeço aos amigos da Caiena, que vivenciaram diariamente a evolução do trabalho, sempre com apoio e paciência. Em especial aos grandes amigos Spantalho, Gueixa, Largato e Cabeça, que me acompanham desde o início, dando todo o suporte necessário.

À De, por todo carinho, apoio e compreensão que sempre empregou.

Aos membros da banca examinadora, Professor Eraldo e Professora Sandra, pelas excelentes contribuições que fizeram para o trabalho.

Agradeço à Deolinda pelos valiosos esclarecimentos e fornecimento do conhecimento adquirido em seu trabalho.

# SUMÁRIO

|                                                                |    |
|----------------------------------------------------------------|----|
| <b>1 Introdução</b>                                            | 1  |
| 1.1 Objetivos do Trabalho                                      | 2  |
| 1.2 Organização do Trabalho                                    | 2  |
| <b>2 Aspectos Conceituais</b>                                  | 4  |
| 2.1 Sistemas Marítimos de Produção de Petróleo                 | 4  |
| 2.2 Raciocínio Baseado em Casos (RBC)                          | 5  |
| 2.2.1 Histórico do RBC                                         | 6  |
| 2.2.2 Vantagens da Abordagem de Raciocínio Baseado em Casos    | 6  |
| 2.2.3 Desvantagens da Abordagem de Raciocínio Baseado em Casos | 9  |
| 2.2.4 Fundamentos de Raciocínio Baseado em Casos               | 9  |
| 2.2.5 O que é um caso?                                         | 11 |
| 2.2.6 O Ciclo do RBC                                           | 14 |
| 2.2.7 Representação dos Casos                                  | 15 |
| 2.2.7.1 Indexação                                              | 16 |
| 2.2.8 Recuperação de Casos                                     | 18 |
| 2.2.8.1 Conceito de similaridade                               | 21 |
| 2.2.9 Reutilização dos casos                                   | 23 |
| 2.2.9.1 Cópia                                                  | 23 |
| 2.2.9.2 Adaptação                                              | 24 |
| 2.2.10 Revisão da solução                                      | 24 |
| 2.2.10.1 Avaliação de Solução                                  | 25 |
| 2.2.10.2 Reparação de Falha                                    | 25 |
| 2.2.11 Retenção dos casos                                      | 25 |
| 2.3 Teoria dos Conjuntos Nebulosos                             | 28 |
| 2.3.1 Definição de conjunto nebuloso                           | 28 |
| 2.3.2 Funções de Pertinência                                   | 30 |
| 2.3.3 Variáveis Linguísticas                                   | 31 |
| 2.3.4 Características nebulosas                                | 32 |
| 2.3.5 Similaridade nebulosa                                    | 33 |
| 2.3.6 Geometria dos conjuntos nebulosos                        | 34 |
| 2.4 Raciocínio Baseado em Casos Flexível                       | 36 |
| <b>3 Abordagem Proposta</b>                                    | 37 |
| 3.1 Representação                                              | 38 |

|                                                                                             |           |
|---------------------------------------------------------------------------------------------|-----------|
| 3.2 Indexação .....                                                                         | 40        |
| 3.3 Recuperação .....                                                                       | 43        |
| 3.3.1 Cálculo da similaridade .....                                                         | 43        |
| 3.3.2 Região de similaridade .....                                                          | 46        |
| 3.4 Adaptação .....                                                                         | 47        |
| 3.4.1 Reinstanciação .....                                                                  | 48        |
| 3.5 Retenção – Manutenção da base de casos .....                                            | 49        |
| <b>4 RBC Nebuloso para Recomendação de Sistemas Marítimos de Produção de Petróleo .....</b> | <b>51</b> |
| 4.1 Arquitetura do sistema .....                                                            | 51        |
| 4.2 Implementação do sistema .....                                                          | 51        |
| 4.2.1 Representação dos sistemas marítimos de produção .....                                | 52        |
| 4.2.2 Atributos para definição do sistema de produção .....                                 | 55        |
| 4.2.3 Definição dos conjuntos nebulosos .....                                               | 56        |
| 4.2.4 Recuperação dos casos similares .....                                                 | 61        |
| 4.2.5 RIGS .....                                                                            | 64        |
| 4.2.5.1 Testes e Avaliação .....                                                            | 69        |
| 4.2.5.2 Resultados .....                                                                    | 83        |
| <b>5 Conclusões .....</b>                                                                   | <b>88</b> |
| <b>Referências Bibliográficas .....</b>                                                     | <b>92</b> |

## LISTA DE ABREVIATURAS E SIGLAS

|                    |                                         |
|--------------------|-----------------------------------------|
| <b><i>CSS</i></b>  | <i>Cascading Style Sheets</i>           |
| <b><i>DOM</i></b>  | <i>Document Object Model</i>            |
| <b><i>JSON</i></b> | <i>JavaScript Object Notation</i>       |
| <b><i>k-NN</i></b> | <i>k-nearest neighbors</i>              |
| <b><i>RBC</i></b>  | <i>Raciocínio baseado em casos</i>      |
| <b><i>SQL</i></b>  | <i>Structured query language</i>        |
| <b><i>UEP</i></b>  | <i>Unidade Estacionária de Produção</i> |
| <b><i>XUL</i></b>  | <i>XML User Interface Language</i>      |

## GLOSSÁRIO

**Framework:** Em desenvolvimento de *software*, é uma estrutura de suporte onde um outro projeto pode ser desenvolvido, compreendendo programas de suporte, bibliotecas de código, linguagens de *script* e outros *softwares* auxiliares no desenvolvimento e composição de um projeto de *software*.

**Layout Submarino:** Arranjo dos equipamentos no fundo do mar.

**Manifold:** Conjunto de tubulações munidas de válvulas (Franco, 2003).

**Riser:** Um duto através do qual o líquido ou o gás escoam para cima até a unidade estacionária de produção (Franco, 2003).

## LISTA DE FIGURAS

|                                                                                                                                                           |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 1:</b> Ciclo do RBC. Fonte: Adaptado de Aamodt & Plaza (1994).....                                                                              | 14 |
| <b>Figura 2:</b> Processos envolvidos no RBC. Fonte: Adaptado de Pal & Shiu (2004).....                                                                   | 21 |
| <b>Figura 3:</b> Distância entre casos. Fonte: Pal & Shiu (2004).....                                                                                     | 27 |
| <b>Figura 4:</b> Função de pertinência do conjunto nebuloso "jovem" (Pal & Shiu, 2004).....                                                               | 29 |
| <b>Figura 5:</b> Função de pertinência triangular.....                                                                                                    | 30 |
| <b>Figura 6:</b> Função de pertinência trapezoidal.....                                                                                                   | 30 |
| <b>Figura 7:</b> Função de pertinência sigmóide ( $a > 0$ ).....                                                                                          | 31 |
| <b>Figura 8:</b> Função de pertinência sigmóide ( $a < 0$ ).....                                                                                          | 31 |
| <b>Figura 9:</b> RBC nebuloso (Pal & Shiu, 2004).....                                                                                                     | 33 |
| <b>Figura 10:</b> Conjuntos como pontos. O conjunto A é um ponto no 2-cubo unitário com coordenadas ou pertinência ( $1/3 \ 3/4$ ) (Kosko, 1992).....     | 36 |
| <b>Figura 11:</b> Exemplo de modelo de compra de imóvel.....                                                                                              | 39 |
| <b>Figura 12:</b> Registro de caso de compra de imóvel.....                                                                                               | 39 |
| <b>Figura 13:</b> Exemplo numérico em três dimensões com dois atributos: Idade e Peso. O ponto no cubo representa uma pessoa com 45 anos e 65 quilos..... | 42 |
| <b>Figura 14:</b> Representação de dois conjuntos nebulosos num hipercubo bidimensional unitário, e a distância entre eles.....                           | 45 |
| <b>Figura 15:</b> Exemplo de hiperesfera de similaridade.....                                                                                             | 47 |
| <b>Figura 16:</b> Exemplo de reinstanciação.....                                                                                                          | 49 |
| <b>Figura 17:</b> Modelo relacional representando os sistemas marítimos de produção.....                                                                  | 54 |
| <b>Figura 18:</b> Representação dos casos contendo os atributos da descrição (índices) e da solução do problema.....                                      | 56 |
| <b>Figura 19:</b> Conjunto nebuloso Área do Reservatório (Franco, 2003).....                                                                              | 57 |
| <b>Figura 20:</b> Conjunto nebuloso Profundidade do Reservatório (Franco, 2003).....                                                                      | 57 |
| <b>Figura 21:</b> Conjunto nebuloso Número de Poços (Franco, 2003).....                                                                                   | 58 |
| <b>Figura 22:</b> Conjunto nebuloso Vazão Diária por Poço (Franco, 2003).....                                                                             | 58 |
| <b>Figura 23:</b> Conjunto nebuloso Lâmina d'água (Franco, 2003).....                                                                                     | 58 |
| <b>Figura 24:</b> Conjunto nebuloso Distância da Costa (Franco, 2003).....                                                                                | 59 |
| <b>Figura 25:</b> Conjunto nebuloso Capacidade de Armazenamento (Franco, 2003).....                                                                       | 59 |
| <b>Figura 26:</b> Exemplo de três dimensões de um caso com 22 poços e 230 Km de distância da costa, representado no espaço nebuloso Q por um ponto.....   | 61 |
| <b>Figura 27:</b> Janela de login do RIGS.....                                                                                                            | 64 |
| <b>Figura 28:</b> Ferramenta principal do RIGS, com ferramentas de busca, navegação, visualização e                                                       |    |

|                                                                                                                                                                                                                         |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| manutenção da base de casos.....                                                                                                                                                                                        | 65 |
| <b>Figura 29:</b> Exemplo da utilização da ferramenta de busca de casos. Neste exemplo, os casos que possuem o atributo condições ambientais igual a amena são exibidos em destaque.....                                | 66 |
| <b>Figura 30:</b> Ferramenta de manutenção dos conjuntos nebulosos.....                                                                                                                                                 | 67 |
| <b>Figura 31:</b> Ferramenta de início do RBC e da recuperação de casos similares.....                                                                                                                                  | 68 |
| <b>Figura 32:</b> Resultado de concordância e discordância obtido entre todos os atributos das soluções recomendadas para os 23 casos consulta.....                                                                     | 83 |
| <b>Figura 33:</b> Resultado de concordância e discordância obtido entre todos os atributos das soluções recomendadas para os 14 casos consulta que possuem casos recuperados com similaridade maior que 0.5.....        | 84 |
| <b>Figura 34:</b> Resultado de concordância e discordância em relação aos casos completos, obtido das soluções recomendadas para os 23 casos consulta.....                                                              | 85 |
| <b>Figura 35:</b> Resultado de concordância e discordância em relação aos casos completos, obtido das soluções recomendadas para os 14 casos consulta que possuem casos recuperados com similaridade maior que 0.5..... | 85 |
| <b>Figura 36:</b> Relação das soluções discordantes e o número de atributos discordantes da solução real. ....                                                                                                          | 86 |

## LISTA DE TABELAS

|                                                                                                                      |    |
|----------------------------------------------------------------------------------------------------------------------|----|
| <b>Tabela 1:</b> Relação das funções de pertinência com os termos linguísticos de cada conjunto nebuloso .....       | 60 |
| <b>Tabela 2:</b> Casos utilizados para o cálculo de similaridade (valores aproximados para facilitar o cálculo)..... | 63 |
| <b>Tabela 3:</b> Similaridade entre o caso A e os casos B, C, D e E.....                                             | 63 |
| <b>Tabela 4:</b> Resultado da recuperação dos casos similares ao caso consulta 6.....                                | 71 |
| <b>Tabela 5:</b> Resultado da recuperação dos casos similares ao caso consulta 27.....                               | 71 |
| <b>Tabela 6:</b> Resultado da recuperação dos casos similares ao caso consulta 42.....                               | 72 |
| <b>Tabela 7:</b> Resultado da recuperação dos casos similares ao caso consulta 51.....                               | 72 |
| <b>Tabela 8:</b> Resultado da recuperação dos casos similares ao caso consulta 98.....                               | 73 |
| <b>Tabela 9:</b> Resultado da recuperação dos casos similares ao caso consulta 130.....                              | 73 |
| <b>Tabela 10:</b> Resultado da recuperação dos casos similares ao caso consulta 162.....                             | 74 |
| <b>Tabela 11:</b> Resultado da recuperação dos casos similares ao caso consulta 164.....                             | 74 |
| <b>Tabela 12:</b> Resultado da recuperação dos casos similares ao caso consulta 165.....                             | 75 |
| <b>Tabela 13:</b> Resultado da recuperação dos casos similares ao caso consulta 166.....                             | 75 |
| <b>Tabela 14:</b> Resultado da recuperação dos casos similares ao caso consulta 168.....                             | 76 |
| <b>Tabela 15:</b> Resultado da recuperação dos casos similares ao caso consulta 219.....                             | 76 |
| <b>Tabela 16:</b> Resultado da recuperação dos casos similares ao caso consulta 173.....                             | 77 |
| <b>Tabela 17:</b> Resultado da recuperação dos casos similares ao caso consulta 175.....                             | 77 |
| <b>Tabela 18:</b> Resultado da recuperação dos casos similares ao caso consulta 178.....                             | 78 |
| <b>Tabela 19:</b> Resultado da recuperação dos casos similares ao caso consulta 185.....                             | 78 |
| <b>Tabela 20:</b> Resultado da recuperação dos casos similares ao caso consulta 196.....                             | 79 |
| <b>Tabela 21:</b> Resultado da recuperação dos casos similares ao caso consulta 222.....                             | 79 |
| <b>Tabela 22:</b> Resultado da recuperação dos casos similares ao caso consulta 235.....                             | 80 |
| <b>Tabela 23:</b> Resultado da recuperação dos casos similares ao caso consulta 276.....                             | 80 |
| <b>Tabela 24:</b> Resultado da recuperação dos casos similares ao caso consulta 308.....                             | 81 |
| <b>Tabela 25:</b> Resultado da recuperação dos casos similares ao caso consulta 328.....                             | 81 |
| <b>Tabela 26:</b> Resultado da recuperação dos casos similares ao caso consulta 355.....                             | 82 |

## RESUMO

Um dos grandes desafios existentes na elaboração de um projeto de desenvolvimento de um campo marítimo de petróleo é a definição do sistema de produção. A dificuldade em escolher um bom sistema marítimo de produção de óleo está ligada a aspectos técnicos, econômicos, ambientais e políticos. A experiência adquirida ao longo dos anos pelos engenheiros é um fator essencial no sucesso da tomada dessas decisões em um projeto de produção.

Diante deste cenário, existem algumas abordagens computacionais, na área de Inteligência Artificial, que podem ser utilizadas para a elaboração e desenvolvimento de sistemas inteligentes que auxiliam na solução de novos problemas por analogia a problemas antigos, baseado nas experiências adquiridas. O Raciocínio Baseado em Casos é uma dessas abordagens, e consiste basicamente na construção de sistemas baseado em conhecimento, tendo como princípio adaptar soluções antigas para resolver novos problemas, utilizando casos antigos para criar uma solução equiparável para um novo problema.

O objetivo deste trabalho é realizar o estudo de diversos métodos e técnicas de Raciocínio Baseado em Casos, principalmente as utilizadas para realizar a indexação e recuperação de casos, considerando a abordagem da Teoria dos Conjuntos Nebulosos, para a posterior aplicação dos resultados na construção de um sistema nebuloso de raciocínio baseado em casos para recomendação de sistemas marítimos de produção de petróleo.

## **ABSTRACT**

A major challenge in the elaboration of a project for development of offshore oilfields is the definition of the production system. The difficulty in choosing a good offshore oil production system is related to technical, economic, environmental and political aspects. The experience acquired over the years by engineers is an essential factor in the success of these decisions in a production project.

In this scenario, there are some computational approaches, in the area of Artificial Intelligence, that can be used for the elaboration and development of intelligent systems that assist in solving new problems by analogy to old problems, based on experiences. The Case-Based Reasoning is one of these approaches, and basically consists in building knowledge-based systems, based on the principle to adapt old solutions to solve new problems, using old cases to create a solution equivalent to a new problem.

The objective of this work is to make the study of various methods and techniques of Case-Based Reasoning, especially those used to perform the indexing and retrieval of cases, considering the approach of Fuzzy Set theory, for the subsequent application of the results in the development of a fuzzy case-based reasoning system for recommendation of offshore oil production systems.

# 1 Introdução

A partir de observações feitas sobre o comportamento cognitivo humano pode-se verificar que as pessoas quando solucionam ou interpretam problemas, muitas vezes se apoiam em experiências prévias.

O Raciocínio Baseado em Casos (RBC) é uma abordagem de Inteligência Artificial utilizada para a construção de sistemas inteligentes baseados em conhecimento e tem como princípio a solução de novos problemas por analogia a problemas antigos. Os sistemas implementados como RBC têm como princípio adaptar soluções antigas para resolver novos problemas, utilizando casos antigos para explicar ou criticar novas situações, raciocinar a partir de precedentes para interpretar uma nova situação ou criar uma solução equiparável para um novo problema (Kolodner, 1993).

Alguns pesquisadores afirmam que o conhecimento ou o significado de certos fatos empíricos nem sempre se comportam de maneira probabilística ou mesmo determinística, envolvendo muitas vezes alguma forma de aproximação ou raciocínio aproximado. Nas situações reais, atributos são frequentemente especificados de maneira incompleta ou com incertezas. Desse modo, surge a necessidade de elaboração de um sistema que permita analisar tal conhecimento de maneira a permitir que estas informações sejam levadas em consideração durante a avaliação do problema. Nas situações em que os atributos são melhor descritos através de termos linguísticos, assumindo características qualitativas e portando contendo incertezas, a Teoria dos Conjuntos Nebulosos pode ser aplicada para facilitar a descrição dos conceitos incertos ou incompletos contidos nesses termos.

Uma importante tomada de decisão de um projeto de desenvolvimento de um campo de petróleo é a definição do sistema de produção. A complexidade dessa decisão é devida a fatores ligados principalmente ao ambiente em que será desenvolvido o novo campo petrolífero, a aspectos técnicos, econômicos, ambientais e políticos. A definição do sistema de produção para um determinado campo depende especificamente dos aspectos do ambiente do campo petrolífero e da configuração do sistema de produção, onde é considerado um grande número de fatores, como lâmina d'água, área e profundidade do reservatório, condições ambientais, entre outros (Franco, 2003).

Um sistema marítimo de produção consiste basicamente de uma ou mais unidades estacionárias de produção e de equipamentos posicionados tanto sobre esta unidade quanto no fundo do mar. A experiência adquirida ao longo dos anos pelos engenheiros é um fator essencial e preponderante no sucesso da tomada dessas decisões na elaboração de um projeto de produção. Levando em consideração que decisões equivocadas podem proporcionar uma série de custos

extras, a análise dos parâmetros mais importantes no processo de escolha de alternativas de sistemas de produção deve ser realizada tendo como principal base de informação o conhecimento gerado por projetos anteriores.

Alguns atributos importantes a serem considerados nessa análise são: o arranjo dos poços e a presença de *manifold*; tipos de Unidade Estacionária de Produção – que leva em consideração a lâmina d'água e a área do reservatório, entre outros -, *risers* e ancoragem utilizadas; tipo de transporte utilizado na exportação do óleo produzido até a costa; entre outros atributos (Franco, 2003).

Diante da problemática da definição de um sistema marítimo de produção, em que as experiências adquiridas são fator fundamental para o seu sucesso, e tendo em vista que a abordagem de Raciocínio Baseado em Casos se utiliza exatamente dessas experiências anteriores para a solução de novos problemas, é válida a elaboração e o desenvolvimento de um sistema baseado em casos, e suas diversas técnicas, para auxiliar a escolha do sistema de produção e facilitar o processo decisório dos especialistas.

Nesse contexto, existem técnicas e abordagem computacionais, na área de Inteligência Artificial, capazes de auxiliar a tomada de decisão para a solução de problemas baseada em experiências anteriores. Uma dessas abordagens e foco de estudo deste trabalho, é o Raciocínio Baseado em Casos.

## **1.1 Objetivos do Trabalho**

Este trabalho tem como objetivo a construção de um sistema nebuloso de raciocínio baseado em casos para recomendação de sistemas marítimos de produção de petróleo. Para alcançar o objetivo proposto, será realizado o levantamento bibliográfico da técnica de Raciocínio Baseado em Casos e da Teoria dos Conjuntos Nebulosos, bem como o estudo das técnicas e métodos utilizados para a construção de sistemas RBC, com o foco principal nas diversas técnicas e métodos de indexação e recuperação de casos, para a posterior aplicação dos resultados.

## **1.2 Organização do Trabalho**

Este trabalho está organizado em 5 capítulos, compreendendo os aspectos conceituais, a abordagem proposta para o desenvolvimento de um sistema nebuloso de RBC e o estudo de caso realizado para verificação e validação das técnicas e métodos utilizados.

O capítulo 2 apresenta conceitos básicos de Sistemas Marítimos de Produção, um breve histórico do RBC, seus conceitos fundamentais, seu ciclo de processos e a descrição destes processos. Também é apresentada a Teoria dos Conjuntos Nebulosos e seus principais aspectos.

No capítulo 3 é apresentada a abordagem proposta para a construção de um sistema nebuloso de RBC, as técnicas e métodos utilizadas nos processos e etapas do RBC.

O capítulo 4 apresenta a construção do sistema nebuloso de raciocínio baseado em casos para recomendação de sistemas marítimos de produção de petróleo, os testes realizados para validar a abordagem proposta e seus resultados.

Por fim, no capítulo 5 são feitas as conclusões e apresentadas algumas extensões que podem ser feitas ao trabalho.

## 2 Aspectos Conceituais

Neste capítulo são apresentados os aspectos conceituais necessários para o entendimento do restante do trabalho, e está organizado da seguinte maneira: na seção 2.1 são apresentados os conceitos básicos dos sistemas marítimos de produção de petróleo; a seção 2.2 aborda os fundamentos do Raciocínio Baseado em Casos e as técnicas e métodos envolvidos em cada etapa do seu ciclo. Apresenta também o conceito de similaridade; na seção 2.3 é apresentada a Teoria dos Conjuntos Nebuloso; finalmente, a seção 2.4 aborda o conceito de Raciocínio Baseado em Casos Flexível.

### 2.1 Sistemas Marítimos de Produção de Petróleo

De acordo com Franco (2003), o desenvolvimento de um campo de petróleo deve ser considerado após o descobrimento de uma jazida em uma determinada região. Após o descobrimento da jazida, diversas informações são importantes para o seu desenvolvimento, tais como o tipo e a quantidade de óleo existentes. Essas informações são essenciais para se analisar a viabilidade econômica do desenvolvimento do campo.

Para realizar a extração do petróleo da rocha matriz de um campo marítimo, poços são perfurados no leito oceânico. A disposição desses poços é feita de modo a otimizar o volume de óleo extraído, podendo ser perfurados de forma perpendicular, oblíqua ou com inclinação nula em relação ao solo (Franco, 2003).

A perfuração de um poço de petróleo é realizada através de uma sonda. Na perfuração rotativa, as rochas são perfuradas pela ação da rotação e peso aplicados a uma broca existente na extremidade de uma coluna de perfuração, a qual consiste basicamente de comandos (tubos de paredes espessas) e tubos de perfuração (tubos de paredes finas) (Franco, 2003).

Todo óleo extraído dos poços é enviado para a Unidade Estacionária de Produção (UEP) por *risers* sendo limitado o número de ligações diretas entre os poços e a UEP. O *riser* é uma tubulação de aço especialmente projetado que interliga o fundo à superfície do mar onde encontra-se posicionada uma UEP podendo ter várias configurações. O controle do número de ligações diretas à estrutura é realizado por *manifolds*. Os *manifolds* são conjuntos de válvulas posicionados no fundo do mar capazes de coletar o óleo produzido de diferentes poços e direcioná-lo para a estrutura (Franco, 2003).

Tendo em vista as diversas fases e complexidade da extração do petróleo, é necessário que a empresa que fará a extração selecione, através de um planejamento, um sistema marítimo de

produção que seja técnica e economicamente viável, sem deixar de analisar aspectos ambientais, capaz de operar durante toda vida útil do campo.

Dessa forma, um sistema de produção consiste em uma ou mais unidades estacionárias de produção e os equipamentos de superfície e subsuperfície, ou seja, equipamentos localizados no convés da UEP e os equipamentos localizados no fundo do mar constituindo o *layout* submarino (Franco, 2003).

Os sistemas de produção podem ter configurações e equipamentos variados, de acordo com o local onde serão instalados, devido a fatores que influenciam direta ou indiretamente na fixação desses equipamentos, tais como a lâmina d'água, o tipo de estrutura empregada, e outros.

## **2.2 Raciocínio Baseado em Casos (RBC)**

A partir da observação sobre o comportamento cognitivo humano pode-se verificar que as pessoas quando solucionam ou interpretam problemas, muitas vezes se apóiam em experiências prévias. De acordo com Leake (1996), os seres humanos são robustos resolvidores de problemas; eles freqüentemente resolvem problemas difíceis, apesar das limitações e incertezas que possam haver no conhecimento, melhorando seu desempenho com a experiência. O estudo e desenvolvimento de sistemas inteligentes que solucionam problemas novos por analogia a problemas antigos, tem sido encorajado por estas observações. É neste contexto que surgem os sistemas baseados em casos – sistemas computacionais inteligentes que fazem uso do raciocínio baseado em casos (Kolodner, 1993; Leake, 1996).

O Raciocínio Baseado em Casos (RBC) (Kolodner, 1993) é uma técnica de Inteligência Artificial (IA) que consiste na construção de sistemas baseados em conhecimento tendo como princípio adaptar soluções antigas para resolver novos problemas, utilizando casos antigos para explicar ou criticar novas situações, raciocinar a partir de precedentes para interpretar uma nova situação ou criar uma solução equiparável para um novo problema.

O RBC é útil para uma ampla variedade de tarefas de resolução de problemas, incluindo planejamento, diagnóstico, e recomendação. Em cada uma dessas tarefas, o armazenamento das experiências em forma de casos permite ao raciocinador se tornar mais eficiente, tornando-se mais capaz ao longo do tempo de solucionar problemas que exigem a realização de várias metas conjuntas ou a consideração de múltiplas restrições que interagem entre si (Kolodner, 1993).

Diferentemente da maioria das metodologias de solução de problemas em Inteligência Artificial (IA), o RBC é baseado em memória, refletindo o comportamento humano de lembrar problemas e soluções para serem utilizados como um ponto inicial na solução de um novo problema (Mántaras *et al.*, 2005).

Solucionar um problema com RBC envolve obter uma descrição do problema; mensurar a similaridade do problema atual com problemas anteriores armazenados em uma base de dados

(ou memória) com suas conhecidas soluções; recuperar um ou mais casos similares, e tentar reutilizar a solução de um dos casos recuperados, possivelmente após adaptá-la para considerar as diferenças nas descrições dos problemas (Mántaras *et al.*, 2005).

### 2.2.1 Histórico do RBC

Segundo Aamodt & Plaza (1994), as raízes do raciocínio baseado em casos em IA são encontradas nos trabalhos de Roger Schank (1982) sobre memória dinâmica e o papel central que a lembrança de situações anteriores (episódios, casos) e modelo de situação (*Scripts*, MOPs – pacotes de organização de memória) têm na solução de problemas e aprendizado.

O primeiro sistema que pode ser chamado como um raciocinador baseado em casos foi o sistema CYRUS, desenvolvido por Janet Kolodner (Kolodner, 1983). O sistema foi baseado no modelo de memória dinâmica de Schank e na teoria MOP de solução de problema e aprendizado. Foi basicamente um sistema de questões e respostas com conhecimento de várias viagens e encontros do antigo secretário de estado dos Estados Unidos, Cyrus Vance. O modelo de memória de casos desenvolvido para o sistema serviu posteriormente como base para diversos outros sistemas de raciocínio baseado em casos, como o MEDIATOR (Simpson, 1985), PERSUADER (Sycara, 1992), CHEF (Hammod, 1989), JULIA (Hinrichs, 1992) e CASEY (Koton, 1989), entre outros (Aamodt & Plaza, 1994).

### 2.2.2 Vantagens da Abordagem de Raciocínio Baseado em Casos

Algumas vantagens de se utilizar raciocínio baseado em casos, segundo Pal & Shiu (2004) são:

1. *Reduzir a tarefa de aquisição de conhecimento.* Eliminando a necessidade de se adquirir um modelo do conhecimento como é necessário nos sistemas baseados em conhecimento, as tarefas de aquisição de conhecimento do RBC consistem principalmente da coleção das experiências/casos existentes e sua representação e armazenamento.
2. *Evitar repetir erros cometidos no passado.* Nos sistemas que armazenam as falhas bem como os sucessos, e talvez a razão das falhas, as informações sobre o que causou falhas no passado podem ser usadas para predizer falhas potenciais no futuro.
3. *Prover flexibilidade na modelagem do conhecimento.* Devido à sua rigidez na formulação do problema e modelagem, sistemas baseados em modelos às vezes não conseguem

resolver um problema que está no limite de seus conhecimentos ou escopo ou quando os dados estão incompletos. Por sua vez, sistemas baseados em casos utilizam a experiência passada como conhecimento de domínio e muitas vezes podem fornecer uma solução razoável, através da adaptação apropriada, para esses tipos de problemas.

4. *Raciocinar em domínios que não foram completamente compreendidos, definidos, ou modelados.* Em uma situação onde não existe conhecimento suficiente para construir um modelo causal de um domínio ou para extrair um conjunto de heurísticas para ele, um raciocinador baseado em casos ainda pode ser desenvolvido utilizando somente um pequeno conjunto de casos a partir do domínio. O conhecimento de domínio não tem que ser quantificado ou compreendido inteiramente para um raciocinador baseado em casos funcionar.
5. *Fazer predições do provável sucesso da solução sugerida.* Quando a informação é armazenada considerando o nível de sucesso das soluções passadas, o raciocinador baseado em casos pode prever o sucesso da solução sugerida para o problema atual. Isto é feito consultando as soluções armazenadas, o nível de sucesso destas soluções, e as diferenças entre o anterior e atual contexto da aplicação dessas soluções.
6. *Aprendizagem ao longo do tempo.* Conforme os sistemas RBC são utilizados, mais situações de problemas são encontradas e soluções criadas. Se os casos de solução são testados posteriormente no mundo real e um nível de sucesso é determinado para essas soluções, esses casos podem ser adicionados à base de casos e utilizados para ajudar na solução de problemas futuros. Como os casos são adicionados, um sistema RBC deve ser capaz de raciocinar em um ampla variedade de situações com um grau maior de refinamento e sucesso.
7. *Raciocinar com dados e conceitos incompletos ou imprecisos.* Conforme os casos são recuperados, eles podem não ser idênticos ao caso atual. No entanto, quando eles estão dentro de alguma medida de similaridade definida para o presente caso, qualquer imperfeição e imprecisão pode ser tratada por um raciocinador baseado em casos. Embora esses fatores possam causar uma ligeira degradação no desempenho, devido ao aumento da disparidade entre o caso atual e o recuperado, o raciocínio pode continuar.
8. *Evitar repetir todos os passos necessários para se chegar a solução.* Em domínios de problema que exigem processos significativos para se criar uma solução a partir do zero, a abordagem alternativa de modificar uma solução utilizada anteriormente pode reduzir

significativamente esta exigência de processamento. Além disso, a reutilização de uma solução anterior também permite que os atuais passos adotados para alcançar essa solução possam ser reutilizados para resolver outros problemas.

9. *Prover um meio de explicação.* Sistemas de raciocínio baseado em casos podem fornecer um caso anterior e sua solução (bem sucedida) para ajudar a convencer um usuário, ou para justificar, de uma proposta de solução para o problema atual. Na maioria dos domínios, haverá ocasiões em que um usuário deseja ter a certeza sobre a qualidade da solução fornecida por um sistema. Explicando como um caso anterior foi bem sucedido em uma situação, utilizando as semelhanças entre os casos e o raciocínio envolvido na adaptação, um sistema RBC pode explicar a sua solução para um usuário. Mesmo para um sistema híbrido, que pode estar usando vários métodos para encontrar uma solução, esse mecanismo de explicação proposto pode aumentar a explicação causal (ou outra) dada a um usuário.
10. *Estender para muitas finalidades diferentes.* Um sistema RBC pode ser implementado de diversas formas e utilizado com muitos propósitos, como para criar um plano, fazer um diagnóstico, e argumentar sobre um ponto de vista. Portanto, os dados tratados por um sistema de RBC são capazes de assumir muitas formas, e os métodos de recuperação e adaptação também irão variar.
11. *Estender para uma ampla gama de domínios.* Raciocínio baseado em casos pode ser aplicado a domínios de aplicação extremamente diversos. Isto se deve ao número aparentemente ilimitado de formas de representação, indexação, recuperação e adaptação dos casos.
12. *Refletir o raciocínio humano.* Como existem muitas situações em que nós, seres humanos, utilizamos uma forma de raciocínio baseado em casos, não é difícil convencer os implementadores, usuários e gestores da validade do paradigma. De maneira similar, os seres humanos podem entender o raciocínio e explicações de um sistema RBC e são capazes de ser convencidos da validade das soluções fornecidas pelo sistema. Se um usuário desconfia da validade de uma solução anterior, ele é menos propenso a usar esta solução. Quanto mais crítico o domínio, menor a chance de uma solução passada ser usada e maior o nível exigido de compreensão e credulidade do usuário.

### **2.2.3 Desvantagens da Abordagem de Raciocínio Baseado em Casos**

Algumas desvantagens de se utilizar a abordagem de raciocínio baseado em casos:

1. Casos antigos podem ter descrições incompletas.
2. Muitos casos apropriados podem não ser recuperados.
3. Necessidade de conhecimento para avaliar a recuperação e a adaptação dos casos.
4. Dificuldade em obter casos disponíveis e confiáveis.
5. Os casos dificilmente cobrem todo o domínio.
6. Dificuldade na elaboração de bons algoritmos de adaptação dos casos.
7. Exige bastante espaço para armazenamento dos casos.

### **2.2.4 Fundamentos de Raciocínio Baseado em Casos**

Em Raciocínio Baseado em Casos, novos problemas são abordados lembrando-se problemas antigos similares e avançando a partir daí. Situações são interpretadas comparando-as e contrastando-as com situações anteriores similares. Relatos são entendidos e inferências são feitas encontrando casos mais próximos na memória, comparando e contrastando com esses, fazendo inferências baseadas nessas comparações, e fazendo perguntas quando inferências não podem ser feitas. E o aprendizado acontece como parte do processo de integração de um novo caso na memória (Kolodner, 1993).

O RBC provê as bases para o modelo cognitivo que nos permite construir uma variedade de sistemas automatizados e interativos que podem nos ajudar nas nossas mais diversas tarefas (Kolodner, 1993).

O RBC pode significar adaptar soluções antigas para satisfazer novas demandas; utilizar casos antigos para explicar novas situações; utilizar casos antigos para criticar novas soluções, ou ainda, raciocinar sobre precedentes para interpretar uma nova situação ou criar uma solução equitativa para um novo problema (Kolodner, 1993).

Se olharmos o modo como as pessoas resolvem problemas, observaremos o RBC em uso constantemente. Advogados são ensinados a utilizarem casos como precedentes para construir e justificar argumentos em novos casos. Mediadores e árbitros são ensinados a fazerem o mesmo. Outros profissionais não são ensinados a utilizarem RBC mas frequentemente acreditam que ele provê um modo eficiente de solucionar problemas (Kolodner, 1993).

Casos provêm contexto para discutir questões mais abstratas, e fornecem ilustrações dessas diretrizes abstratas que as pessoas se lembram e aplicam em raciocínios posteriores. RBC também é utilizado extensivamente no raciocínio de senso comum do dia-a-dia. Por exemplo,

quando pedimos uma refeição em um restaurante, frequentemente tomamos a decisão sobre qual refeição poderá ser boa baseados em nossas outras boas experiências neste restaurante. Conforme planejamos nossas atividades, nós lembramos o que funcionou e o que não funcionou anteriormente e utilizamos isso para criar nossos novos planejamentos.

Em geral, a segunda vez que solucionamos algum problema ou realizamos alguma tarefa é mais fácil que a primeira vez pois nos lembramos de nossos erros e tomamos algumas medidas para evitá-los (Kolodner, 1993).

O RBC sugere um modelo de raciocínio que incorpora solução de problemas, entendimento, e aprendizagem e integra tudo com processos de memória.

A qualidade do raciocínio do sistema RBC depende de cinco aspectos (Kolodner, 1993):

1. As experiências que o sistema RBC teve.
2. Sua habilidade em entender novas situações em termos das experiências antigas.
3. Sua capacidade de adaptação.
4. Sua capacidade de avaliar e reparar.
5. Sua habilidade em integrar novas experiências em sua memória apropriadamente.

É importante que um sistema RBC comece com um conjunto representativo de casos. Um conjunto representativo abrange os objetivos e subobjetivos que surgem no raciocínio e ambas as tentativas bem sucedidas e mal sucedidas para alcançar esses objetivos. As tentativas bem sucedidas são utilizadas para propor soluções para o novo problema. As tentativas mal sucedidas são utilizadas para alertar sobre a potencialidade de falha.

A habilidade em entender um novo problema em termos de experiências antigas contém duas partes (Kolodner, 1993):

- Recordando experiências antigas

Chamado de problema de indexação, significa encontrar na memória a experiência mais próxima da nova situação. Pode-se pensar nisso como o problema de associar índices a experiências armazenadas na memória de modo que possam ser recordadas sob circunstâncias apropriadas. Recordar casos apropriadamente esta na essência do RBC.

- Interpretação da nova situação em termos das experiências recordadas

Interpretação é o processo de comparação da nova situação com as experiências recordadas. Quando as situações de problema são interpretadas, são comparadas e contrastadas com as situações de problema antigas. O resultado é uma interpretação da nova situação, a adição do conhecimento inferido sobre a nova situação, ou uma classificação da situação. Quando as soluções para problemas são comparadas com

antigas soluções, o sistema ganha um entendimento dos prós e contras em fazer alguma coisa de um jeito particular. Geralmente há a utilização do processo de interpretação quando problemas não são bem entendidos e quando existe uma necessidade de criticar a solução. Quando o problema é bem conhecido, há uma pequena necessidade do processo interpretativo.

A habilidade de aprender com as experiências é uma das marcas do RBC. Para aprender com as experiências, um sistema RBC requer respostas para então conseguir interpretar o que foi certo e o que foi errado com suas soluções. Sem respostas, o sistema pode ser rápido na solução de problemas mas poderá repetir seus erros e nunca aumentar suas capacidades. Portanto, a avaliação e o consequente reparo são contribuições importantes para a especialidade de um sistema RBC. A avaliação pode ser realizada no contexto das saídas dos outros casos similares, pode ser baseada em respostas, ou pode ser baseada em simulação.

Em Inteligência Artificial, quando se fala em aprendizagem normalmente significa a aprendizagem de generalizações, tanto através de significados indutivos ou baseados em explicação (Kolodner, 1993). Embora a memória de um sistema RBC observe similaridades entre casos e possa, portanto, perceber quando generalizações devem ser formadas, a formação indutiva de generalização é responsável por somente algumas das aprendizagens em um sistema RBC. O RBC realiza a maioria de seus aprendizados de duas outras maneiras: através da acumulação de novos casos ou através da atribuição de índices.

Novos casos dão ao sistema contextos familiares adicionais para solucionar problemas ou avaliar situações. Um sistema RBC cujos casos abrangem mais que o domínio será melhor que um cujos casos abrangem menos que o domínio. Um sistema RBC cujos casos abrangem instâncias de falhas tão bem quando de sucesso será melhor que um outro

cujos casos abrangem somente sucessos. Novos índices permitem ao sistema RBC afinar seu aparato de recordação para que se lembre de casos em vezes mais apropriadas (Kolodner, 1993).

Um sistema RBC que se lembra de suas experiências aprende conforme seus raciocínios; respostas de experiências anteriores lhe dão visão para solucionar problemas posteriores.

Neste contexto, o primeiro aspecto a ser considerado é a definição de um caso.

### **2.2.5 O que é um caso?**

Quando discutimos representação de conhecimento em IA, geralmente pensamos em tornar o conhecimento o mais geral possível para que se possa ter uma ampla aplicação. Os modelos nos dizem como os dispositivos funcionam, e podemos tentar aplicá-los em diversas situações. O objetivo na formulação de modelos é capturar o que é comum através de uma variedade de objetos similares. Embora algumas regras sejam muito específicas, a ideia é formular regras que

são de aplicação geral (Kolodner, 1993).

Certamente o conhecimento geral possui vantagens. Uma delas é a economia de armazenamento que ele permite. Se podemos representar todas as instâncias de um objeto utilizando um modelo, menos terá que ser armazenado. O conhecimento geral também nos permite lidar com a incerteza, onde o modelo nos dá uma maneira de aproximá-la o bastante para que possamos raciocinar sobre a situação sem precisar conhecer todos os detalhes.

Porém, o conhecimento geral também tem suas desvantagens, sendo uma das principais a questão da operacionalização. Apesar do conhecimento geral poder fornecer uma estrutura para o raciocínio, muitas vezes fica dificultada a aplicação de alguma parte do conhecimento geral em uma situação específica. Outra desvantagem do conhecimento geral é que apesar de abranger o normal, ele não nos diz muita coisa sobre como lidar com as situações que são diferentes do normal.

Resumindo, modelos fazem um pressuposto de “mundo fechado” que é bastante difícil de transcender quando o conhecimento geral necessário não está disponível (Kolodner, 1993). Em contraste, um sistema RBC lida com o conhecimento incompleto que tem adicionando casos que descrevem as situações que seus outros casos não consideram.

Os casos, que podem ser de muitas formas e tamanhos, representam um conhecimento específico relacionado à uma situação específica, ou seja, representam o conhecimento no nível operacional (Kolodner, 1993). Eles explicitam como uma tarefa foi efetuada, como uma parte do conhecimento foi aplicado ou quais estratégias foram utilizadas para alcançar um objetivo. Podem representar uma situação que evolui ao longo do tempo, um instante ou qualquer outra fatia de tempo entre estes dois extremos. O comum em todos os casos é que eles representam uma situação vivida. Quando surge uma situação similar a um dos casos armazenados, ao relembrar o caso o conhecimento embutido e as decisões tomadas nos dão um ponto de partida para interpretar a nova situação e resolver os problemas que ela apresenta.

Algumas questões importantes que surgem em relação aos casos são a relação entre eles e o conhecimento geral e se vale a pena armazenar todas as experiências vividas – se não, quais devem ser armazenadas.

Casos que valem a pena ser armazenados parecem ser as experiências que são diferentes do normal ou diferente do que era esperado. As experiências “normais”, aquelas em que tudo foi como o esperado, devem ser unidas na memória como generalizações ou casos compostos. As menores diferenças são combinadas para definir estas generalizações ou casos compostos, e aquelas que nos lembramos distintamente são as que são diferentes (Kolodner, 1993).

Outra questão importante é a sobrecarga do sistema RBC, que pode ocorrer quando são armazenadas as diversas experiências conforme é utilizado e, para se tornar um melhor raciocinador, queremos guardar toda experiência que seja um pouco diferente daquelas que estão na biblioteca de casos. Para evitar a sobrecarga do sistema devemos considerar quais realmente

valem a pena serem lembradas como casos distintos. Saber distinguir quais experiências com algumas diferenças valem a pena serem armazenadas como casos distintos é uma tarefa um pouco complicada. Uma maneira de se fazer isso é verificando se as suas diferenças realmente fazem a diferença (Kolodner, 1993). Se a diferença não causou nenhum efeito prejudicial, não houve necessidade de raciocínio extra e não causou nenhum efeito que não pudesse ser previsto, não há a necessidade de se armazenar esta experiência como um caso separado.

O principal neste ponto é: se a diferença é instrutiva de tal forma que ensina uma lição para o futuro que não poderia ter sido inferida facilmente dos casos já gravados, então grave isso como um caso (Kolodner, 1993). Casos não representam somente diferenças, mas também podem representar normas que foram diferentes do conhecimento prévio no momento em que foram experimentadas. Armazenar casos quando suas diferenças são instrutivas mas não armazenar quando podem ser facilmente inferidas é uma forma de utilizar uma única representação para o conhecimento normativo e conhecimento único.

Alguns princípios importantes de casos são (Kolodner, 1993):

- Um caso é a representação de um conhecimento específico ligado a um contexto, representando o conhecimento no nível operacional.
- Casos associam soluções com problemas, resultados com situações, ou ambos.
- Um caso armazena experiências que são diferentes do que é esperado – somente as que são importantes, isto é, se ensinar uma lição útil.
- Lições úteis são aquelas que têm o potencial de ajudar o sistema RBC a atingir um objetivo ou um conjunto deles com mais facilidade no futuro, ou aquelas que alertam sobre a possibilidade de falhas ou apontam para um problema imprevisto.

Abrangendo toda essa discussão, pode-se definir um caso como sendo um pedaço contextualizado de conhecimento representando uma experiência que ensina uma lição fundamental para alcançar os objetivos do raciocinador (Kolodner, 1993).

Um caso pode ser dividido em duas partes:

- As lições que ele ensina.
- O contexto no qual ele pode ensinar suas lições.

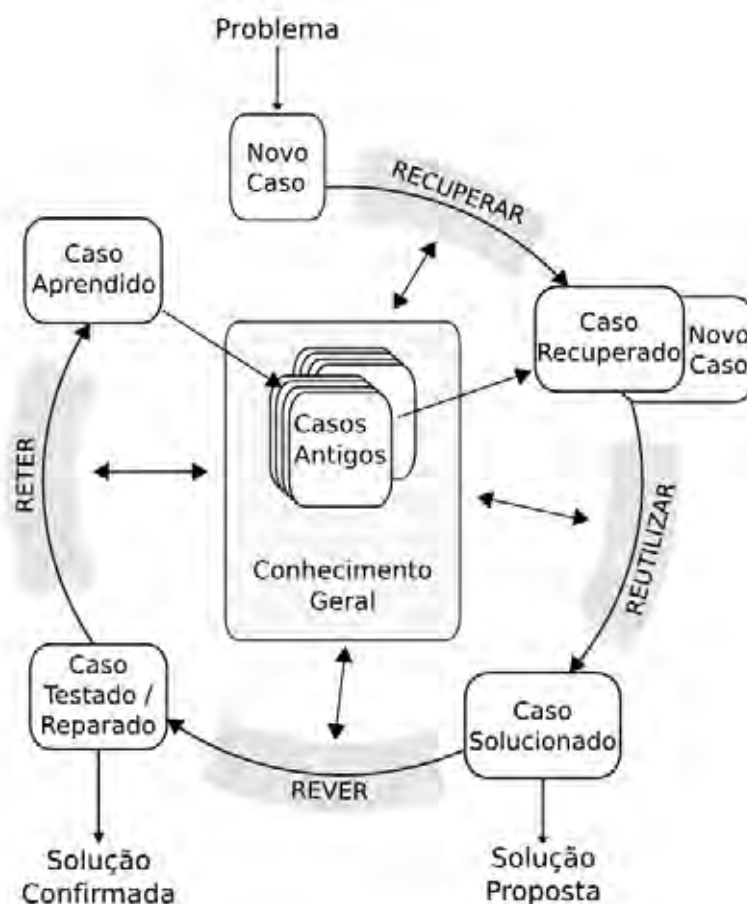
As lições que o caso ensina são o seu conteúdo, e os contextos dos casos são os índices (Kolodner, 1993).

## 2.2.6 O Ciclo do RBC

Segundo Aamodt & Plaza (1994), o ciclo do RBC pode ser representado como um ciclo esquemático (Figura 1) e descrito através dos seguintes processos:

- RECUPERAR os caso(s) mais similar(es).
- REUTILIZAR das informações e conhecimento do(s) caso(s) recuperado(s) para tentar solucionar o problema.
- REVISAR a solução proposta se necessário.
- RETER a nova solução como parte de um novo caso.

Um novo problema é solucionado recuperando um ou mais casos antigos, reutilizando estes casos de alguma maneira, revisando a solução baseado na reutilização de um caso antigo, e retendo (armazenando) a nova experiência, incorporando-a em uma existente base de conhecimento (base de casos).



**Figura 1:** Ciclo do RBC. Fonte: Adaptado de Aamodt & Plaza (1994).

Uma descrição inicial do problema define um novo caso. Este novo caso é utilizado para RECUPERAR casos antigos, que são então combinados com o novo caso – através da

REUTILIZAÇÃO – em uma solução proposta para o problema inicial. Através do processo de REVISÃO, a solução proposta é testada sendo aplicada ao mundo real ou avaliada por um especialista, e ajustada em caso de falhas. Durante o processo de RETENÇÃO, as experiências úteis são retidas para futuras reutilizações e a base de casos é atualizada com um novo caso aprendido, ou com modificações em alguns casos existentes (Aamodt & Plaza, 1994).

A visão de processo representada pela Figura 1 enfatiza o RBC como um ciclo sequencial de passos.

### 2.2.7 Representação dos Casos

Um sistema baseado em casos é fortemente dependente da estrutura e conteúdo de sua coleção de casos – frequentemente referida como sua memória de casos. O problema de representação em RBC é essencialmente o problema de decidir o que armazenar em um caso, encontrar uma estrutura apropriada para descrever o conteúdo do caso, e decidir como a memória de casos deve ser organizada e indexada para uma eficaz recuperação e reutilização (Aamodt & Plaza, 1994).

Talvez uma das mais importantes e comumente utilizada técnica de representação de casos seja a abordagem relacional de banco de dados.

O modelo relacional de banco de dados é simples e flexível, e tem sido bastante utilizado nas aplicações de raciocínio baseado em casos. Cada objeto (ou caso) é representado por uma linha em uma tabela relacional onde as colunas são usadas para definir os atributos dos objetos. Desse modo, podemos dividir uma tabela relacional em duas partes: uma parte da descrição do problema e uma parte da solução.

Seja qual for a forma escolhida de representação dos casos, a coleção dos casos deve ser estruturada de maneira a facilitar a recuperação dos casos apropriados. Diversas abordagens têm sido usadas para indexar os casos para uma recuperação eficiente. Uma estrutura comum utilizada é a base de casos *flat*. Nessa estrutura, os índices são escolhidos para representar os aspectos importantes do caso, e a recuperação envolve a comparação das características importantes do caso com cada caso na base de casos. Outra estrutura comum é a hierárquica, que armazena os casos agrupando-os em categorias apropriadas para reduzir o número de casos que precisam ser procurados durante uma comparação. O modelo de memória para uma forma escolhida de representação irá depender de um número de fatores (Pal & Shiu, 2004):

- A representação usada na base de casos.
- O propósito do sistema RBC. Por exemplo, uma estrutura hierárquica é a escolha natural para um sistema envolvendo problemas de classificação.

- O número e a complexidade dos casos que serão armazenados. Conforme o número de casos armazenados cresce na base de casos, uma estrutura que é buscada sequencialmente irá consumir mais tempo durante a recuperação (por exemplo, uma base de casos *flat*).
- O número de características que são utilizadas para realizar a comparação durante a procura por casos.
- Se alguns casos são semelhantes o suficiente para serem agrupados. Quando os casos se encaixam em agrupamentos naturais, alguma facilidade de estruturação pode ser útil.
- O quanto é conhecido sobre um domínio específico. Isto influencia a habilidade em determinar se os casos são similares. Por exemplo, se pouco conhecimento sobre o domínio está disponível, a estruturação dos casos tende a ser errada.

Concluindo, os casos são assumidos possuindo dois componentes: a especificação do problema e a solução. Normalmente, a especificação do problema consiste de um conjunto de atributos e valores. Os atributos de um caso devem definir a unicidade do caso e devem ser suficientes para predizer a solução para aquele caso. A representação pode ser uma simples estrutura de dados *flat* ou uma complexa hierarquia de objetos (Pal & Shiu, 2004).

### **2.2.7.1 Indexação**

O problema de indexação é um problema central e muito focado em RBC. Se baseia em decidir que tipo de índices utilizar para recuperações futuras e como estruturar o espaço de busca dos índices. Este é um problema de aquisição de conhecimento, e deve ser analisado sendo parte das etapas de análise do conhecimento do domínio e modelagem. Uma solução trivial para o problema é utilizar todas as características de um caso como índices. Esta é a abordagem de métodos baseados em sintaxe com raciocínio baseado em instância e baseado em memória (Aamodt & Plaza, 1994).

A indexação de casos é feita a partir de um conjunto de características que representam um caso. A função da indexação é orientar a avaliação da similaridade dos casos da base. Os índices devem (Delpizzo, 1997):

- Ser preditivos.

- Endereçar os propósitos dos casos que serão usados.
- Ser abstrato o bastante para permitir a ampliação dos casos da base.
- Ser concreto o bastante para serem reconhecidos no futuro.
- Variar com o domínio e com a tarefa do sistema.

Os índices representam uma interpretação da situação, a maneira como alguém pensa sobre determinada situação e a circunstância no qual ela ou ele querem lembrar o fato. Dois aspectos devem ser enfocadas ao tratar os índices: o primeiro é definir o vocabulário e o segundo é como estes índices vão ser valorados. O sucesso do vocabulário de índices é determinado pelo conhecimento do domínio que o especialista possui, portanto este processo pode ser considerado uma representação do conhecimento (Delpizzo, 1997).

A indexação de casos e o armazenamento estão fortemente ligados. A maneira como os casos são armazenados na memória restringe os esquemas de indexação que podem ser utilizados na recuperação. Os principais aspectos que influenciam na organização da base de casos são: conteúdo dos casos; paradigma utilizado na indexação dos casos; e visualização dos casos. Outros dois fatos importantes na definição da indexação dos casos e estrutura são (Gomes, 2003):

- Flexibilidade, pois um caso pode ter que ser recuperado de diferentes modos, dependendo da perspectiva da recuperação.
- Eficiência, devido ao tamanho da base de casos poder impor uma forte limitação computacional.

O processo de indexação consiste em associar um ou mais índices a um caso, definindo situações nas quais o caso é relevante. A seleção do índice é um processo de interpretação baseado no contexto no qual o caso é considerado útil. Há três aspectos a serem considerados no mecanismo de indexação (Gomes, 2003):

- Vocabulário de indexação.
- Processo de indexação.

- Organização do índice .

O índice pode ser considerado um rótulo associado a um caso que estabelece uma relação entre o caso indexado e a situação particular que o índice representa. Os índices são normalmente associados com especificações do problema e podem ter diferentes níveis de abstração. Um índice que pode ser retirado diretamente da especificação da descrição do caso é chamado de índice específico. Índices abstratos são derivados da descrição do caso, mas não correspondem diretamente às características do caso (Gomes, 2003).

Índices específicos são fáceis de serem extraídos dos casos, mas geralmente não são bons índices pois representam aspectos superficiais da situação descrita no caso. Índices abstratos são mais difíceis de serem extraídos, pois a extração envolve um processo de inferência, porém são mais precisos para determinar a relevância do caso.

A indexação de casos pode ser realizada manualmente ou de um modo automático utilizando-se algoritmos de indexação. Existem três principais modos de selecionar índices automaticamente: baseado em listas, baseado em diferenças e baseado em explicações. Na indexação baseada em listas existe uma lista de atributos que são utilizados como índices. A indexação baseada em diferenças utiliza as diferenças entre as descrições dos casos para determinar os índices relevantes. A indexação baseada em explicações extrai os índices utilizando modelos do domínio de trabalho (Gomes, 2003).

### **2.2.8 Recuperação de Casos**

A recuperação de casos é o processo de encontrar, em uma base de casos, aqueles casos que são mais próximos ao caso atual. Este processo começa com uma descrição parcial do problema e acaba quando a melhor escolha de um caso semelhante é encontrada. As subetapas da Recuperação podem ser chamadas de Identificação das Características, Comparação Inicial, Procura e Seleção, executadas nesta ordem (Aamodt & Plaza, 1994).

A tarefa de identificação consiste em uma descrição do problema de modo a “entendê-lo” melhor em um determinado contexto. O objetivo da tarefa da comparação inicial é retornar um conjunto de casos que sejam suficientemente similares ao novo caso através de uma medida de similaridade.

A tarefa de seleção trabalha nesse conjunto de casos, escolhendo o melhor caso, ou ao menos o primeiro caso a ser testado. Nessa etapa, a medida de similaridade define como será calculada a similaridade entre o caso atual e um determinado caso na base de casos. Esta é aplicada repetidamente, par a par, no cálculo da similaridade entre a descrição do problema do caso atual e a descrição de cada caso na base de casos. Feito isso, os casos na base de casos são ordenados de acordo com o seu valor de similaridade e o(s) caso(s) mais similar(es) é(são) sugerido(s) como

solução potencial para o problema presente (Aamodt & Plaza, 1994).

A habilidade de distinguir quais dos vários casos parcialmente similares têm o potencial de serem mais úteis que outros é a chave para fazer com que sistemas de raciocínio baseado em casos funcionem.

A recuperação pode ser decomposta em duas tarefas: a recuperação de casos relevantes da base de casos (chamada de tarefa de recuperação), e o ranqueamento desses casos em relação ao problema alvo. Em geral, os sistemas RBC incorporam essas duas tarefas.

A tarefa de recuperação é normalmente feita utilizando-se os índices, e é basicamente um processo de procura na estrutura de indexação. A recuperação deve ser eficiente e precisa para poupar tempo nas fases subsequentes do raciocínio.

Depois de um conjunto de casos ter sido identificado pela tarefa de recuperação, esses casos são ranqueados em relação ao problema alvo. Normalmente os casos são ranqueados utilizando-se uma métrica de similaridade, a qual avalia o grau de similaridade entre o problema e o caso sendo ranqueado (Gomes, 2003).

Um aspecto importante da recuperação é a correspondência (*matching*) de atributo. Uma simples maneira de realizar a correspondência de atributo quando os casos são representados por pares atributo/valor, é simplesmente corresponder (*match*) atributos com o mesmo nome. Porém mesmo nesta abordagem bastante simples pode haver situações em que ocorra falha, como por exemplo quando é permitido que um atributo apareça mais que uma vez em uma descrição de caso. Esquemas mais elaborados utilizam similaridade semântica de atributos, ou uma correspondência estrutural em uma representação hierárquica ou de grafo, que apresenta uma grande complexidade para a métrica de similaridade (Gomes, 2003).

A etapa de recuperação é a principal área de pesquisa em RBC. As técnicas de recuperação mais comumente investigadas são a *k*-vizinhos mais próximos (*k*-NN – *k*-nearest neighbors), árvores de decisão, e seus derivados. Essas técnicas envolvem desenvolvimento de uma métrica de similaridade que permita verificar a proximidade (isto é, a similaridade) entre os casos a serem medidos.

Algumas considerações importantes, de acordo com Pal & Shiu (2004):

1. *Recuperação através do vizinho mais próximo.* Nesse tipo de recuperação, o caso recuperado é escolhido quando a soma ponderada de suas características que correspondem ao caso atual é maior do que os outros casos na base de casos. Em termos simples, com todas as características ponderadas igualmente, um caso que corresponde ao caso atual, sobre as  $n$  características, será recuperada ao invés de um caso que corresponde a apenas  $k$  características, sendo  $k < n$ . Características que são consideradas mais importantes numa situação de resolução de problema podem ter sua importância explicitada atribuindo a elas um peso maior no processo de correspondência dos casos.

2. *Abordagens indutivas.* Quando abordagens indutivas são utilizadas para determinar a estrutura da base de casos, as quais determinam a importância relativa das características para discriminar entre os casos similares, a estrutura hierárquica resultante da base de caso fornece um espaço de busca reduzido para o recuperados de caso. Isto pode, por sua vez, reduzir o tempo de consulta de pesquisa.
3. *Abordagens guiadas por conhecimento.* Abordagens guiadas por conhecimento para recuperação utilizam o conhecimento de domínio para determinar quais características de um caso são importantes para recuperar o caso no futuro. Em algumas situações, características diferentes de um caso terão diferentes níveis de importância ou de contribuição para os níveis de sucesso associados ao caso. Tal como acontece com as abordagens indutivas para a recuperação, a indexação guiada por conhecimento podem resultar em uma estrutura hierárquica, a qual pode ser mais eficaz para a pesquisa.
4. *Recuperação validada.* Tem havido numerosas tentativas de melhorar a recuperação. Uma dessas tentativas é a recuperação validada, proposta por Simoudis (Pal & Shiu, 2004) que consiste de duas fases. A fase 1 envolve a recuperação de todos os casos que parecem ser relevantes para o problema, baseado nas características principais do caso atual. A fase 2 envolve derivar mais características diferenciadoras do grupo inicial de casos recuperados para determinar se estes casos são válidos na situação atual. A vantagem de recuperação validada é que os métodos de baixo custo computacional podem ser usados para fazer a recuperação inicial da base caso, enquanto os métodos de maior custo computacional podem ser utilizados na segunda fase, onde são aplicados em apenas um subconjunto da base de casos.

Existem vários fatores que devem ser considerados para determinar qual método de recuperação utilizar (Pal & Shiu, 2004):

- O número de casos a serem procurados.
- A quantidade de conhecimento de domínio disponível.
- A facilidade de determinar os pesos para características individuais.
- Se todos os casos devem ser indexados pelas mesmas características ou se cada caso pode ter características que variam em importância.

Normalmente, quando um caso é recuperado, existe uma análise para determinar se o caso é próximo o suficiente ao caso do problema ou se os parâmetros de busca precisam ser modificados e a busca ser realizada novamente. Ao se tomar a decisão correta durante essa análise, pode haver um significativo ganho de tempo no restante do processo.

Se revisarmos os processos envolvidos no RBC apresentados até o momento, podemos representá-los como mostrado na Figura 2:



**Figura 2:** Processos envolvidos no RBC. Fonte: Adaptado de Pal & Shiu (2004)

### 2.2.8.1 Conceito de similaridade

É importante considerar que similaridade pode ter diferentes significados para diferentes pessoas, ou diferentes significados para a mesma pessoa em diferentes momentos (Pal & Shiu, 2004).

O significado de similaridade sempre depende no contexto fundamental de uma aplicação particular, e ela não carrega uma característica fixa que se aplica a qualquer contexto comparativo. Em RBC, a computação da similaridade se tornou uma questão bastante importante no processo de recuperação de casos. A efetividade de uma medida de similaridade é determinada pela utilidade de um caso recuperado em resolver um novo problema. Portanto, estabelecer uma função de similaridade apropriada é uma tentativa de lidar com as relações profundas ou escondidas entre os objetos relevantes associados com os casos.

Uma medida de similaridade pode ser a soma ponderada do grau de similaridade entre os atributos do caso. Um exemplo de uma função de similaridade pode ser representada pela fórmula 2.1, onde  $P$  é o problema alvo,  $C$  é o caso,  $w_i$  são os pesos associados ao atributos (somatório de

$\omega_i = 1$ ), AttrSim é a similaridade do atributo,  $p_i$  é atributo do problema e  $c_i$  é atributo do caso

$$\text{Similaridade}(P, C) = \sum_{i=1}^n \omega_i \cdot \text{AttrSim}(p_i, c_i) \quad (2.1)$$

Os pesos são uma forma de atribuir importâncias diferentes aos atributos. Existem outras medidas de similaridade, como medidas recursivas para representação de caso hierárquica, ou medidas menos sofisticadas, como contar o número de ocorrências de palavras em uma descrição textual de um caso (Gomes, 2003).

Existem duas abordagens principais para o cálculo da similaridade entre os casos. A primeira é baseada na computação de distância entre casos, utilizando o conceito de distância para construir a noção de similaridade, sendo que o caso mais similar é determinado pela avaliação de uma medida de similaridade (isto é, uma métrica). A segunda abordagem está mais relacionada a estruturas de representação / indexação dos casos, onde a estrutura de indexação pode ser percorrida para procurar um caso similar (Pal & Shiu, 2004).

Alguns conceitos e características básicas de medidas de distância utilizadas no cálculo da similaridade são descritas a seguir.

### **Distância Euclidiana ponderada**

De acordo com Pal & Shiu (2004), o tipo mais comum de medida de distância é baseado no local dos objetos no espaço Euclidiano (isto é, um conjunto ordenado de números reais), onde a distância é calculada através da raiz quadrada da soma dos quadrados das diferenças aritméticas entre as coordenadas correspondentes de dois objetos. Mais formalmente, a distância Euclidiana ponderada entre casos pode ser expressa da maneira a seguir. Consideramos  $BC = \{c_1, c_2, \dots, c_N\}$  como sendo uma base de casos tendo  $N$  casos. Cada caso pode ser identificado por um índice das características correspondentes. De maneira mais formal, uma coleção de características  $\{F_j (j = 1, 2, \dots, n)\}$  é utilizada para indexar os casos. O  $i$ -ésimo caso  $c_i$  na base de casos pode ser representado como um vetor  $(n + 1)$ -dimensional, ou seja,  $c_i = (x_{i1}, x_{i2}, \dots, x_{in})$ , onde  $x_{ij}$  corresponde ao valor da característica  $F_j$  ( $1 \leq j \leq n$ ).

Suponha que para cada característica  $F_j$  ( $1 \leq j \leq n$ ), um peso  $w_j$  ( $w_j \in [0,1]$ ) foi atribuído à  $j$ -ésima característica para indicar a sua importância. Sendo assim, para qualquer par de casos  $c_p$  e  $c_q$  na base de casos, uma medida de distância ponderada pode ser definida como:

$$d(c_p, c_q) = \left[ \sum_{j=1}^n w_j^2 (x_{pj} - x_{qj})^2 \right]^{1/2} \quad (2.2)$$

Utilizando a distância ponderada, uma medida de similaridade entre dois casos,  $SM(c_p, c_q)$ , pode ser definida por:

$$SM(c_p, c_q) = \frac{1}{1 + \alpha d(c_p, c_q)} \quad (2.3)$$

onde  $\alpha$  é uma constante positiva. Quanto maior o valor de  $d(c_p, c_q)$ , menor a similaridade entre  $c_p$  e  $c_q$ .

### **Princípio $k$ -Vizinhos mais Próximos ( $k$ -NN – $k$ -Nearest Neighbor)**

O princípio  $k$ -NN é um método de classificação baseado na busca pelos  $k$  casos mais próximos ao caso de entrada atual utilizando uma medida de distância (como a distância Euclidiana) e então selecionar a classe da maioria desses  $k$  casos como a classe recuperada. Em outras palavras, para a classificação do caso de consulta, um fator de ocorrência de cada classe é computado como  $m_i/k$ , onde  $m_i$  é o número de casos entre os casos mais próximos que pertence à classe  $i$ . A classe com o maior fator de ocorrência é então atribuída ao caso de consulta. Normalmente, para melhorar as chances de uma decisão correta para os casos estão próximos ao limite entre duas classes, um limiar  $\beta$  é definido de modo que ao menos  $\beta$  entre os  $k$  vizinhos mais próximos concordem com a classificação (Pal & Shiu, 2004).

## **2.2.9 Reutilização dos casos**

A reutilização de uma solução de um caso no contexto de um novo caso é focada em dois aspectos: (a) nas diferenças entre o caso antigo e o caso corrente e (b) em qual parte do caso recuperado pode ser transferida para o novo caso. As principais subtarefas da Reutilização são a Cópia e a Adaptação (Aamodt & Plaza, 1994).

### **2.2.9.1 Cópia**

Em um processo de Cópia simples, as diferenças entre o novo caso e o caso recuperado são desconsideradas, enquanto as similaridades são consideradas relevantes e a parte da

solução do caso recuperado é transferida ao novo caso como uma parte da solução. Este é um tipo trivial de reutilização. Entretanto, alguns sistemas de RBC levam em conta as diferenças entre o caso atual e os antigos de modo que a parte reutilizada não pode ser diretamente transferida ao novo caso, mas requer um processo de adaptação que leva em conta estas diferenças (Sá, 2007).

#### **2.2.9.2 Adaptação**

De acordo com Kolodner (1993), como nenhuma solução antiga é exatamente a igual à uma nova, devem normalmente ser adaptadas para se tornarem aplicáveis a novas situações.

O processo de adaptação compreende duas etapas: determinar o que precisa ser adaptado e realizar a adaptação. A identificação das partes de um caso recuperado que necessitam de mudança pode ser feita pela comparação entre os resultados obtidos com este caso e os que eram desejados. A diferença obtida nesta comparação irá fornecer estas partes.

Os métodos e estratégias de adaptação podem ser classificados em (Aamodt & Plaza, 1994):

- Substituição: substitui valores apropriados na nova situação a partir de valores da antiga situação;
- Transformação: transforma uma antiga solução em outra adequada a nova situação;
- Reutilização derivacional: reutiliza o método antigo de solução para construir a nova solução
- Outros métodos: usados para realizar alterações estruturais não cobertas por outros métodos.

Métodos de substituição trocam valores ou componentes dos casos recuperados. Em um método de transformação, regras ou procedimentos são utilizados para modificar valores ou componentes do caso. A reutilização derivacional assume que o caso recuperado inclui o método ou procedimento utilizado para gerar a solução do caso. Esse método de geração é aplicado ao problema alvo para gerar uma nova solução que se ajuste ao problema (Gomes, 2003).

A maioria dos métodos de adaptação utiliza o conhecimento de domínio para realizar as modificações na solução do caso. Os conhecimentos de domínio mais utilizados pelos métodos de adaptação são: heurísticas, modelos de domínio, modelos causais, regras e redes semânticas.

#### **2.2.10 Revisão da solução**

Após a etapa de adaptação, a solução adaptada deve ser validada e, se necessário, corrigida. Critérios para a revisão da solução podem ser a qualidade da solução e o quão correta ela esta,

bem como critérios específicos da aplicação em questão, como a facilidade de compreensão da solução por parte do usuário ou a quantidade de esforço para se implementar a solução sugerida. A verificação da solução é, em muitos casos, realizada durante sua aplicação prática, embora possa ser substituída por uma simulação (Aamodt & Plaza, 1994).

Quando uma solução de caso gerada pela fase de reutilização não está correta, surge uma oportunidade para aprender com a falha. Esta fase é chamada de revisão de caso e consiste em duas tarefas: (1) avaliar a solução do caso gerada pela reutilização - se for bem sucedida, aprender com o sucesso (retenção de caso), (2) caso contrário corrigir a solução do caso utilizando conhecimento de domínio específico ou utilizar as informações de entrada (Aamodt & Plaza, 1994).

#### **2.2.10.1 Avaliação de Solução**

A avaliação da solução obtém o resultado da aplicação da solução em um ambiente real (consultando um especialista no assunto ou testando a solução no mundo real). Este é usualmente um passo fora do sistema de RBC, ou seja, necessita a intervenção humana. Os resultados da aplicação da solução podem levar bastante tempo, dependendo da aplicação. O caso ainda pode ser aprendido e ficar disponível na base de casos neste período intermediário, mas deve ser marcado como um caso não-avaliado. A solução ainda pode ser aplicada em um programa de simulação que seja capaz de gerar uma solução correta (Aamodt & Plaza, 1994).

#### **2.2.10.2 Reparação de Falha**

A reparação de falha envolve a detecção de erros, recuperação e/ou geração de explicações para solução corrente. Uma das formas utilizadas para a geração de explicações é denominada técnica do aprendizado baseado em explicações, onde o conhecimento é utilizado para gerar uma explicação do motivo pelo qual determinados objetivos do plano de solução não foram alcançados. Um segundo passo da etapa de revisão utiliza as explicações das falhas para modificar a solução de modo que tais falhas não ocorram no futuro (Aamodt & Plaza, 1994).

#### **2.2.11 Retenção dos casos**

Em um sistema RBC, raciocínio e aprendizado estão intimamente ligados. Toda vez que um problema é resolvido, a nova experiência pode ser retida e integrada na base de casos, tornando-a imediatamente disponível para problemas futuros. Dessa forma, o conhecimento presente em

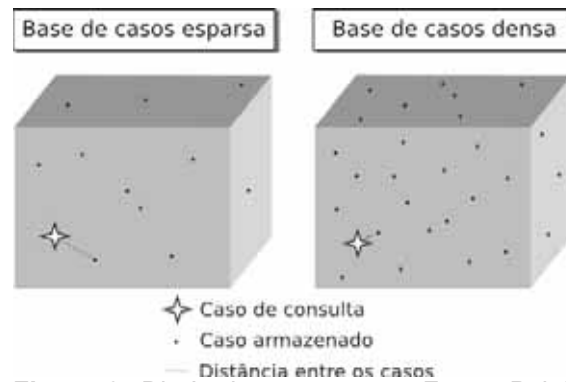
um sistema de RBC é continuamente atualizado, à medida que novas experiências dão origem a novos casos armazenados na base de casos (Sá, 2007).

De acordo Aamodt & Plaza (1994), a etapa de retenção pode ser definida pelo processo de incorporar o que é útil da solução do problema ao conhecimento existente. O aprendizado a partir do sucesso ou da falha da solução proposta é disparado pelo resultado da avaliação e dos possíveis reparos. Isto envolve selecionar qual informação sobre o caso será armazenada, de que forma isso será armazenado, como indexar o caso para uma recuperação posterior e como integrar o novo caso na estrutura da base de casos.

Segundo Aamodt & Plaza (1994), a base de casos RBC é atualizada não importando de qual maneira o problema foi resolvido. Se o problema foi solucionado através do uso de um caso anterior, um novo caso pode ser construído ou o caso antigo pode ser generalizado para ser incluído no caso presente. Se o problema foi solucionado através de outros métodos, incluindo com interação do usuário, um caso completamente novo terá que ser construído. De qualquer maneira, uma decisão deve ser feita sobre o que usar como fonte de aprendizagem. As informações relevantes do problema e as soluções são candidatos óbvios. Porém, uma explicação ou outra forma de justificativa do porquê uma solução é uma solução para o problema também pode ser marcada para inclusão em um novo caso. Um outro tipo de estrutura que pode ser extraído para aprendizagem é o método de solução do problema, ou seja, o caminho de raciocínio estratégico, tornando o sistema adequado para reutilização derivacional.

As falhas podem também ser extraídas e armazenadas, tanto como casos de falha separados ou como casos totalmente problemáticos. Quando uma falha é encontrada, o sistema pode então se lembrar de uma falha anterior similar, e utilizar o caso de falha para melhorar sua compreensão da falha atual e corrigi-la (Aamodt & Plaza, 1994).

De acordo com Pal & Shiu (2004), o aprendizado pode ocorrer de duas maneiras. A adição de um novo problema, sua solução, e o resultado na base de caso é um método comum. A adição de casos na base de casos irá aumentar o alcance das situações cobertas pelos casos armazenados e reduzir a distância média entre o vetor de entrada e o vetor armazenado mais próximo (veja Figura 3).



**Figura 3:** Distância entre casos. Fonte: Pal & Shiu (2004)

Um segundo método de aprendizado em sistemas RBC é utilizar a avaliação da solução para modificar os índices dos casos armazenados ou para modificar o critério para a recuperação do caso. Se um caso tem índices que não são relevantes aos contextos específicos em que devem ser recuperados, ajustar os índices pode aumentar a correlação entre as ocasiões em que um caso é de fato recuperado e as ocasiões em que deveria ter sido recuperado. Da mesma forma, a avaliação do desempenho de uma solução pode levar a uma melhor compreensão do base do modelo causal do domínio que pode ser usado para melhorar o processo de adaptação. Se melhores maneiras podem ser encontradas para modificar os casos com relação à distância entre os casos atuais e os recuperados, a solução fornecida provavelmente será melhorada.

Quando o aprendizado envolve a adição de novos casos na base de casos, há algumas considerações (Pal & Shiu, 2004):

1. *Em quais situações um caso deve ser adicionado à base de casos, e quais situações ele deve ser desconsiderado?* Para determinar a resposta, devemos considerar o nível de sucesso da solução, o quão similar é aos outros casos na base de casos, e se existem importantes lições a serem aprendidas com o caso.
2. *Se um caso deve ser adicionado à base de casos, os índices do novo caso devem ser determinados e como esse caso deve ser adicionado à base de casos.* Se a estrutura da base de casos e o método de recuperação são fortemente estruturados (isto é, uma estrutura hierárquica determinada indutivamente ou um conjunto de redes neurais), a incorporação de um novo caso pode requerer planejamento significativo e reestruturação da base de casos.

## 2.3 Teoria dos Conjuntos Nebulosos

A Teoria dos Conjunto Nebulosos, proposta por Zadeh (1965), foi motivada pela convicção de que os métodos numéricos de análise eram inadequados para tratar sistemas onde a razão humana envolve o uso de variáveis que assumem termos linguísticos. Na proposta de Zadeh (1965), as variáveis assumem termos linguísticos que são representados como conjuntos nebulosos. Como exemplo, podemos citar termos bastante utilizados pela linguagem humana: alto, pesado, quente, muito frio, muito longe, etc.

A teoria de conjuntos nebulosos admite a possibilidade de uma pertinência parcial, diferente da noção clássica, onde há uma clara distinção dos elementos que pertencem e os que não pertencem a um dado conjunto. Logo, há uma generalização da noção clássica onde um elemento pode pertencer a um dado conjunto com um determinado grau de pertinência (Mendes, 2001).

Ela tem sido largamente utilizada nas diversas áreas da Computação e da Engenharia. No contexto do RBC, quando atributos quantitativos são utilizados para criar índices, envolve conversão de atributos numéricos em termos qualitativos para indexação e recuperação. Esses termos qualitativos podem ser representados como conjuntos nebulosos. Além disso, uma das principais questões na teoria dos conjuntos nebulosos é a medição de similaridade. A idéia de medida de similaridade em RBC é inerentemente nebulosa por natureza. Por exemplo, a distância Euclidiana entre dois atributos é comumente utilizada para representar a similaridade entre os casos. Entretanto, o uso da teoria dos conjuntos nebulosos para indexação e recuperação tem muitas vantagens sobre tais medidas utilizadas na teoria dos conjuntos convencionais (*crisp*):

- Atributos numéricos podem ser convertidos para termos nebulosos para simplificar a comparação.
- Conjuntos nebulosos permitem a indexação múltipla de um caso sobre um único atributo com diferentes graus de pertinência.
- Conjuntos nebulosos tornam mais fácil a transferência de conhecimento entre domínios.
- Conjuntos nebulosos permitem a utilização de termos modificadores para aumentar a flexibilidade na recuperação de casos.

### 2.3.1 Definição de conjunto nebuloso

Formalmente um conjunto nebuloso é definido como:

**Definição:** um conjunto nebuloso  $A$  é caracterizado por uma função de pertinência  $\mu_A(x)$  que mapeia os elementos  $x$  de um espaço  $X$  para um intervalo unitário  $[0,1]$ , (Zadeh, 1965). Isto é:

$$\mu_A(x): X \rightarrow [0,1] \quad (2.7)$$

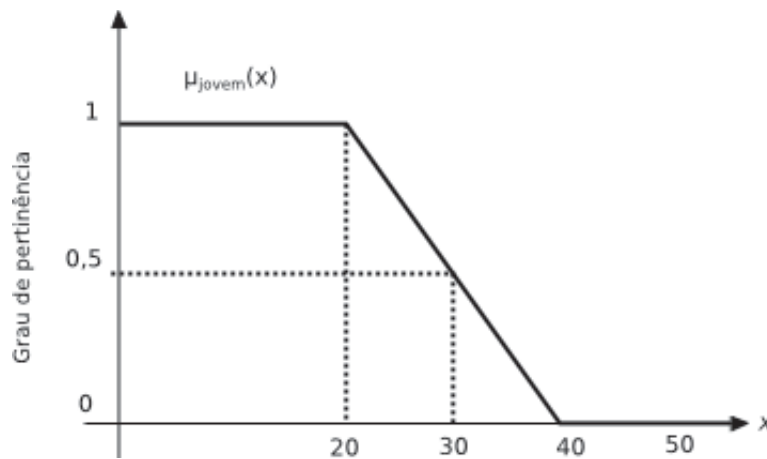
O valor de  $\mu_A(x)$  fornece o grau de pertinência - muitas vezes também chamado de grau de compatibilidade ou grau de verdade - de um elemento  $x \in X$  com as propriedades de  $A$ . Esse conjunto nebuloso  $A$  pode ser representado por um conjunto de pares ordenados de elementos  $x \in X$  e seus graus de pertinência, ou seja (Mendes, 2001):

$$A = \{(x, \mu_A(x)) \mid x \in X\} \quad (2.8)$$

Para exemplificar, consideremos um conjunto de pessoas jovens utilizando um conjunto nebuloso. De maneira geral, se uma pessoa tem menos que 20 anos de idade, ela será considerada “jovem”. O conjunto nebuloso “jovem”, denotado por  $A$ , pode ser caracterizado pelas seguintes equações (Pal & Shiu, 2004):

$$\mu_{jovem} = \begin{cases} 1 & x < 20 \\ \frac{-1}{20}x + 2 & 20 \leq x < 40 \\ 0 & x \geq 40 \end{cases}$$

onde  $x$  é um elemento de  $X = \{0, 1, \dots, 39\}$ .



**Figura 4:** Função de pertinência do conjunto nebuloso "jovem" (Pal & Shiu, 2004)

A partir da Figura 4, podemos verificar que pessoas no grupo de idade  $[0,20]$  (isto é,  $0 \leq x \leq 20$ ) são membros absolutos do conjunto “jovem” (têm um grau de pertinência de 1). Pessoas no grupo

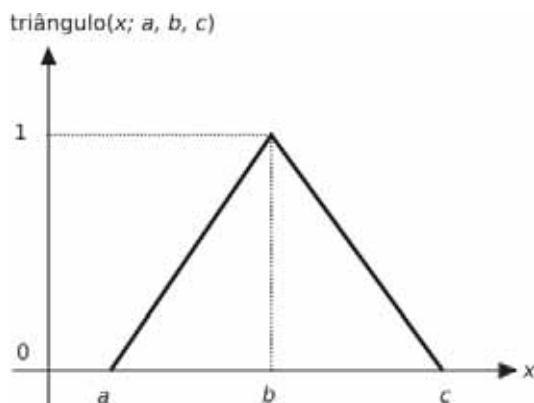
[20,30] são membros parcial do conjunto “jovem” (por exemplo, uma pessoa de 30 anos é “jovem” com um grau de pertinência de 0,5). Pessoas acima de 40 anos não são membros do conjunto “jovem”. É importante notar que existem diversas funções de pertinência que podem ser usadas para caracterizar o conjunto nebuloso “jovem”.

### 2.3.2 Funções de Pertinência

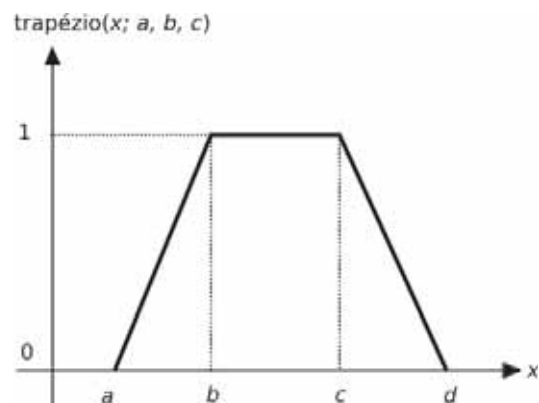
Segundo Pal & Shiu (2004), uma função de pertinência especifica o grau de pertinência de um elemento em um conjunto nebuloso. A determinação de funções de pertinência é naturalmente subjetiva e dependente de contexto. Algumas funções de pertinência populares são a função de pertinência triangular, trapezoidal, função de pertinência sigmóide Gaussiana, função S, e a função  $\pi$ . As mais simples são descritas a seguir.

- *Função de pertinência Triangular.* Essa função de pertinência (veja na Figura 5) é definida como

$$\text{triângulo}(x; a, b, c) = \max \left\{ \min \left\{ \frac{x-a}{b-a}, \frac{c-x}{c-b} \right\}, 0 \right\}$$



**Figura 5:** Função de pertinência triangular



**Figura 6:** Função de pertinência trapezoidal

- *Função de pertinência trapezoidal.* Essa função de pertinência (veja Figura 6) é definida como

$$\text{trapézio}(x; a, b, c, d) = \max \left\{ \min \left\{ \frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right\}, 0 \right\}$$

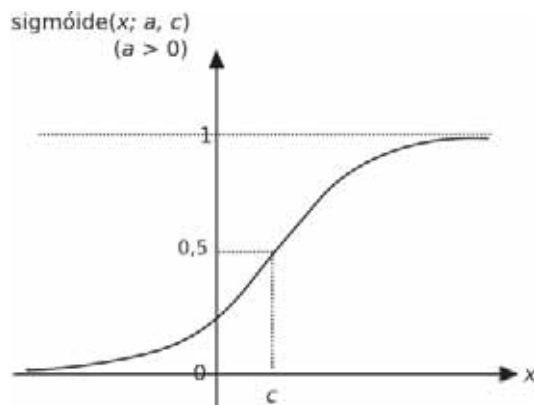
Note que a função de pertinência triangular é uma forma especial da função de pertinência

trapezoidal.

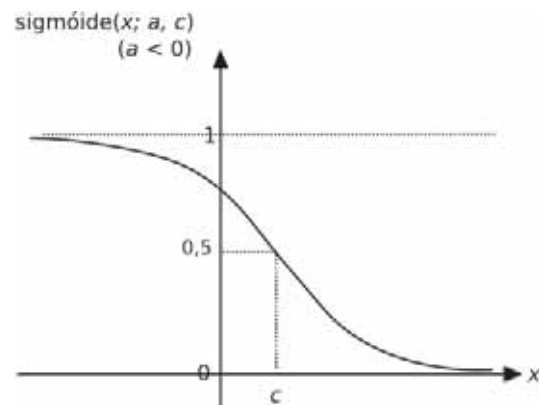
- *Função de pertinência sigmóide.* Essa função de pertinência (veja Figuras 7 e 8) é definida como

$$\text{sigmóide}(x; a, c) = \frac{1}{1 + e^{-a(x-c)}}$$

Das Figuras 7 e 8, se o parâmetro  $a > 0$ , essa função é contínua e aberta no lado direito, enquanto se  $a < 0$ , essa função é contínua e aberta no lado esquerdo. Essa função é utilizada para descrever conceitos tais como “grande” ou “pequeno”.



**Figura 7:** Função de pertinência sigmóide ( $a > 0$ )



**Figura 8:** Função de pertinência sigmóide ( $a < 0$ )

Embora as funções de pertinência sejam subjetivas, elas não podem ser definidas arbitrariamente. Por exemplo, um conjunto nebuloso “inteiros em volta de  $b$ ” pode ser representada por alguma função como na Figura 5, na qual o valor de pertinência primeiro aumenta uniformemente até  $x = b$ , onde alcança o valor máximo, e então começa a diminuir; entretanto esse conjunto nebuloso não pode ser representado por uma função como as das Figuras 7 ou 8.

### 2.3.3 Variáveis Linguísticas

Kaufmann & Gupta (1988), afirmam que o conhecimento ou o significado de certos fatos empíricos nem sempre se comportam de maneira probabilística ou mesmo determinística, envolvendo muitas vezes alguma forma de aproximação ou raciocínio aproximado.

A frase a seguir ilustra esta idéia: *O investimento inicial para possibilitar a implementação completa do empreendimento é extremamente elevado.* Os conceitos envolvidos no significado

das palavras **completa** e **elevado**, vão muito além do que os aspectos determinísticos ou probabilísticos podem analisar. Assim fica clara a necessidade de elaboração de um sistema que permita analisar tais dados de maneira a permitir que estas informações sejam levadas em consideração durante a avaliação do problema.

Uma variável linguística é uma variável cujos valores são palavras (termos linguísticos) ou sentenças em uma linguagem natural ou sintética. Como exemplo, para uma variável linguística como “idade”, seus valores podem ser “muito jovem”, “jovem”, “meia idade”, “velho”, “muito velho”, e assim por diante. Uma variável linguística pode ser interpretada como um rótulo de um conjunto nebuloso que é caracterizado por uma função de pertinência, e sua definição formal é dada a seguir (Pal & Shiu, 2004).

**Definição:** uma *variável linguística* é caracterizada pela quintupla  $(X, T, U, g, s)$ , na qual  $X$  é o nome da variável cujos valores variam sobre o conjunto universal  $U$ ,  $T$  é um conjunto finito de termos linguísticos  $\{t_0, t_1, \dots, t_n\}$  que descreve  $X$ ,  $g$  é uma gramática para termos linguísticos, e  $s$  é uma regra semântica que mapeia cada termo  $t \in T$  para o conjunto nebuloso (isto é,  $s(t)$ ) em  $U$  (Pal & Shiu, 2004).

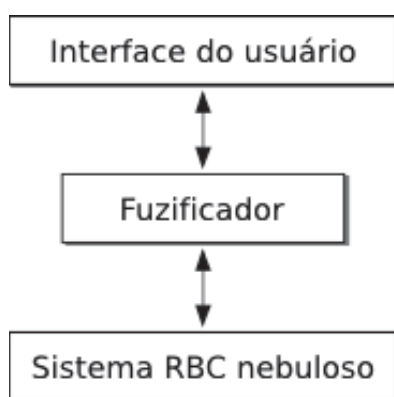
Por exemplo, para descrever a idade das pessoas, a variável linguística “idade” pode ser utilizada. Da definição de variável linguística, temos  $X = \text{idade}$  e  $U$  consiste dos possíveis valores de idade das pessoas (por exemplo,  $U = \{0, 1, 2, \dots, 110\}$ ). Esses valores são descritos por um conjunto pré-ordenado  $T$ , tal como  $T = \{\text{jovem, meia idade, velho}\}$ . A gramática  $g$  é utilizada para gerar termos adicionais, como “muito jovem”, e “muito velho”. A regra semântica  $s$  mapeia cada termo, tal como “jovem” para o conjunto nebuloso (ou seja,  $s(\text{jovem})$ ).

#### 2.3.4 Características nebulosas

As características de um caso podem ser *fuzificadas* em termos linguísticos nebulosos e números nebulosos. Um número nebuloso é utilizado para lidar com quantidades inexatas e é formulado por uma expressão linguística que reflete o quão perto está de um número real, por exemplo, “por volta de 100” ou “mais ou menos 100” (Pal & Shiu, 2004).

A *fuzificação* é primeiro passo do sistema lógico nebuloso, e corresponde à transformação dos dados de entrada iniciais em suas respectivas variáveis linguísticas. Nesta etapa, todas as informações relativas à imprecisão ou incerteza associada a estas variáveis devem ser consideradas. É fundamental e necessário que especialistas da área estudada sejam consultados durante a atribuição de valores relacionados aos graus de pertinência para cada uma das variáveis em estudo, contribuindo assim para maior precisão nos resultados.

O sistema RBC deve definir quais atributos são passíveis de serem *fuzificados*. Geralmente esses atributos são os que o usuário tem dificuldade para especificar com exatidão o seu valor, por exemplo o preço de um apartamento que varia de acordo com o tamanho, localização, quantidade de cômodos. Nessa situação, um sistema RBC nebuloso poderia aceitar uma consulta como “Estou procurando um apartamento que tenha 3 quartos, tenha um preço moderado e esteja localizado em uma região não movimentada”. A arquitetura de um sistema RBC nebuloso é mostrada na Figura 9. O *fuzificador* consiste de um conjunto de algoritmos de *fuzificação* (e de *defuzificação*) que são utilizados para gerar as entradas nebulosas a partir da entrada comum do usuário. O sistema RBC nebuloso contém funções de similaridade nebulosas para comparar de consulta com os casos na base de casos.



**Figura 9:** RBC nebuloso (Pal & Shiu, 2004)

### 2.3.5 Similaridade nebulosa

Os conceitos e formas da similaridade tradicional, como a distância Euclidiana ponderada apresentada na seção 2.2.8.1, podem somente lidar com características com valores reais ou valores específicos (Pal & Shiu, 2004). Entretanto, nas situações reais, características do caso são frequentemente especificadas de maneira incompleta ou com incertezas. Como exemplo, uma das características dos casos em um sistema RBC pode ser descrita por termos linguísticos tais como *baixo*, *médio*, *alto*. Assim, para a implementar o processo de correspondência dos casos e recuperação, é necessário definir uma métrica de similaridade apropriada. O conceito de conjunto nebuloso provê uma boa ferramenta para lidar com o problema de um modo natural. Na teoria dos conjuntos nebuloso, podemos considerar o termo linguístico como um número nebuloso, que é um tipo de conjunto nebuloso. Então uma função de pertinência é determinada em relação ao dado termo linguístico. Quando um valor real da característica de um dado problema é inserido, o valor correspondente de pertinência dos diferentes termos linguísticos são obtidos através das funções

de pertinência. Isto é, depois da correspondência aproximada, as características com valores reais são transformadas em características linguísticas. Então, para se escolher o conjunto de casos com melhor correspondência, é necessário definir algumas medidas de similaridade e algoritmos para computar a similaridade nebulosa, que serão apresentados na seção 3, de acordo com a abordagem proposta.

### 2.3.6 Geometria dos conjuntos nebulosos

Após a proposição da sua teoria dos conjuntos nebulosos, Zadeh introduziu uma interpretação geométrica dos *conjuntos nebulosos como pontos*, observando que, uma vez que as funções de pertinência fazem o mapeamento de elementos para um intervalo unitário, essas funções poderiam ser representadas por ponto em hipercubos unitários (Kosko, 1992; Sadegh-Zadeh<sup>(a)</sup>, 1999).

Para a completa compreensão, primeiramente um pouco de terminologia.

O conjunto  $X$  com  $n$  elementos tem  $2^n$  subconjuntos. O conjunto desses subconjunto é referido como o conjunto potência (*power set*) de  $X$  e denotado por  $2^X$ . Para exemplificar, se  $X = \{x_1, x_2\}$ , teremos o conjunto potência  $2^X = \{\emptyset, X, \{x_1\}, \{x_2\}\}$ .

O conjunto potência nebuloso (*fuzzy power set*), denotado por  $F(2^X)$ , é o conjunto contendo todos os subconjuntos nebulosos de  $X$ . Este conjunto é infinito e também contém todos os elementos de  $2^X$ . Importante notar que  $F(2^X)$  não é um conjunto nebuloso (Sadegh-Zadeh<sup>(a)</sup>, 1999).

#### Hipercubo unitário

O produto cartesiano de um intervalo fechado é chamado de *retângulo fechado* ou simplesmente *caixa*. Assim, se  $a_i < b_i$ , temos (Sadegh-Zadeh<sup>(a)</sup>, 1999):

$$[a_1, b_1] \times \dots \times [a_n, b_n] = \text{uma caixa em } R^n$$

com  $n \geq 1$ . Uma *caixa* em  $R^n$  é um cubo se  $a_1 = a_2 = \dots = a_n = 0$  e  $b_1 = \dots = b_n$ . Se ainda  $b_1 = \dots = b_n = 1$ , o cubo é um cubo unitário em  $R^n$ , ou seja:

$$[0, 1] \times \dots \times [0, 1]$$

chamado de *hipercubo unitário*  $[0, 1]^n$ . Esse hipercubo  $[0, 1]^n$  é

|                                  |               |
|----------------------------------|---------------|
| a linha unitária entre 0 e 1     | se $n = 1$ ;  |
| o quadrado unitário              | se $n = 2$ ;  |
| o cubo unitário comum            | se $n = 3$ ;  |
| o cubo unitário $n$ -dimensional | se $n \geq 1$ |

### Hipercubo unitário nebuloso

De acordo com Sadegh-Zadeh<sup>(a)</sup> (1999), se  $\{(x_1, a_1), \dots, (x_n, a_n)\}$  é um conjunto nebuloso, a  $n$ -upla  $\langle a_1, \dots, a_n \rangle$  é referida como o vetor de pertinência, onde  $a_i$  é o grau de pertinência de  $x_i$ .

Portanto, podemos utilizar para qualquer conjunto nebuloso  $A = \{(x_1, a_1), \dots, (x_n, a_n)\}$ , com  $n \geq 1$ , a notação vetorial e representá-lo pelo seu vetor de pertinência  $n$ -dimensional  $(a_1, \dots, a_n)$  com seus componentes em  $[0, 1]$ . Por exemplo, se temos o conjunto  $X = \{x_1, x_2, x_3\}$ , utilizamos a notação:

$(a_1, a_2, a_3)$  para o conjunto nebuloso  $\{(x_1, a_1), (x_2, a_2), (x_3, a_3)\}$ .

O  $i$ -ésimo componente  $a_i$ , com  $i \geq 1$ , no vetor do conjunto nebuloso  $(a_1, \dots, a_n)$ , representa o grau de pertinência  $\mu_A(x_i) = a_i$  do correspondente elemento  $x_i$ .

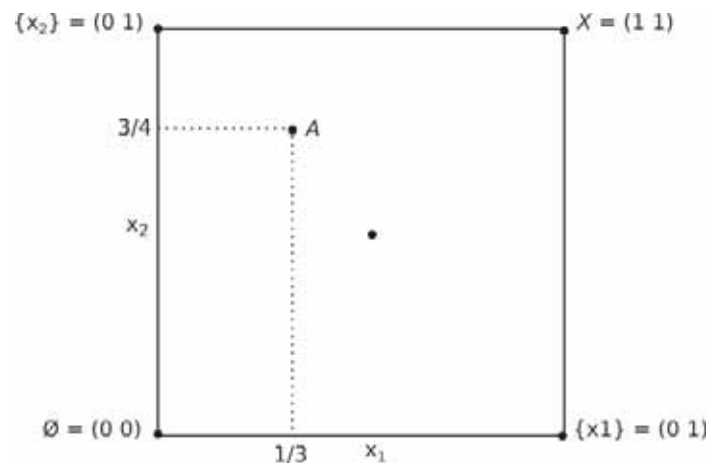
Uma função de pertinência  $\mu_A$ , portanto, define um conjunto nebuloso como um vetor  $n$ -dimensional  $A = (\mu_A(x_1), \dots, \mu_A(x_n))$  com  $\mu_A(x_i) \in [0, 1]$ . Levando em consideração que geometricamente um vetor  $n$ -dimensional  $(a_1, \dots, a_n)$  de números reais define

|                          |                 |
|--------------------------|-----------------|
| um ponto em uma linha    | se $n = 1$ ;    |
| um ponto em um retângulo | se $n = 2$ ;    |
| um ponto em um cubo      | se $n = 3$ ;    |
| um ponto em um hipercubo | se $n \geq 3$ , |

atingimos a ideia de representação geométrica de Zadeh mencionada anteriormente: Um conjunto nebuloso  $\{(x_1, a_1), \dots, (x_n, a_n)\}$  como um vetor  $n$ -dimensional  $(a_1, \dots, a_n)$ , com seus componente em  $[0, 1]$ , é um ponto em um hipercubo unitário  $n$ -dimensional  $I^n = [0, 1]^n$ , denominado como *hipercubo nebuloso* (Sadegh-Zadeh<sup>(a)</sup>, 1999).

Como exemplo, consideremos o conjunto  $X = \{x_1, x_2\}$ . Um conjunto nebuloso  $A$ , poderia ser representado por seu vetor de pertinência  $(1/3, 3/4)$ , significando que o elemento  $x_1$  pertence pouco ao conjunto  $A$  – com grau de pertinência  $1/3$  – enquanto que o elemento  $x_2$  pertence um

pouco mais – grau pertinência de  $3/4$ . Na Figura 10 podemos observar a visão de conjunto nebuloso como um ponto, que geometricamente representa o conjunto nebuloso A como um ponto no hipercubo nebuloso  $I^2$ .



**Figura 10:** Conjuntos como pontos. O conjunto A é um ponto no 2-cubo unitário com coordenadas ou pertinência  $(1/3 \ 3/4)$  (Kosko, 1992).

## 2.4 Raciocínio Baseado em Casos Flexível

De acordo com Zadeh (1994), computação flexível (*soft computing*) é “uma abordagem emergente de computação, que se assemelha à notável capacidade da mente humana de raciocinar e aprender em um ambiente de incerteza e imprecisão”. De maneira geral, consiste na associação de ferramentas e técnicas computacionais, como lógica nebulosa, teoria das redes neurais, computação evolutiva, raciocínio probabilístico, entre outras, para abordar os problemas que envolvem incertezas e imprecisões onde a utilização das técnicas clássicas apresentam alguma dificuldade (Pal & Shiu, 2004).

Na computação flexível, a associação das técnicas permite que elas atuem de maneira independente ou de maneira colaborativa, não competitiva. Quando utilizadas de maneira colaborativa podem melhorar o domínio de aplicação umas das outras, integrando os seus méritos individuais (por exemplo, a capacidade de manipulação de incerteza de conjuntos nebulosos, capacidade de aprendizagem de redes neurais artificiais, e as características robustas de busca e otimização dos algoritmos genéticos). O principal objetivo desta área é criar sistemas de processamento de informação flexíveis que podem se utilizar de uma tolerância para a imprecisão, a incerteza, o raciocínio aproximado, e a verdade parcial, a fim de alcançar tratabilidade, robustez, solução de baixo custo, e uma maior semelhança com tomada de decisão humana (Pal & Shiu, 2004).

### 3 Abordagem Proposta

Nesta seção, considerando principalmente as etapas de representação, indexação, e recuperação dos casos do ciclo do RBC, será realizada a definição dos métodos e técnicas utilizados visando a construção de um sistema RBC nebuloso para recomendação de sistemas marítimos de produção de petróleo.

De maneira geral, a construção de um sistema RBC envolve a realização de algumas atividades, de maneira ordenada, a fim de atender às principais etapas do ciclo do RBC.

Baseadas nos conceitos apresentados na seção 2, essas atividades são descritas a seguir:

1. *Aquisição do conhecimento do domínio da aplicação.* Inicialmente na concepção e criação de um sistema RBC é feita a definição do domínio da aplicação, o levantamento dos principais aspectos do domínio e suas características. A aquisição do conhecimento poderá ser feita junto a especialistas do domínio.
2. *Definição do modelo de representação dos casos.* Uma vez que o conhecimento do domínio da aplicação está disponível, é feita a definição da estrutura apropriada para representar e descrever os casos e de como deve ser organizada a base de casos. O modelo de representação dos casos deve ser pautado pela tentativa de garantir a eficiente recuperação e reutilização dos casos.
3. *Definição e tratamento dos índices.* A definição dos índices que serão utilizados para a realização da recuperação dos casos e o tratamento a eles dados tem um grande impacto na eficiência do sistema RBC como um todo. Neste sentido, os índices devem ser definidos a partir do conjunto de atributos que melhor representam um caso. Tendo em vista que alguns índices podem ser melhor descritos como valores linguísticos, a teoria dos conjuntos nebulosos pode ser aplicada para o tratamento desses atributos.
4. *Definição do método de recuperação.* Considerando que a principal tarefa da etapa de recuperação é a seleção de um conjunto de caso similares ao problema, a escolha do método de recuperação deve ser baseada na eficiência e capacidade de reconhecimento de similaridade entre os casos. Existem diversas técnicas e medidas de similaridades que podem ser utilizadas, sendo que uma boa parte delas utiliza o conceito de distancia.

5. *Definição da estratégia de adaptação da solução.* Após a recuperação dos casos mais similares ao problema, suas soluções devem ser adaptadas ao novo problema. Diversas técnicas de adaptação podem ser utilizadas, tais como as técnicas tradicionais da reinstânciação e da transformação, bem como as técnicas de aprendizagem de máquina. A escolha da estratégia de adaptação deve considerar a relação entre a eficiência e a complexidade de desenvolvimento.
6. *Revisão da solução proposta.* Para a utilização da solução adaptada, uma validação deve ser realizada a fim de corrigir eventuais erros ou violação de restrições. A verificação da solução pode ser feita através da sua aplicação na prática, da consulta de um especialista com o conhecimento e experiência do domínio ou através de simulação.
7. *Definição do método de retenção.* A adição de novos casos na base de casos pode ser entendida como uma estratégia de aprendizado, aumentando assim a quantidade de situações coberta pelos casos armazenados. O processo de adição dos novos casos à base de casos, bem como a manutenção da base, devem ser definidos levando em consideração algumas tarefas envolvidas, tais como garantir a consistência e precisão do sistema, garantir a eficiência da resolução do problema e controle o tamanho da base de casos.

Nas próximas seções, essas atividades, seus métodos e estratégias adotados, serão detalhados para a compreensão e utilização na construção do sistema RBC nebuloso para recomendação de sistemas marítimos de produção de petróleo.

### 3.1 Representação

Como apresentado na seção 2.2.7, a abordagem relacional de banco de dado para a representação de casos é interessante e bastante utilizada. Essa abordagem tem como principais vantagens a facilidade de integrar dados disponíveis em outros formatos, a simplicidade da utilização por profissionais da área de Computação e a flexibilidade. Entretanto, uma desvantagem é que os *designers* da base de casos devem gastar um bom tempo e energia para desenvolver um modelo relacional que representa de maneira adequada os casos.

No modelo relacional de banco de dados cada objeto (ou caso) é representado por uma linha em uma tabela relacional onde as colunas são usadas para definir os atributos dos objetos. Desse modo, podemos dividir uma tabela relacional em duas partes: uma parte da descrição do problema e uma parte da solução.

Para ilustrar esse método de representação, consideremos o exemplo de compra de um imóvel. Para vender o imóvel, o vendedor precisa conhecer todas as informações relevantes que possam interessar ao comprador, como por exemplo, tipo do imóvel, tamanho, quantidade de cômodos, localização. Outras informações relacionadas ao comprador, como o orçamento disponível, preferências, também podem ser consideradas. O modelo de representação do exemplo de compra de imóvel pode ser visualizado na Figura 11.



**Figura 11:** Exemplo de modelo de compra de imóvel.

As tabelas do modelo relacional, que representam o domínio da aplicação, são relacionadas através de chaves primárias e estrangeiras. Através de instruções SQL (*structured query language*) podemos criar essas tabelas relacionais, inserir os casos e recuperá-los através de consultas SQL. Desse modo, os casos não necessariamente devem ser representados por uma única tabela relacional. Os casos podem ter suas propriedades representadas por um conjunto de tabelas que descrevem entidades do domínio e que se relacionam.

O caso para esse modelo contém informações do comprador, do interesse do comprador e informações do imóvel. Um típico registro desse caso é mostrado na Figura 12.

|                 |            |              |           |
|-----------------|------------|--------------|-----------|
| Nome Comprador: | Eduardo    | Tamanho:     | Médio     |
| ID Comprador:   | 123        | Cômodos:     | 5         |
| Idade:          | 38         | Quartos:     | 3         |
|                 |            | Banheiros:   | 2         |
| Preço:          | 100.000,00 |              |           |
|                 |            | Recomendação |           |
| Cidade:         | Rio Claro  | Tipo Imóvel: | Casa      |
| Movimentação:   | Moderada   | Bairro:      | Santana   |
|                 |            | Venda:       | Realizada |

**Figura 12:** Registro de caso de compra de imóvel.

### 3.2 Indexação

A indexação de casos é feita a partir de um conjunto de características que representam um caso. A função da indexação é orientar a avaliação da similaridade dos casos da base.

Talvez o maior problema do desenvolvimento de sistemas baseados em casos seja a indexação e a recuperação apropriada dos casos (Kolodner, 1993). Para minimizar esse problema, a escolha de atributos que melhor representem o contexto da aplicação deve ser realizada com cuidado e com o conhecimento específico do domínio, geralmente feita baseando-se nas experiências de especialistas do domínio da aplicação. A escolha desses atributos tem um grande impacto na eficiência da etapa de recuperação, e consequentemente no sistema como um todo.

Em algumas aplicações os atributos são melhor descritos como valores não numéricos, tais como palavras e sentenças, assumindo características qualitativas. Essa descrição linguística é menos específica que a numérica, porém, permite agregar informações de uma maneira mais apropriada para o raciocínio diante da possível complexidade do domínio da aplicação. Nestas situações, o conceito de conjuntos nebulosos aplicado na modelagem dos atributos pode facilitar e simplificar a etapa de recuperação (Mendes *et al.*, 2001).

Uma vez definidos os atributos para a indexação e tendo em vista o tratamento dos atributos através da Teoria de Conjuntos Nebulosos, apresentada na seção 2.3, a base de casos pode ser construída como um espaço nebuloso (Mendes, 2001), utilizando a interpretação geométrica apresentada na seção 2.3.6.

Considere uma base de casos na qual os casos são indexados por um conjunto de atributos  $\{atributo_1, atributo_2, \dots, atributo_N\}$ , os quais possuem valores numéricos. Podemos representar o valor numérico do  $atributo_i$  por  $u_i$ , sendo que  $u_i$  pertence ao universo  $U_i$ . Em abordagens tradicionais, os casos em um espaço  $S = U_1 \times U_2 \times \dots \times U_N$  podem ser representados por um vetor  $(u_1, u_2, \dots, u_N)$  e o processo de recuperação pode ser realizado.

Para se construir um espaço nebuloso de casos, denominado Q, primeiramente descrevemos os  $N$  atributos de indexação dos casos como variáveis linguísticas, possuindo como valores os seguintes termos linguísticos (Mendes *et al.*, 2001):

$$\begin{aligned} T(atributo_1) &= \{T_{11}, T_{12}, \dots, T_{1m(1)}\} \\ T(atributo_2) &= \{T_{21}, T_{22}, \dots, T_{2m(2)}\} \\ &\vdots \\ T(atributo_N) &= \{T_{N1}, T_{N2}, \dots, T_{Nm(N)}\} \end{aligned}$$

Onde,  $T_{ij}$  é o nome do conjunto nebuloso descrito por uma função de pertinência  $\mu_{ij}(u_i)$ ,  $i$  indica a variável linguística  $i$  (*atributo<sub>i</sub>*),  $j$  indica o termo da variável linguística  $i$  e  $m(i)$  é o número de termos da variável linguística  $i$ .

Sendo assim, temos  $n$  conjuntos nebulosos descritos por funções de pertinência  $\mu_{ij}(u_i): U_i \rightarrow [0,1]$ , sendo:

$$n = \sum_{i=1}^N m(i)$$

Para podermos representar os casos no espaço nebuloso  $Q$ , que é representado por um hipercubo nebuloso  $I^n = [0, 1]^n$  como definido na seção 2.3.6, precisamos de uma função que faça de mapeamento de  $S$  para  $Q$ , ou  $F: S \rightarrow Q$ . Como temos  $S = U_1 \times U_2 \times \dots \times U_N$  e  $Q = [0, 1]^n$ , então  $F: U_1 \times U_2 \times \dots \times U_N \rightarrow [0, 1]^n$ .

Considerando cada atributo de maneira individual, podemos reduzir a função  $F$  para uma função  $F_i: U_i \rightarrow [0,1]^{m(i)}$ , para qualquer  $u_i \in U_i$ , determinada por (Mendes *et al.*, 2001):

$$F_i(u_i) = (\mu_{i1}(u_i), \mu_{i2}(u_i), \dots, \mu_{im(i)}(u_i)) \quad (3.1)$$

De modo semelhante, podemos definir, então, a função  $F$  como:

$$F(u_1, u_2, \dots, u_N) = (F_1(u_1), F_2(u_2), \dots, F_N(u_N)) \quad (3.2)$$

Se substituirmos os  $F_i$  's na função  $F$ , temos:

$$F = (\mu_{11}(u_1), \mu_{12}(u_1), \dots, \mu_{1m(1)}(u_1), \mu_{21}(u_2), \mu_{22}(u_2), \dots, \mu_{2m(2)}(u_2), \dots, \mu_{N1}(u_N), \mu_{N2}(u_N), \dots, \mu_{Nm(N)}(u_N)) \quad (3.3)$$

Tal função  $F$  permite mapear os casos do espaço  $S$  - espaço no qual os atributos de indexação assumem valores numéricos - para um hipercubo nebuloso  $n$ -dimensional. No espaço nebuloso  $Q$ , um caso pode ser interpretado como um conjunto nebuloso ordenado, geometricamente representado por um ponto, e descrito como um vetor de pertinência  $n$ -dimensional, como visto na seção 2.3.6.

Para exemplificar, tomemos como exemplo dois atributos de indexação: Idade e Peso, com  $U_1 = [0,100]$  e  $U_2 = [0,80]$  respectivamente. Os conjuntos de termos linguísticos podem ser definidos como:

$$T(Idade) = \{jovem, velho\}$$

$$T(Peso) = \{leve\}$$

As funções de pertinência dessas variáveis linguísticas são exibidas na Figura 13. Neste exemplo, uma pessoa com 45 anos de idade e 65 quilos de peso pode ser representada no espaço S pelo vetor (45,65). Utilizando as funções descritas anteriormente, temos:

$$F(45,65) = (\mu_{jovem}(45), \mu_{velho}(45), \mu_{leve}(65))$$

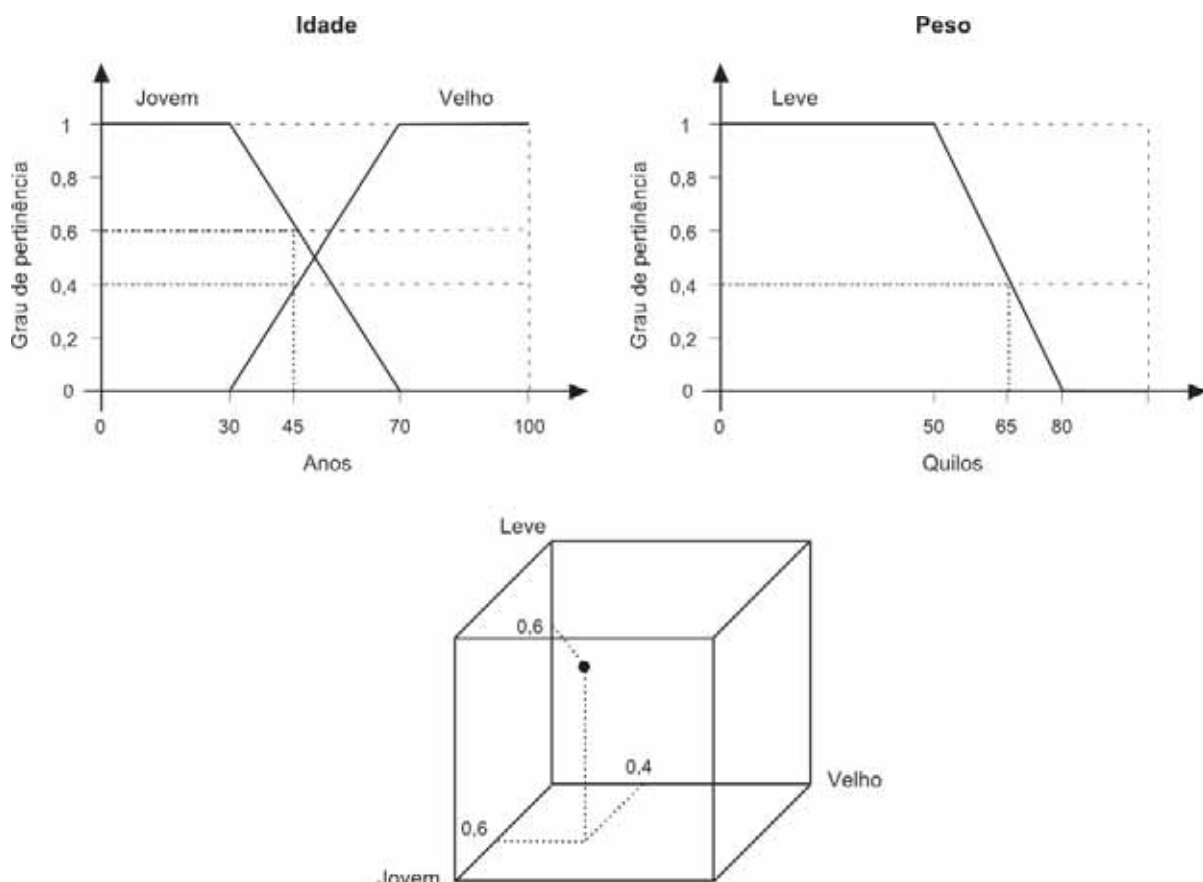
Deste modo, o caso pode ser representado no espaço nebuloso Q pelo vetor de pertinência (0.6, 0.4, 0.4).

Sendo assim:

$$(45,65) \rightarrow (0.6, 0.4, 0.4)$$

Este caso é um ponto no hipercubo nebuloso tridimensional mostrado na Figura 13.

Com essa abordagem, os casos representados com a notação vetorial no espaço nebuloso Q podem ser interpretados como conjuntos nebulosos ordenados.



**Figura 13:** Exemplo numérico em três dimensões com dois atributos: Idade e Peso. O ponto no cubo representa uma pessoa com 45 anos e 65 quilos.

### 3.3 Recuperação

Como descrito na seção 2.2.8, boa parte do sucesso de sistemas baseado em casos depende da etapa de recuperação. Portanto, ela pode ser considerada a principal área de pesquisa em RBC. Basicamente, a etapa de recuperação tem como objetivo principal selecionar na base de casos o caso, ou um conjunto de caso, mais próximo (similar) ao novo caso.

Considerando o espaço nebuloso, o processo de busca de casos também deve ter a capacidade de reconhecer o quanto um caso é semelhante ao novo problema. Uma das formas de se mensurar a semelhança entre os casos no espaço nebuloso é através do conceito de distância apresentado nas seções 2.1.8.1 e 2.2.6.

#### 3.3.1 Cálculo da similaridade

As medidas de similaridade que serão utilizadas no estudo de caso para recuperar os casos foram construídas utilizando-se a noção de distância entre conjuntos nebulosos no espaço nebuloso. Quando os casos são representados como pontos no espaço nebuloso  $Q$ , suas posições irão refletir a similaridade entre eles. Portanto, quanto mais próximos os casos estão, mais similares serão (Mendes *et al.*, 2001).

A distância entre dois conjuntos nebulosos  $A = (a_1, \dots, a_n)$  e  $B = (b_1, \dots, b_n)$ , pertencentes ao mesmo universo, pode ser descrita pelas distâncias mais comumente utilizadas:

- *Distância Euclidiana*, dada por:

$$d(A, B) = \left( \sum_{i=1}^n (a_i - b_i)^2 \right)^{1/2} \quad (3.4)$$

- *Distância Manhattan*, dada por:

$$d(A, B) = \sum_{i=1}^n |a_i - b_i| \quad (3.5)$$

- *Distância Minkowski*, dada por:

$$d(A, B) = \left( \sum_{i=1}^n |a_i - b_i|^q \right)^{1/q} \quad (3.6)$$

Esta distância é a generalização das duas distâncias anteriores. Quando o valor de  $q$  é igual a 1, esta distância representa a distância de Manhattan, e quando  $q = 2$ , a distância Euclidiana.

Quando os atributos dos casos possuem um peso relativo à sua importância, podemos representar, por exemplo, a distância Euclidiana ponderada como:

$$d(A, B) = \left( \sum_{i=1}^n w_i^2 (a_i - b_i)^2 \right)^{1/2} \quad (3.7)$$

sendo  $w_i$  o peso atribuído ao atributo  $i$ .

Com o auxílio do conceito de distância entre conjuntos nebulosos, podemos definir a noção de similaridade entre dois conjuntos nebulosos como sendo o complemento de sua distância, tendo no denominador um fator de normalização para obtermos o resultado no intervalo  $[0, 1]$  (adaptado de Mendes, 2001):

$$SM(A, B) = 1 - \frac{d(A, B)}{d(A \cup B, \emptyset)} \quad (3.8)$$

O conceito de união de conjuntos nebulosos é convenientemente representado tomando-se o máximo dos correspondentes graus de pertinência dos conjuntos (Mendes, 2001):

$$A \cup B = (\max(a_1, b_1), \dots, \max(a_n, b_n))$$

Logo, a medida de similaridade pode ser representada por:

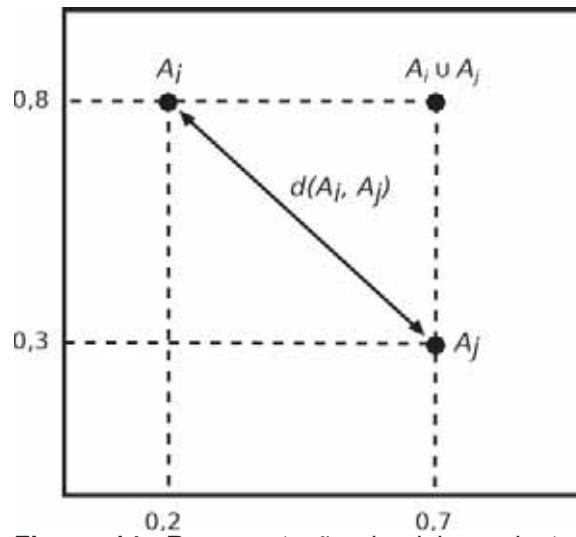
$$SM(A, B) = 1 - \frac{d(A, B)}{d(\max(A, B), 0)} \quad (3.9)$$

Utilizando a distância Euclidiana ponderada, essa expressão de similaridade pode ser definida

como:

$$SM(A, B) = 1 - \frac{\left( \sum_{i=1}^n w_i^2 (a_i - b_i)^2 \right)^{1/2}}{\left( \sum_{i=1}^n w_i^2 \cdot \max(a_i, b_i)^2 \right)^{1/2}} \quad (3.10)$$

Como exemplo de cálculo de similaridade, consideremos os conjuntos nebulosos  $A_i = (0,2; 0,8)$  e  $A_j = (0,7; 0,3)$ . A Figura 14 ilustra a representação desses conjuntos em um hipercubo bidimensional unitário, bem como a distância entre eles.



**Figura 14:** Representação de dois conjuntos nebulosos num hipercubo bidimensional unitário, e a distância entre eles.

Para calcular a similaridade dos conjuntos nebulosos  $A_i$  e  $A_j$ , representados na Figura 14, utilizaremos a medida de similaridade descrita pela equação (3.10):

$$d(A_i, A_j) = \sqrt{(0,2-0,7)^2 + (0,8-0,3)^2} = 0,71$$

$$A_i \cup A_j = (0,7; 0,8)$$

$$d(A_i \cup A_j, \emptyset) = \sqrt{(0,7-0)^2 + (0,8-0)^2} = 1,06$$

Portanto:

$$SM(A_i, A_j) = 1 - \frac{0,71}{1,06} = 0,33$$

Como neste exemplo todos os atributos têm a mesma importância, todos os pesos  $w_i$  são iguais a 1 e, desta maneira, foram omitidos do cálculo da similaridade.

### 3.3.2 Região de similaridade

Como descrito na seção 2.2.8.1, o princípio  $k$ -vizinhos mais próximos envolve buscar pelos  $k$  casos mais próximos ao caso de entrada atual utilizando uma medida de distância (como a distância Euclidiana) e então selecionar a classe da maioria desses  $k$  casos como a classe recuperada, no caso de problemas de classificação, ou então selecionar os  $k$  casos como sendo os casos mais similares.

Na abordagem proposta, quando um novo problema é fornecido, o processo de recuperação pode estabelecer uma posição do problema no espaço nebuloso, representada por uma  $n$ -tupla ordenada de graus de pertinência nebulosa, como visto anteriormente. Todos os casos próximos à posição do problema, ou seja, casos similares, podem ser considerados candidatos à solução.

Para determinar qual a proximidade mínima para que o caso seja considerado candidato à solução, podemos utilizar o conceito de hiperesfera. Matematicamente, a hiperesfera é definida como o conjunto de pontos que distam a mesma distância a um determinado ponto fixo a uma dada dimensão. Assim uma 2-esfera contém 2 dimensões, chamada de círculo, a 3-esfera contém 3 dimensões, chamada de esfera, e assim por diante.

Podemos construir uma esfera  $n$ -dimensional no espaço nebuloso  $Q$ , centrada no ponto que representa o problema, com o raio definido de acordo com o mínimo valor de similaridade desejado, visto que a distância no espaço  $Q$  é equivalente à similaridade. Os casos que podem ser considerados candidatos à solução do problema estarão situados dentro dessa hiperesfera (Mendes *et al.*, 2001).

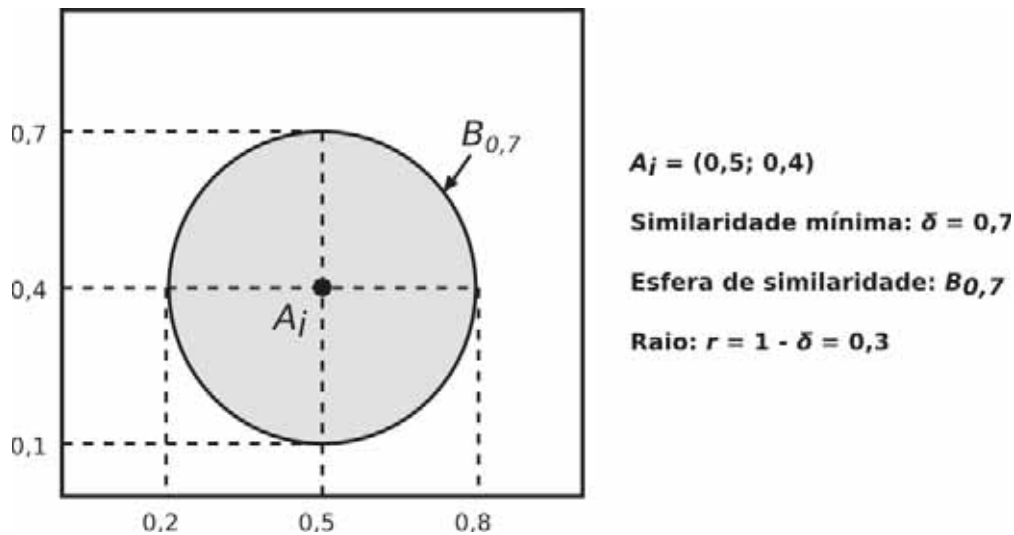
De uma maneira formal, a definição de uma hiperesfera de similaridade (Sadegh-Zadeh<sup>(b)</sup>, 1999):

**Definição:** dado um ponto  $A_i \in Q$  e um valor de similaridade  $\delta$ , uma hiperesfera de raio  $r = 1 - \delta$  centrada no ponto  $A_i$ , denotada por  $B_\delta(A_i)$ , pode ser definida como:

$$B_\delta(A_i) = \{ A_j \mid \text{Similaridade}(A_i, A_j) \geq \delta \}$$

Como podem haver diversos casos dentro dessa hiperesfera, podemos utilizar o princípio  $k$ -vizinhos mais próximos para selecionar os  $k$  casos mais próximos, para refinar o método de recuperação.

Para ilustrar, tomemos como exemplo o conjunto nebuloso  $A_i$  descrito pelo vetor de pertinência (0,5, 0,4). Na Figura 15 está representado do conjunto nebuloso  $A_i$  como o centro de uma hiperefera de similaridade com  $\delta = 0,7$ .



**Figura 15:** Exemplo de hiperesfera de similaridade

### 3.4 Adaptação

Uma vez que a etapa de recuperação selecionou e recuperou os casos mais similares ao problema, e assumindo que os casos similares têm soluções similares, podemos derivar as soluções dos casos recuperados para construir a solução do novo problema. Normalmente precisamos ajustar as soluções para se adequarem à nova situação, o que é feito na etapa de adaptação.

Existem duas maneiras de se realizar a adaptação. A abordagem tradicional é feita através de entrevistas com especialistas do domínio e codificando as tarefas específicas da adaptação manualmente no sistema RBC. Essas tarefas podem ser representadas por tabelas de decisões, árvores semânticas ou regras IF-THEN. Outra abordagem utilizada é a adaptação realizada através de técnicas de aprendizagem de máquina (Pal & Shiu, 2004).

Adquirir o conhecimento para a adaptação através de entrevistas com especialistas do domínio pode ser desgastante e consumir muito tempo.

Recentemente, devido à disponibilidade de casos e ao aumento de poder de processamento do computador, muitas abordagens de aprendizagem de máquina para derivar o conhecimento para a adaptação estão sendo desenvolvidas. Entre estas, as técnicas de computação flexível desempenham um papel único na aquisição de conhecimento para a adaptação que é impreciso,

incerto, e aproximado por natureza (Pal & Shiu, 2004).

Para o desenvolvimento do estudo de caso, utilizaremos as técnicas tradicionais de adaptação, tendo em vista que este trabalho tem foco principal nas etapas de indexação e recuperação dos casos.

De acordo com Pal & Shiu (2004), os principais tipos de estratégias tradicionais de adaptação de casos são descritos a seguir:

- *Reinstanciação* é a mais simples maneira de adaptação, na qual a solução do novo problema é simplesmente copiada do caso recuperado e utilizada diretamente, sem modificações.
- *Substituição* substitui partes dos atributos da solução antiga que são inválidos por serem conflitantes com os requisitos no novo problema ou o contradiz.
- *Transformação* é utilizada quando nenhum item substituto apropriado está disponível. A solução adequada será derivada baseada nas restrições e características da solução exigida. Uma restrição descreve ou define as propriedades de um(ns) componente(s) da solução. Ele especifica que propriedades a solução deve ou não ter. O componente da solução deve estar de acordo com as limitações, e nenhuma contradição ou conflito é permitido. Para identificar as restrições, alguns conhecimentos específicos ou heurística predefinidos devem estar disponíveis.

A fim de simplificar o desenvolvimento da adaptação no sistema RBC, visto que esta etapa não é o foco principal deste trabalho, a estratégia da reinstanciação será utilizada no estudo de caso, e será explicada com mais detalhes em seguida.

### **3.4.1 Reinstanciação**

A reinstanciação envolve a cópia direta e uso de uma solução antiga de um caso recuperado. Se a similaridade entre os casos recuperados é alta e nenhuma restrição ou requisito são impostos à solução exigida, a estratégia da reinstanciação pode ser utilizada. As vantagens da utilização dessa estratégia são o baixo custo computacional e a rápida respostas aos usuários (Pal & Shiu, 2004).

A Figura 16 ilustra um exemplo de escolha de imóvel utilizando a estratégia de reinstanciação. O caso A é o novo problema e o caso B é o caso mais similar recuperado da base de casos. Como os dois casos são bastante similares e não existem restrições, a solução do caso B é copiada e utilizada como solução para o caso A.

Nessa estratégia, o fator mais importante é a medida de similaridade utilizada para recuperar os casos, uma vez que a solução do caso mais similar será utilizada.



**Figura 16:** Exemplo de reinstanciação

Os passos da reinstanciação são descritos a seguir (Pal & Shiu, 2004):

1. Determinar uma medida de similaridade adequada e um limiar de similaridade apropriado
2. Recuperar o caso mais similar
3. Se o caso recuperado tem um valor de similaridade acima do limiar determinado e não existem restrições na solução exigida, a antiga solução é copiada e aplicada ao novo problema.

### 3.5 Retenção – Manutenção da base de casos

Como visto na seção 2.2.11, uma das estratégias de aprendizado em sistemas RBC é a adição de novos casos na base de casos.

Nesta estratégia, a adição do novo problema, sua solução, e seu resultado à base de casos aumenta o alcance das situações coberta pelos casos armazenados.

Com a utilização dessa estratégia de aprendizado, a manutenção da base de casos se torna indispensável para a evolução do conhecimento e eficiência do sistema, em se considerando o aumento do tamanho da base de casos e o custo computacional para a solução do problema (aumento do número de busca de casos, cálculo de similaridade, etc.).

A manutenção da base de casos envolve políticas e técnicas de adição, remoção e atualização dos casos, índices e outras estruturas (por exemplo, similaridades e adaptação) em um sistema RBC a fim de garantir sua eficácia e desempenho ao longo do tempo. Essas atividades podem ser

divididas em duas categorias: manutenção qualitativa e manutenção quantitativa. Manutenção qualitativa é realizada para assegurar a efetividade do sistema RBC, o que inclui consistência, precisão, e plenitude do sistema. Por sua vez, a manutenção quantitativa tem como objetivo garantir a eficiência da resolução do problema, como a melhoria do tempo médio de resolução, controle do tamanho da base de casos, e reorganização da estrutura de índices dos casos (Pal & Shiu, 2004).

Neste trabalho, a estratégia quantitativa será utilizada para a manutenção da base de dados.

Algumas tarefas envolvidas na manutenção quantitativa são descritas a seguir (Pal & Shiu, 2004):

- *Controle do tamanho da base de casos.* Geralmente a base de dados de sistemas RBC cresce rapidamente, causando ineficiência na etapa de recuperação dos casos. Não necessariamente a solução será melhorada conforme mais casos são adicionados. Porém, a remoção de muitos casos na base de casos pode comprometer a competência do sistema na resolução dos problemas. O controle do tamanho da base de dados deve sempre levar em consideração.
- *Revisão da estrutura de índice.* A estrutura de índice e a representação dos casos precisam ser reavaliadas e atualizadas quando existem mudanças no conhecimento do domínio ou regras de classificação. Além disso, pode ser necessária a criação de índices dos novos casos e os pesos dos atributos podem precisar serem recalculados.
- *Estatísticas de desempenho.* Alguns indicadores de desempenho podem ser coletados para indicar a efetividade do sistema para ajustes das atividades de manutenção, tais como uso dos casos, custo da recuperação, frequência de acesso aos casos, entre outros.
- *Identificação de casos obsoletos.* Casos obsoletos são casos que não têm mais valor para o sistema, ou seja, as suas soluções não são mais válidas por qualquer motivo. Esses casos devem ser removidos da base de casos para melhorar a eficiência geral do sistema.

## **4 RBC Nebuloso para Recomendação de Sistemas Marítimos de Produção de Petróleo**

A partir dos conceitos e técnicas apresentadas na seção 2, e considerando a abordagem proposta da seção 3, nesta seção é apresentado o desenvolvimento do sistema RBC nebuloso para recomendação de sistemas marítimos de produção de petróleo, denominado RIGS. Também são apresentados os testes e resultados obtidos com o sistema.

### **4.1 Arquitetura do sistema**

A arquitetura do sistema nebuloso RBC para recomendação foi concebida considerando os aspectos da abordagem proposta, descritos na seção 3.

O sistema foi construído de maneira flexível, permitindo a interação com o usuário desde o momento da definição do domínio da aplicação, possibilitando:

- Configuração da representação dos casos.
- Definição dos índices e seus pesos.
- Definição dos conjuntos nebulosos e suas funções de pertinência.
- Escolha das medidas de similaridade e métricas de distância utilizadas na recuperação.
- Definição da região de similaridade do número de casos similares recuperados.
- Visualização dos casos mais similares recuperados para análise da melhor solução encontrada.
- Adaptação e retenção do novo caso contendo a solução sugerida.
- Visualização e manutenção da base de casos.

A construção do sistema e suas ferramentas foi feita sobre uma arquitetura cliente-servidor, aumentando seu desempenho, facilitando o acesso e manutenção e centralizando as informações.

### **4.2 Implementação do sistema**

Para a implementação do sistema, as seguintes tecnologias e ferramentas foram utilizadas:

## Java

Linguagem de programação orientada a objetos, mantida pela Oracle, utilizadas na construção da camada servidor do sistema.

## jCOLIBRI

O jColibri é um *framework* desenvolvido em Java para a construção de sistemas com Raciocínio Baseado em Casos, desenvolvido e mantido pelo *Group for Artificial Intelligence Applications (GAIA)* (jCOLIBRI, 2011). Ele inclui mecanismos para a Recuperação, Reutilização, Revisão e Retenção dos casos, compreendendo assim todo o ciclo RBC.

## CAppCon

O CAppCon (Foster, 2007) é um *framework* desenvolvido e mantido pela empresa Caiena, contendo um conjunto de tecnologias, padrões e técnicas que permitem o desenvolvimento rápido de soluções. As tecnologias envolvidas no *framework* são: JavaScript, CSS (*Cascading Style Sheets*), DOM (*Document Object Mode*), JSON (*JavaScript Object Notation*), XUL (*XML User Interface Language*), entre outras.

## PostgreSQL

Sistema gerenciador de banco de dados objeto-relacional de código aberto, utilizado para representar a base de casos dos sistema.

### 4.2.1 Representação dos sistemas marítimos de produção

Os casos utilizados correspondem a sistemas marítimos de produção que foram utilizados anteriormente em desenvolvimento de campos de petróleo (Morooka & Carvalho, 2011). Inicialmente os casos foram representados através de tabelas em um banco de dados relacional. Utilizando essa abordagem de representação, um modelo relacional foi criado compreendendo diversos atributos relacionados aos aspectos físicos e ambientais de sistemas marítimos de produção.

A definição desses atributos bem como toda a coleta de dados para a construção dessa base de casos foram realizados no trabalho de Morooka & Carvalho (2011). Essa base de casos será utilizada no estudo de caso a fim de validar a abordagem proposta, realizar os testes e comparar os resultados obtidos.

Na criação da base de casos, foram inseridos 377 casos, coletados e disponíveis no levantamento feito por Morooka & Carvalho (2011).

A Figura 17 ilustra o modelo relacional criado, representando a base de casos e contendo 26

tabelas relacionais que representam os sistemas marítimos de produção, contendo diversos atributos relacionados à sua configuração, tais como atributos relacionados aos poços, condições ambientais, sistemas de amarração e ancoragem, tipos de plataformas, informações sobre o reservatório, *risers*, etc.

É importante observar que o modelo criado contém muitos outros atributos além dos atributos essenciais que devem ser analisados para a configuração dos sistemas marítimos de produção, que serão definidos na seção seguinte.

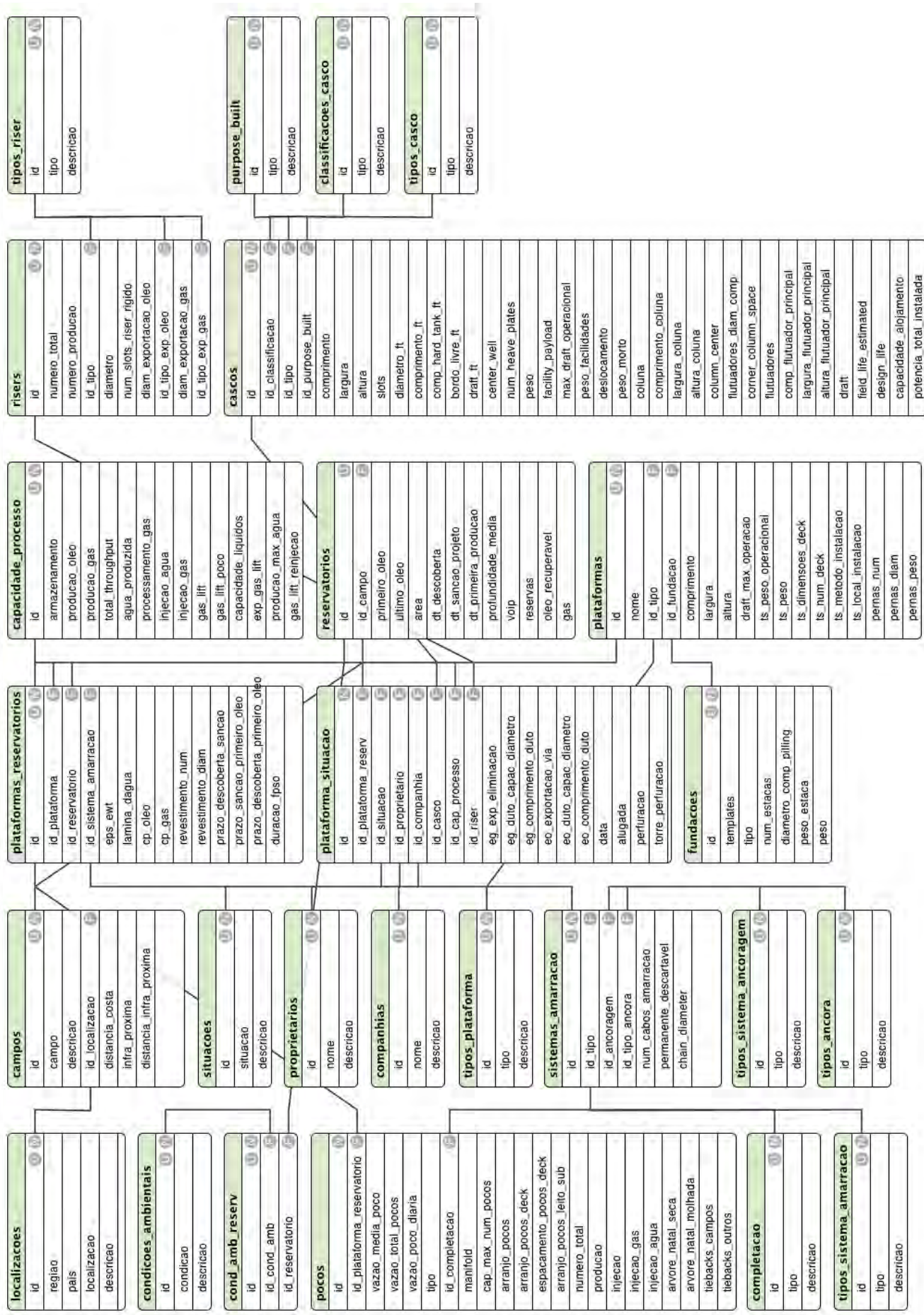


Figura 17: Modelo relacional representando os sistemas maritimos de producao

#### 4.2.2 Atributos para definição do sistema de produção

De acordo com Franco (2003), a escolha da melhor configuração de um sistema de produção entre as diversas opções existentes, depende de alguns atributos que influenciam diretamente essas opções. Em seu trabalho, Franco (2003) esses atributos foram selecionados com a ajuda de especialistas do domínio e de alguns trabalhos que têm o intuito de propor métodos para avaliar melhores projetos de desenvolvimento de campos petrolíferos.

Os principais atributos selecionados no trabalho de Franco (2003), que constituem um sistema marítimo de produção, e que fazem parte da representação dos casos neste trabalho, são: arranjo dos poços, uso de *manifold*, tipo da unidade estacionária de produção (UEP), tipo de ancoragem, tipo de *riser* e tipo de escoamento e armazenamento de óleo.

Tendo em vista que a representação dos casos pode ser dividida em duas partes, como descrito na seção 3.1, sendo uma delas a descrição do problema e a outra a sua solução, esses atributos podem ser considerados como a parte que representa a solução, uma vez que são os principais atributos da configuração de um sistema marítimo de produção e devem ser recomendados pelo sistema RBC.

Para que o sistema RBC possa definir e recomendar tais atributos como solução ao problema, a descrição do problema deve ser fornecida pelo usuário. A descrição do problema é representada pelos atributos que melhor descrevem um sistema marítimo de produção, e, desta forma, são utilizados como atributos de indexação dos casos.

Os atributos de indexação utilizados pelo sistema RBC são os definidos pelo trabalho de Franco (2003), sendo eles a área do reservatório, profundidade do reservatório, número de poços, tipo de poço, vazão diária por poço, lâmina d'água, condições ambientais, distância da costa, existência de infra-estrutura próxima ao campo e capacidade de armazenamento.

A representação dos casos, contendo os atributos da descrição (índices) e da solução, é apresentada na Figura 18.

## CASO

|                                                                                                                                                                                                                                                                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>DESCRIÇÃO</b><br><br>Área do reservatório<br>Profundidade do reservatório<br>Número de poços<br>Tipo de poço<br>Vazão diária por poço<br>Lâmina d'água<br>Condições ambientais<br>Distância da costa<br>Infra-estrutura próxima<br>Capacidade de armazenamento |
| <b>SOLUÇÃO</b><br><br>Arranjo dos poços<br>Manifold<br>Tipo da UEP<br>Tipo de ancoragem<br>Tipo de riser<br>Escoamento e armazenamento de óleo                                                                                                                    |

**Figura 18:** Representação dos casos contendo os atributos da descrição (índices) e da solução do problema.

### 4.2.3 Definição dos conjuntos nebulosos

Como definido na seção 3.2, após a definição dos atributos de indexação e considerando a teoria dos conjuntos nebuloso, o espaço nebuloso  $Q$  pode ser construído. Inicialmente, os atributos de indexação devem ser descritos como variáveis linguísticas, com os seus termos linguísticos. Então, funções de pertinência são criadas para fazerem o mapeamento de cada termo para um conjunto nebuloso.

A descrição dos atributos de indexação como variáveis linguísticas, bem como seus termos, foi feita de acordo com o trabalho de Franco (2003) e Morooka & Carvalho (2011), como se segue:

$T(\text{área do reservatório}) = (\text{pequena}, \text{média}, \text{grande})$

$T(\text{profundidade do reservatório}) = (\text{raso}, \text{médio}, \text{profundo})$

$T(\text{número de poços}) = (\text{baixo}, \text{médio}, \text{alto})$

$T(\text{vazão diária por poço}) = (\text{baixa}, \text{média}, \text{alta})$

$T(\text{lâmina d'água}) = (\text{rasa}, \text{média}, \text{profunda}, \text{ultraprofunda})$

$T(\text{distância da costa}) = (\text{pequena}, \text{média}, \text{grande})$

$T(\text{capacidade de armazenamento}) = (\text{baixa}, \text{média}, \text{alta})$

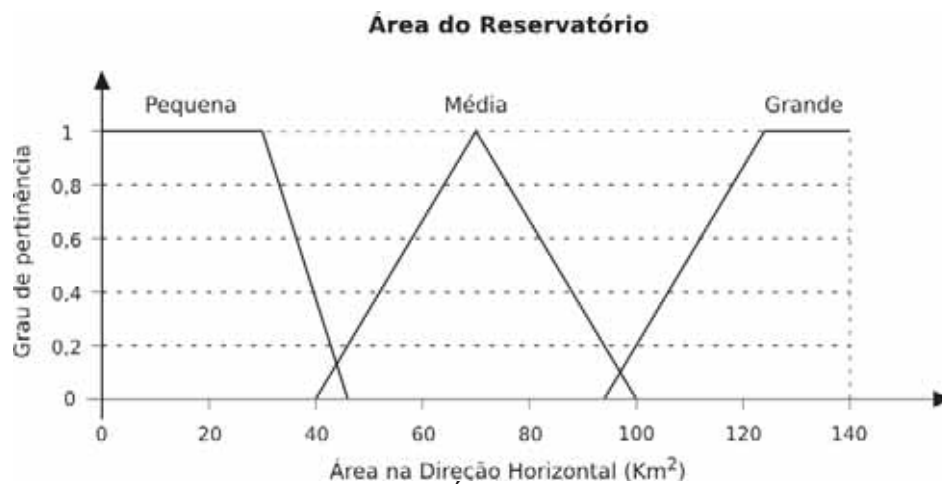
$T(\text{tipo de poço}) = (\text{vertical}, \text{horizontal})$

$T(\text{condições ambientais}) = (\text{amena}, \text{moderada}, \text{severa})$

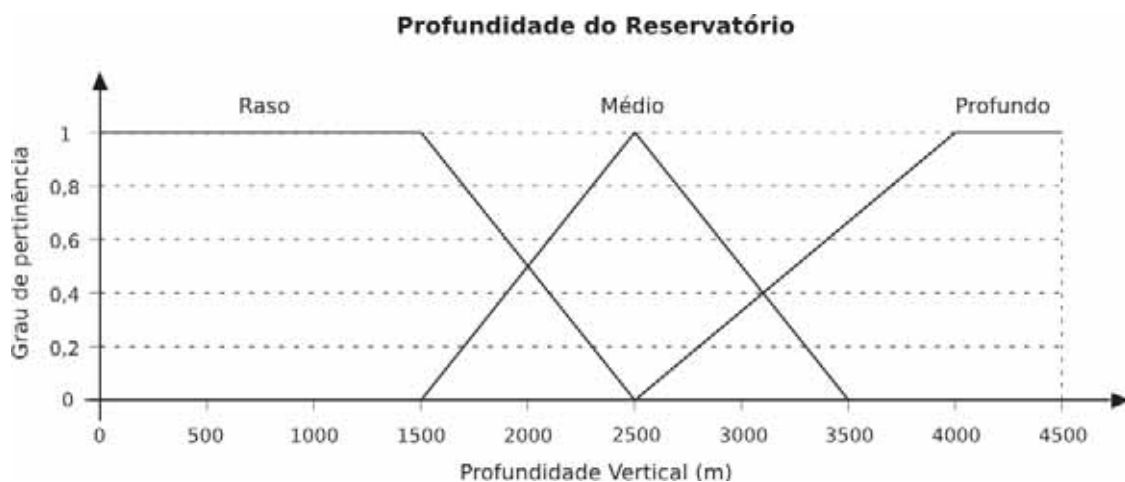
$T(\text{infraestrutura próxima}) = (\text{sim}, \text{não})$

Os conjuntos nebulosos associados aos atributos que possuem valores numéricos – ou seja, para área do reservatório, profundidade do reservatório, número de poços, vazão diária por poço, lâmina d'água, distância da costa e capacidade de armazenamento – e as funções de pertinência que os descrevem estão definidos nas Figuras 19 a 25.

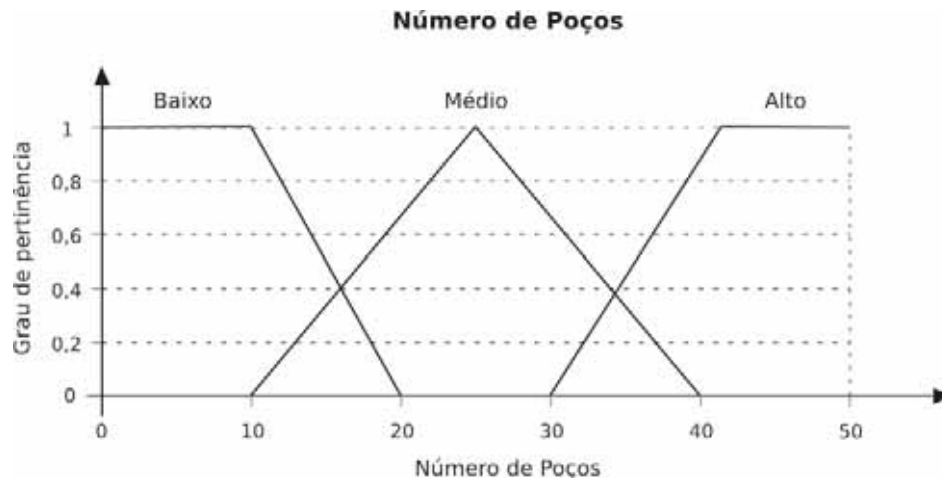
Os atributos que são expressos em um conceito lingüístico determinístico não precisam ser mapeados para os conjuntos nebulosos através de funções de pertinência, ou seja, *tipo de poço* pode ser vertical ou horizontal, *condições ambientais* pode assumir os valores amena, moderada ou severa e *infraestrutura próxima* é existente ou não.



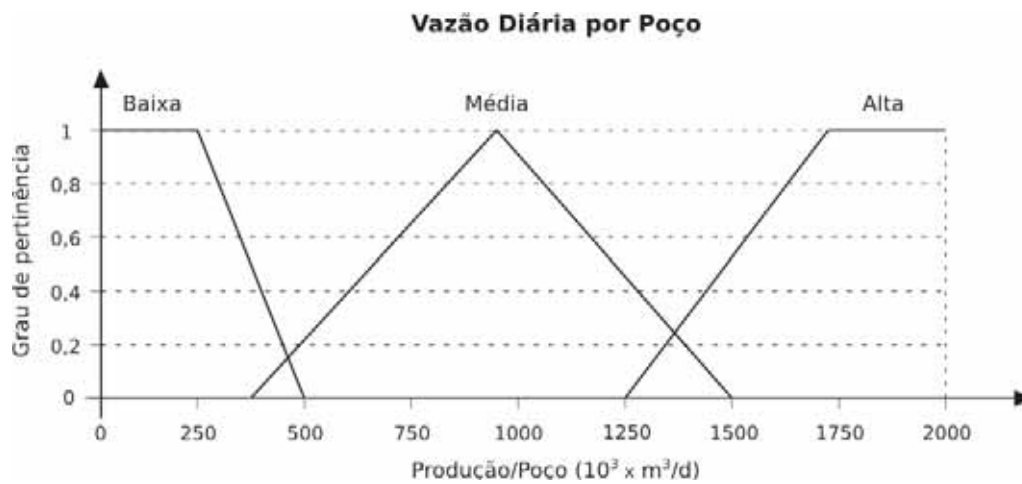
**Figura 19:** Conjunto nebuloso Área do Reservatório (Franco, 2003)



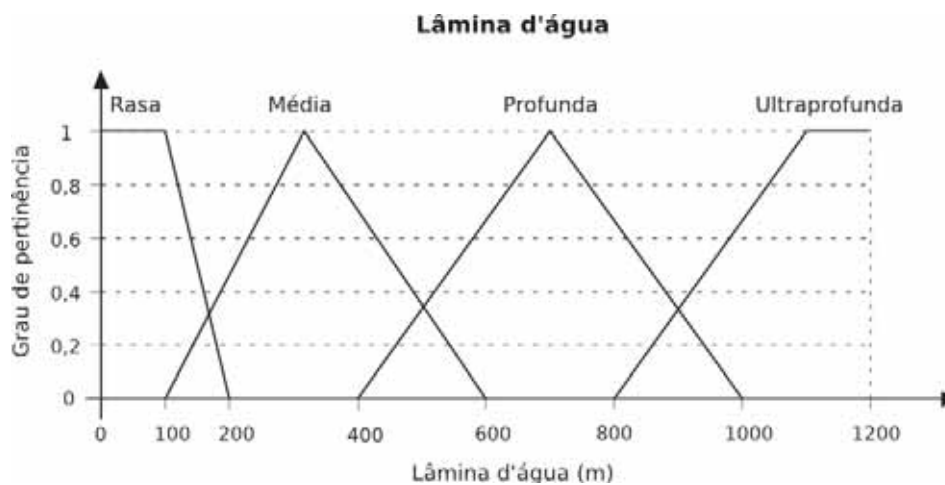
**Figura 20:** Conjunto nebuloso Profundidade do Reservatório (Franco, 2003)



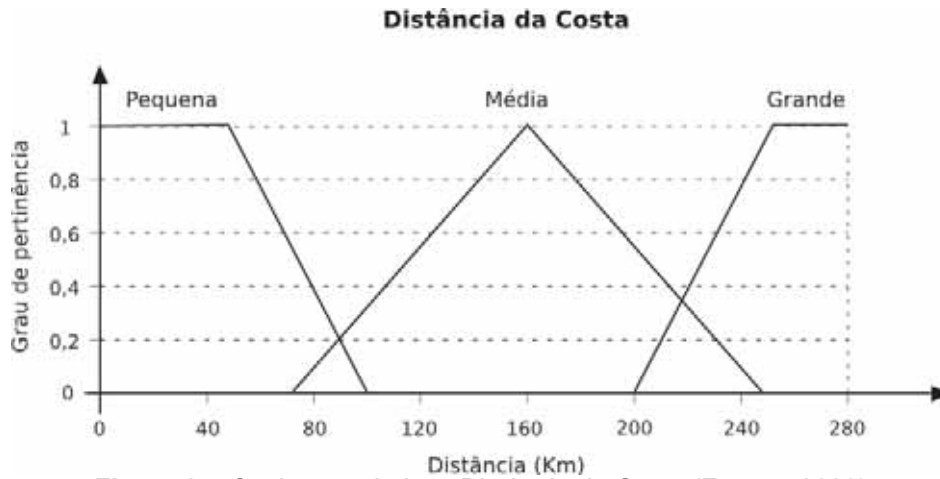
**Figura 21:** Conjunto nebuloso Número de Poços (Franco, 2003)



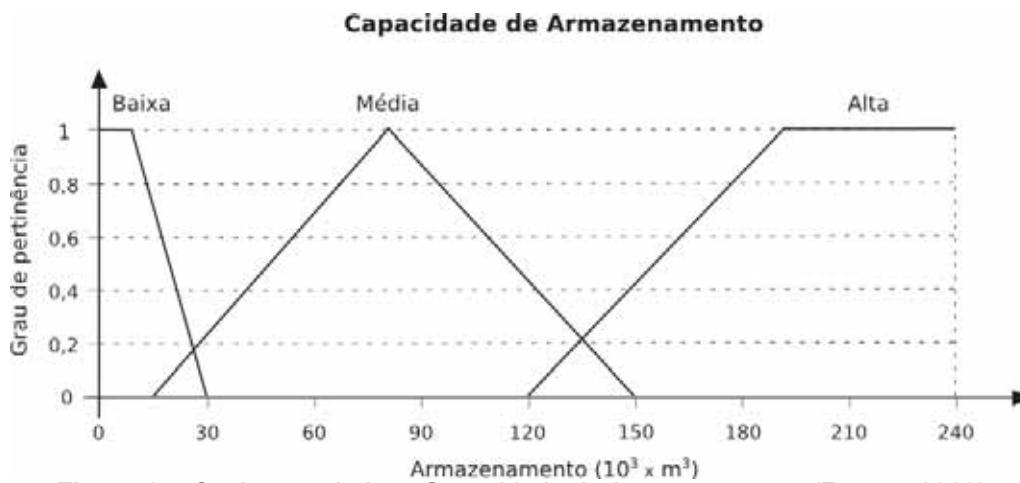
**Figura 22:** Conjunto nebuloso Vazão Diária por Poço (Franco, 2003)



**Figura 23:** Conjunto nebuloso Lâmina d'água (Franco, 2003)



**Figura 24:** Conjunto nebuloso Distância da Costa (Franco, 2003)



**Figura 25:** Conjunto nebuloso Capacidade de Armazenamento (Franco, 2003)

Uma vez tendo definido o espaço nebuloso, precisamos de uma função que faça o mapeamento do espaço  $S$  - no qual os atributos de indexação assumem valores numéricos - para o espaço nebuloso  $Q$ , para podermos representar os casos. O espaço nebuloso  $Q$  é representado por um hipercubo unitário  $I^n = [0, 1]^n$  como definido na seção 3.2. O número de dimensões desse hipercubo é igual a soma de todos os termos linguísticos de todos os atributos que tiveram conjuntos nebulosos associados, ou seja,  $n = 22$ .

Essa função de mapeamento  $F: S \rightarrow Q$  é definida pela fórmula (3.3), e introduzindo os atributos de indexação temos a seguinte função  $F$  (adaptado de Mendes, 2001):

$$F = (\mu_{11}(u_1), \mu_{12}(u_1), \mu_{13}(u_1), \mu_{21}(u_2), \mu_{22}(u_2), \mu_{23}(u_2), \mu_{31}(u_3), \mu_{32}(u_3), \mu_{33}(u_3), \mu_{41}(u_4), \mu_{42}(u_4), \mu_{43}(u_4), \mu_{51}(u_5), \mu_{52}(u_5), \mu_{53}(u_5), \mu_{54}(u_5), \mu_{61}(u_6), \mu_{62}(u_6), \mu_{63}(u_6), \mu_{71}(u_7), \mu_{72}(u_7), \mu_{73}(u_7))$$

sendo  $u_1, u_2, u_3, u_4, u_5, u_6, u_7$ , respectivamente aos valores numéricos de área do reservatório,

profundidade do reservatório, número de poços, vazão diária por poço, lâmina d'água, distância da costa e capacidade de armazenamento. As funções  $\mu_{ij}(u_i)$  são as funções de pertinência que descrevem os conjuntos nebulosos associados aos atributos. As funções de pertinência são denotadas na Tabela 1.

**Tabela 1:** Relação das funções de pertinência com os termos linguísticos de cada conjunto nebuloso

| Atributos                    | Termos        | Funções de Pertinência |
|------------------------------|---------------|------------------------|
| Área do reservatório         | Pequena       | $\mu_{11}(u_1)$        |
|                              | Média         | $\mu_{12}(u_1)$        |
|                              | Grande        | $\mu_{13}(u_1)$        |
| Profundidade do reservatório | Raso          | $\mu_{21}(u_2)$        |
|                              | Médio         | $\mu_{22}(u_2)$        |
|                              | Profundo      | $\mu_{23}(u_2)$        |
| Número de poços              | Baixo         | $\mu_{31}(u_3)$        |
|                              | Médio         | $\mu_{32}(u_3)$        |
|                              | Alto          | $\mu_{33}(u_3)$        |
| Vazão diária por poço        | Baixa         | $\mu_{41}(u_4)$        |
|                              | Média         | $\mu_{42}(u_4)$        |
|                              | Alta          | $\mu_{43}(u_4)$        |
| Lâmina d'água                | Baixa         | $\mu_{51}(u_5)$        |
|                              | Média         | $\mu_{52}(u_5)$        |
|                              | Profunda      | $\mu_{53}(u_5)$        |
|                              | Ultraprofunda | $\mu_{54}(u_5)$        |
| Distância da costa           | Pequena       | $\mu_{61}(u_6)$        |
|                              | Média         | $\mu_{62}(u_6)$        |
|                              | Grande        | $\mu_{63}(u_6)$        |
| Capacidade de armazenamento  | Baixa         | $\mu_{71}(u_7)$        |
|                              | Média         | $\mu_{72}(u_7)$        |
|                              | Alta          | $\mu_{73}(u_7)$        |

Este caso é representado por um vetor 22-dimensional de valores de pertinência, e corresponde a um ponto no espaço nebuloso Q.

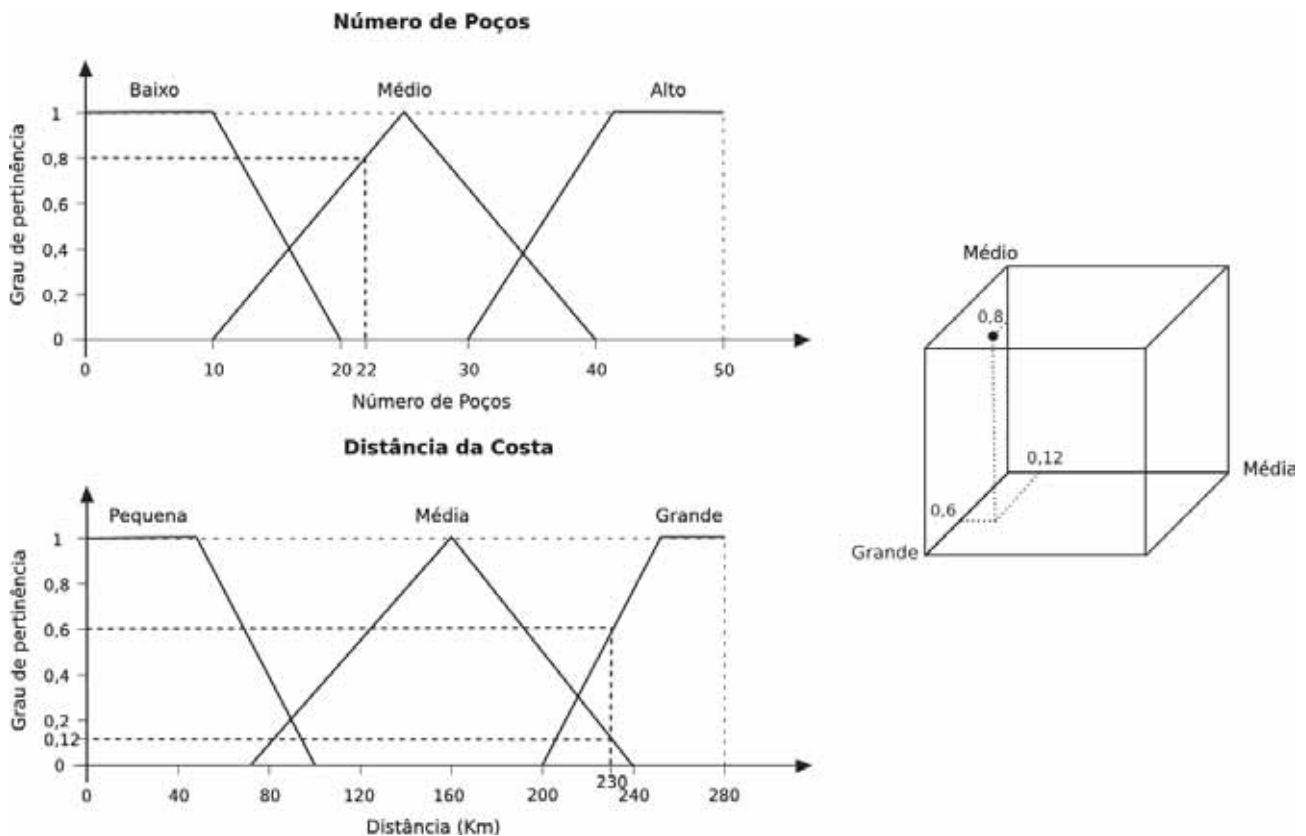
Além dos atributos que têm conjunto nebuloso associado, os casos possuem mais três atributos de indexação que não são mapeados através de funções de pertinência, sendo  $u_8$ ,  $u_9$  e

$u_{10}$  respectivamente os valores dos atributos *tipo de poço*, *condições ambientais* e *infraestrutura próxima*. Portanto, o caso passa a ser indexado por uma 25-upla ordenada de valores, sendo que a função F representa 22 valores de pertinência:

$$(F(u_1, u_2, u_3, u_4, u_5, u_6, u_7), u_8, u_9, u_{10})$$

Para ilustrar, considere o exemplo da Figura 26 de três dimensões no espaço nebuloso para facilitar a visualização. Neste exemplo, o caso pode ser representado no espaço S pela dupla ordenada (22, 230). Aplicando-se a função F, o seguinte mapeamento para o espaço nebuloso Q é obtido:

$$(22, 230) \rightarrow (0.8, 0.12, 0.6)$$



**Figura 26:** Exemplo de três dimensões de um caso com 22 poços e 230 Km de distância da costa, representado no espaço nebuloso Q por um ponto.

#### 4.2.4 Recuperação dos casos similares

De acordo com a abordagem definida na seção 3.3, utilizaremos o conceito de hipersfera de similaridade para delimitar uma região de casos similares. A medida de similaridade utilizada na

recuperação será a definida pela equação (3.11), que faz uso da distância Euclidiana ponderada para medir a distância entre os casos.

Após a recuperação dos casos similares pertencentes à hiperesfera de similaridade, o princípio do  $k$ -vizinhos mais próximos será utilizado como ferramenta de ajuste para ter como resultado somente os  $k$  casos mais similares ao problema.

Para ilustrar essa etapa de recuperação dos casos similares, considere o exemplo em que se deseja recuperar os casos similares ao caso  $A$ , exibido na Figura 26, acrescido do atributo *condições ambientais*, representado pelo vetor ordenado de valores  $A = (22, 230, amena)$ , correspondendo respectivamente aos valores dos atributos *número de poço*, *distância da costa* e *condições ambientais*, ao qual não é associado um conjunto nebuloso.

Para os atributos que possuem conjuntos nebulosos associados, tem-se a função  $F$  que faz o mapeamento do espaço  $S$  para o espaço nebuloso  $Q$ ,  $F : S \rightarrow Q$ , logo a representação do caso no espaço nebuloso pode ser denotada como  $A = (F(22, 230), amena)$ . Aplicando-se a função  $F$  de acordo com as funções de pertinência da Figura 26, a 7-upla ordenada de valores que indexa o caso  $A$ :

$$(F(22, 230), amena) \rightarrow (\mu_{baixo}(22), \mu_{médio}(22), \mu_{alto}(22), \mu_{pequena}(230), \mu_{média}(230), \mu_{grande}(230), amena)$$

$$(F(22, 230), amena) \rightarrow (0, 0.8, 0, 0, 0.12, 0.6, amena)$$

Para se considerar o valor referente ao atributo *condições ambientais* no cálculo de similaridade, é necessário alterar a medida de similaridade da equação (3.11) (adaptado de Mendes, 2001):

$$SM_{AB}^{(w)} = 1 - \frac{\left( \sum_{i=1}^6 w_i^2 (a_i - b_i)^2 \right)^{1/2} + w_7 \cdot \lambda(a_7, b_7)}{\left( \sum_{i=1}^6 w_i^2 \cdot \max(a_i, b_i)^2 \right)^{1/2} + w_7 \cdot \lambda(a_7, b_7)} \quad (4.1)$$

onde  $a_7$  e  $b_7$  são os valores correspondentes ao atributo *condições ambientais* dos casos  $A$  e  $B$ , e a função  $\lambda(a_7, b_7)$  é definida como (adaptado de Mendes, 2001):

$$\lambda(a_7, b_7) = \begin{cases} 0 & \text{se } a_7 = b_7 \\ 1 & \text{se } a_7 \neq b_7 \end{cases}$$

Considere outros 4 casos,  $B$ ,  $C$ ,  $D$  e  $E$  para serem utilizados nos cálculos de similaridade e no

processo de recuperação, descritos na Tabela 2. Considerando uma hiperesfera de similaridade  $\delta = 0,5$ , centrada no caso *A* (representado por um ponto no espaço nebuloso *Q*), denotada por  $B_{0,5}(A)$ , podemos aplicar a medida de similaridade entre o caso *A* e os demais para recuperar os casos pertencentes à hiperesfera de similaridade *B*.

**Tabela 2:** Casos utilizados para o cálculo de similaridade (valores aproximados para facilitar o cálculo)

|                      |       |         | Valores de Pertinência |       |        |       |       |
|----------------------|-------|---------|------------------------|-------|--------|-------|-------|
| Atributos            | Pesos | Termos  | A                      | B     | C      | D     | E     |
| Número de poços      | 2     | Baixo   | 0                      | 0.2   | 0.3    | 0     | 0     |
|                      |       | Médio   | 0.8                    | 0.7   | 0.45   | 0.6   | 0.5   |
|                      |       | Alto    | 0                      | 0     | 0      | 0.1   | 0.2   |
| Distância da costa   | 1     | Pequena | 0                      | 0     | 0      | 0     | 0     |
|                      |       | Média   | 0.12                   | 0.15  | 0      | 0.2   | 0.3   |
|                      |       | Grande  | 0.6                    | 0.5   | 0.8    | 0.5   | 0.3   |
|                      |       |         | Valores                |       |        |       |       |
| Condições ambientais | 1     |         | amena                  | amena | severa | amena | amena |

Para exemplificar o cálculo de similaridade, tomemos os casos *A* e *B*. Aplicando-se a medida de similaridade da equação (4.1), temos:

$$SM_{AB}^{(w)} = 1 - \frac{\sqrt{4(0-0.2)^2 + 4(0.8-0.7)^2 + 4(0-0)^2 + (0-0)^2 + (0.12-0.15)^2 + (0.6-0.5)^2} + \lambda(amena, amena)}{\sqrt{4(0.2)^2 + 4(0.8)^2 + 4(0)^2 + (0)^2 + (0.15)^2 + (0.6)^2} + \lambda(amena, amena)}$$

$$SM_{AB}^{(w)} = 1 - \frac{\sqrt{0.16 + 0.04 + 0 + 0 + 0.01 + 0}}{\sqrt{0.16 + 2.56 + 0 + 0 + 0.02 + 0.36} + 0}$$

$$SM_{AB}^{(w)} = 1 - \frac{0.46}{1.76} = 0.74$$

Os cálculos de similaridade podem ser visualizados na Tabela 3.

**Tabela 3:** Similaridade entre o caso *A* e os casos *B*, *C*, *D* e *E*.

|          |  | Similaridade |          |          |          |
|----------|--|--------------|----------|----------|----------|
| Casos    |  | <i>B</i>     | <i>C</i> | <i>D</i> | <i>E</i> |
| <i>A</i> |  | 0.74         | 0.32     | 0.72     | 0.55     |

Sendo o raio da hiperesfera de similaridade  $B_{0,5}(A)$  definido por  $r = 1 - \delta = 0.5$ , e realizados os

cálculos de similaridade, podemos verificar que os casos *B*, *D* e *E* pertencem à região delimitada por  $B_{0,5}(A)$  e, portanto, são considerados similares ao caso *A* (similaridade  $> 0.5$ ).

Se utilizarmos o princípio do *k*NN, com  $k = 2$ , somente os casos *B* e *D* seriam recuperados por serem os 2 casos mais similares ao caso *A*.

#### 4.2.5 RIGS

O propósito principal do sistema é auxiliar o usuário na tomada de decisão sobre a configuração do sistema marítimo de produção com maior confiabilidade e agilidade, tendo em vista que toda a base de conhecimento utilizada pelo mecanismo de raciocínio do sistema foi construída com a aquisição do conhecimento de especialistas da área e da literatura, realizada no trabalho de Morooka & Carvalho (2011).

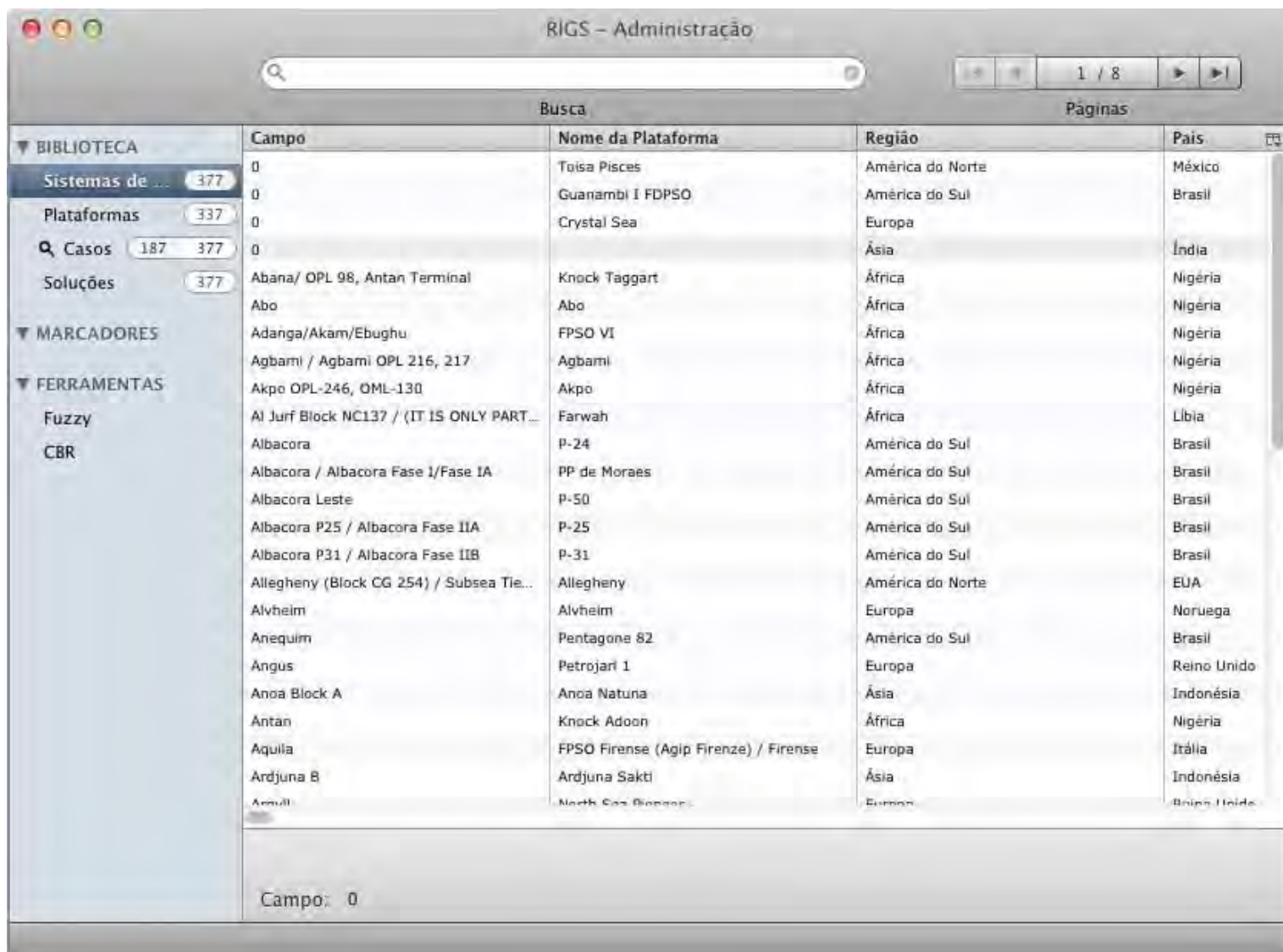
A seguir, as principais ferramentas e interfaces do RIGS são apresentadas.

A Figura 27 apresenta a janela inicial do RIGS, onde é feita a configuração do sistema e realizado o acesso.



**Figura 27:** Janela de login do RIGS

A Figura 28 exibe a ferramenta principal do sistema, onde são exibidas as informações da base de casos, como os Sistemas de Produção, Plataforma e todos os seus atributos, bem como a representação dos casos, seus atributos e soluções.



**Figura 28:** Ferramenta principal do RIGS, com ferramentas de busca, navegação, visualização e manutenção da base de casos.

A Figura 29 mostra a utilização da ferramenta de busca da janela principal do sistema. Com ela, o usuário pode procurar por qualquer valor de qualquer atributo relativo aos casos da base de casos, informando o termo que deseja ou somente parte dele. Também é possível realizar buscas avançadas, fazendo a composição de buscas por diversos atributos. No exemplo exibido, o usuário informa o valor “amena” na ferramenta de busca, e o sistema destaca todos os casos que contém, em qualquer um de seus atributos, o valor “amena”. Neste caso, somente são exibidos os casos em que o atributo *condições ambientais* possui o valor “amena”. Com essa ferramenta, a procura de informações da base de casos é bastante facilitada, colaborando com as atividades envolvidas na sua manutenção e verificação.

RIGS – Administração

Q

amena

1

4

Busca

Páginas

▼ BIBLIOTECA

Sistemas de ... 377

Plataformas 337

Q Cas... 187 377

Soluções 377

▼ MARCADORES

▼ FERRAMENTAS

Fuzzy

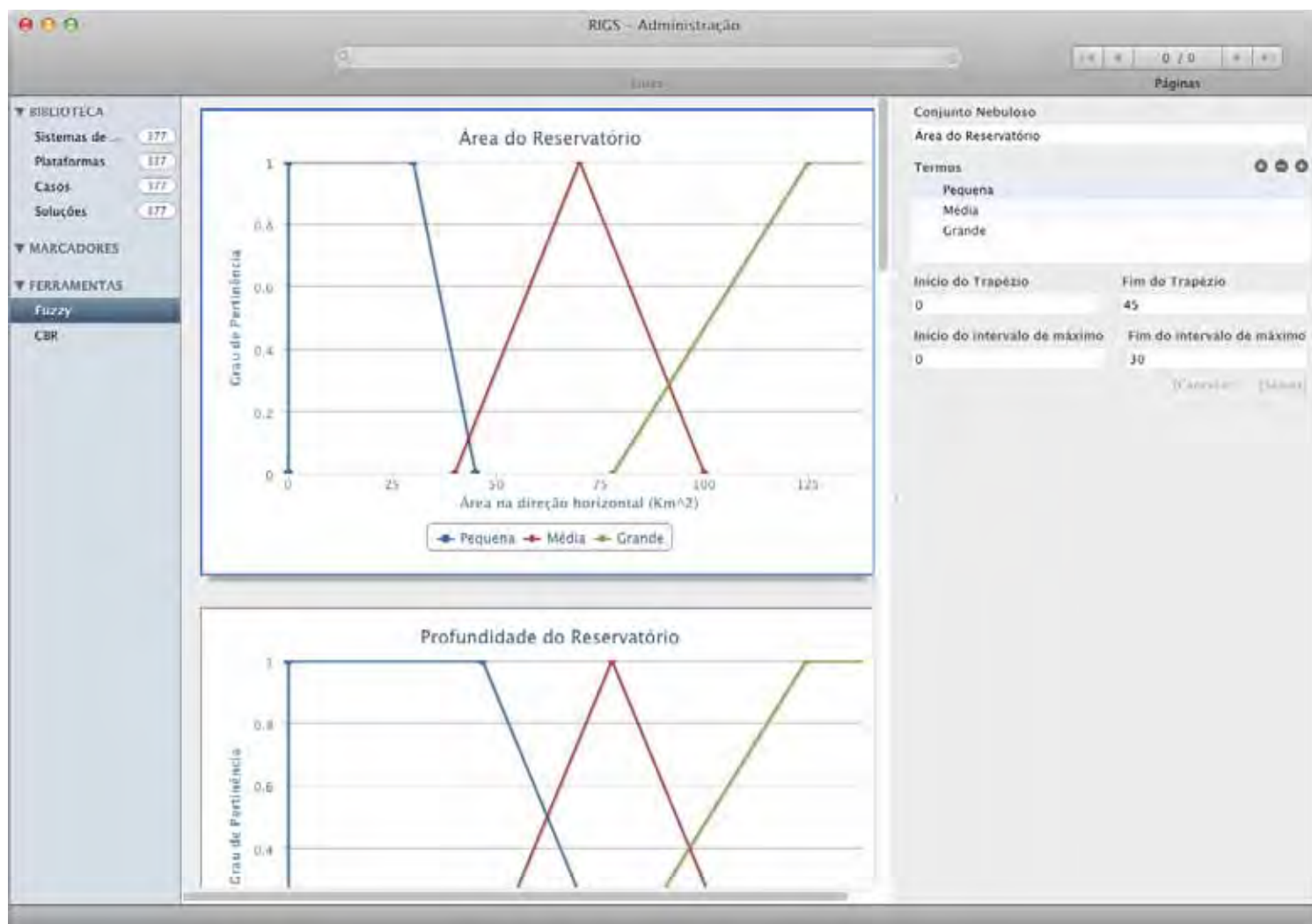
CBR

| Área do Reservatório | Profundidade... | Número de P... | Tipo de Poço | Vazão Diária... | Lâmina d'á |
|----------------------|-----------------|----------------|--------------|-----------------|------------|
| 884.11               | 2799.89         | 0              |              | 0               | 1020       |
| 884.11               | 2799.89         | 2              |              | 0               | 1215       |
| 884.11               | 2799.89         | 21             |              | 0               | 1080       |
| 884.11               | 2799.89         | 11             |              | 0               | 1180       |
| 884.11               | 2799.89         | 21             |              | 0               | 1700       |
| 884.11               | 2799.89         | 24             |              | 0               | 1255       |
| 884.11               | 2799.89         | 2              |              | 0               | 1430       |
| 720                  | 3150.11         | 17             |              | 0               | 950        |
| 457.69               | 2749.91         | 1              |              | 0               | 600        |
| 457.69               | 2749.91         | 22             | Horizontal   | 1576.92         | 1080       |
| 457.69               | 2749.91         | 0              | Direcional   | 1666.67         | 1400       |
| 455                  | 3631.08         | 18             | Horizontal   | 300             | 230        |
| 455                  | 3631.08         | 29             | Horizontal   | 300             | 575        |
| 455                  | 3631.08         | 12             | Horizontal   | 300             | 224        |
| 455                  | 3631.08         | 36             | Horizontal   | 300             | 330        |
| 290                  | 1000            | 28             | Vertical     | 1000            | 373        |
| 263.6                | 0               | 6              |              | 1023.57         | 1560       |
| 261                  | 2749.91         | 21             |              | 0               | 1040       |
| 233                  | 2799.89         | 34             | Horizontal   | 23000           | 800        |
| 233                  | 2799.89         | 11             |              | 0               | 840        |
| 215                  | 2499.97         | 32             | Horizontal   | 1400            | 1240       |
| 180                  | 1499.77         | 48             | Vertical     | 1000            | 1400       |
| 134.19               | 0               | 21             | Horizontal   | 0               | 100        |
| 132.5                | 0               | 4              |              | 0               | 1250       |

Nenhum caso selecionado

**Figura 29:** Exemplo da utilização da ferramenta de busca de casos. Neste exemplo, os casos que possuem o atributo condições ambientais igual a amena são exibidos em destaque

Na Figura 30 é exibida a ferramenta de manutenção dos conjuntos nebulosos, termos linguísticos dos conjuntos e suas funções de pertinência. Nela, o usuário pode adicionar ou remover termos linguísticos dos conjuntos nebulosos, modificar suas funções de pertinência, atribuindo os valores utilizados pelas funções trapezoidais. Desse modo, a representação do conhecimento pode ser refinada para ajustar os resultados do sistema.



**Figura 30:** Ferramenta de manutenção dos conjuntos nebulosos.

A ferramenta de inicialização da base de casos, início do ciclo do RBC e recuperação dos casos similares é ilustrada na Figura 31. Nela, pode-se observar que os atributos de representação dos casos (índices) são exibidos para que o usuário forneça os valores respectivos ao novo caso de consulta, juntamente com o peso de cada atributo. Uma vez que os valores desejados dos atributos foram fornecidos, o sistema realiza o processo de recuperação dos casos similares, apresentando o resultado para a avaliação do usuário. Neste exemplo, o sistema está recuperando os 5 casos mais similares ao problema informado, exibindo as informações dos casos recuperados, suas soluções e similaridades.

RBCS - Administração

0 / 0

Páginas

BIBLIOTECA

Sistemas de... 377
Plataformas 330
Casos 377
Soluções 177

MARCADORES
FERRAMENTAS
Fuzzy
CBE

Área do Reservatório
111
Peso
1

Profundidade Média
3200.1
Peso
1

Número de Poços
9
Peso
1

Tipo de Poço
Horizontal
Peso
1

Vazão diária / poço
2000
Peso
1

Lâmina d'água
1290
Peso
1

Condição Ambiental
Amenas
Peso
1

Distância da Costa
125
Peso
1

Infra-estrutura Próxima
Sim
Peso
1

Capacidade de Armazenamento
1700
Peso
1

Comparar Base de Casos
Resultado

Casos

| id | Área do Rese | Profundidade | Número de P | Tipo de Poço | Vazão Diária | Lâmina d'água |
|----|--------------|--------------|-------------|--------------|--------------|---------------|
| 0  | 111          | 3200.1       | 9           | Horizontal   | 2000         | 1290          |

Casos Similares

| id  | Similaridade   | Área do Rese | Profundidade | Número de P | Tipo de Poço | Vazão Diária |
|-----|----------------|--------------|--------------|-------------|--------------|--------------|
| 42  | 1              | 111          | 3200.1       | 9           | Horizontal   | 2000         |
| 219 | 1              | 111          | 3200.1       | 9           | Horizontal   | 2000         |
| 178 | 0.636951393267 | 111          | 3200.1       | 17          | Horizontal   | 2000         |
| 125 | 0.540152806505 | 111          | 3200.1       | 19          | Horizontal   | 2000         |
| 179 | 0.51037863432  | 111          | 3200.1       | 8           | Horizontal   | 2000         |

Soluções

| Arranjo de Poço | Manifold | Tipo de UEP | Tipo de Anco... | Tipo de Riser | Transporte    |
|-----------------|----------|-------------|-----------------|---------------|---------------|
| Satélite        | Não      | PPSO        |                 | Rigid         | Tanker / Duto |

Figura 31: Ferramenta de início do RBC e da recuperação de casos similares.

#### 4.2.5.1 Testes e Avaliação

Para poder avaliar o desempenho do RIGS, testes foram realizados com casos reais para comparar suas soluções com as soluções recomendadas pelo sistema. Foram considerados 23 casos reais, disponíveis na base de caso. Esses 23 casos foram escolhidos por possuírem sua descrição completa, ou seja, possuem informações para todos os atributos de indexação do caso.

Os testes se deram através dos seguintes passos:

1. O caso consulta, que servirá de entrada para o sistema, é selecionado dentre os 23 casos reais.
2. Os atributos da descrição do caso consulta são informados ao sistema.
3. Inicia-se a etapa de Recuperação do RBC, buscando os 4 casos mais similares na base de casos.
4. As descrições e soluções dos casos recuperados, bem como seu valor de similaridade (0 significando dissimilaridade total, e 1 significando similaridade total), são apresentados ao usuário.
5. O caso mais similar (valor mais próximo a 1) ao caso consulta é então recomendado como solução.
6. O próximo caso consulta é selecionado e o processo é reiniciado.

As tabelas a seguir mostram os resultados obtidos. As soluções recomendadas pelo sistema são comparadas com as soluções reais do caso consulta, e as diferenças são destacadas na cor cinza para facilitar a avaliação.

É importante salientar que a diferença foi observada e destacada quando os valores dos atributos da solução são diferentes, não importando se o seus significados são semelhantes, pois para tal constatação é necessário o conhecimento e avaliação de um especialista.

Isso se deve pela falta de especificação e padronização dos valores permitidos dos atributos dos casos na base de casos, fazendo com que um mesmo valor possua formas distintas de escrita.

Outro fator importante na avaliação da diferença entre os valores é que não foi considerado diferença quando um atributo da solução possui outros valores além do valor da solução real. Para exemplificar, o atributo “Tipo de *Riser*” pode ser multivalorado, possuindo, por exemplo, os valores “Flexível” e “Rígido” (denotado por “Flexível / Rígido”). Sendo assim, se o “Tipo de *Riser*” da solução real for “Flexível” e o da solução recomendada conter ao menos “Flexível”, como exemplo “Flexível / Rígido”, ambos são considerados equivalentes. Esta abordagem se dá para todos os atributos que podem possuir mais que um valor.

Na prática, a possibilidade de múltiplos valores do atributo significa que, para um determinado atributo, existe mais de uma solução possível ou então que várias soluções foram utilizadas para o mesmo problema. Por exemplo, num mesmo sistema de produção marítimo podem ser utilizados tanto *risers* flexíveis quanto rígidos. Nessa situação é necessário o conhecimento de especialistas para tomar a decisão mais adequada.

Outro ponto importante na constatação da diferença entre os atributos é que quando um atributo da solução real não possui valor determinado, todos os valores do atributo na solução recomendada são considerados equivalentes. Isso se dá pela falta de informações das soluções dos casos na base de casos, impedindo que a comparação seja realizada.

Os casos são identificados por um número, correspondente ao seu identificador na base de casos, para facilitar a avaliação dos testes.

A legenda utilizada nas tabelas é apresentada a seguir:

#### **Transporte**

|            |                  |
|------------|------------------|
| T: Tanker  | TS: Tandem Stern |
| D: Duto    | SS: Side by Side |
| N: Navio   | TB: Tandem Bow   |
| TA: Tandem |                  |

#### **Tipo de Ancoragem**

|                       |                                   |
|-----------------------|-----------------------------------|
| SM: Spread Mooring    | TT: Tendões Tensionados           |
| ET: External Turret   | CALM: Catenary Anchor Leg Mooring |
| IT: Internal Turret   | SALM: Single Anchor Leg Mooring   |
| CL: Catenary Legs     | DP: Dynamic Positioning           |
| JSY: Jacket Soft Yoke | TL: Taut Leg System               |
| TC: Taut Catenary     |                                   |

#### **Tipo de UEP**

|             |                                             |
|-------------|---------------------------------------------|
| TLP/4 CC:   | Tension Leg Plataform/4 Column Conventional |
| FPSO:       | Floating Production Storage Offloading      |
| FSO:        | Floating Storage Offloading                 |
| Semi ou SS: | Semisubmersível                             |

**Tabela 4:** Resultado da recuperação dos casos similares ao caso consulta 6

|                                            | CASO CONSULTA   | CASOS SIMILARES – RIGS |               |          |         |
|--------------------------------------------|-----------------|------------------------|---------------|----------|---------|
| CASOS (Identificador)                      | 6               | 170                    | 355           | 222      | 194     |
| Similaridade                               |                 | 0.2537                 | 0.2372        | 0.2362   | 0.2154  |
| Atributos de indexação                     | DESCRIÇÃO       |                        |               |          |         |
| Área do Reservatório (Km <sup>2</sup> )    | 13              | 111                    | 53            | 41       |         |
| Profundidade Média (m)                     | 3500            | 3200.1                 | 4500          | 2590.8   | 3849.42 |
| Número de Poços                            | 2               |                        | 10            | 28       | 9       |
| Tipo de Poço                               | Horizontal      | Horizontal             | Horizontal    | Vertical |         |
| Vazão Diária / Poço (m <sup>3</sup> / d)   | 2900            | 2000                   | 5000          | 2000     |         |
| Lâmina d'água (m)                          | 800             | 815                    | 910           | 800      | 120     |
| Condição Ambiental                         | Amena           | Amena                  | Amena         | Amena    | Amena   |
| Distância da Costa (Km)                    | 50              | 125                    | 306           | 41       |         |
| Infraestrutura Próxima                     | Não             | Sim                    | Sim           | Não      |         |
| Capacidade Armazenamento (m <sup>3</sup> ) | 550             | 1600                   | 15000         | 2000     | 340     |
| Atributos da solução                       | SOLUÇÃO         |                        |               |          |         |
| Arranjo de Poço                            | Satélite        | Satélite               | Satélite      | Satélite |         |
| Manifold                                   | Sim             | Não                    | Sim           | Sim      |         |
| Tipo de UEP                                | FPSO            | FSO                    | TLP/4 C C TLP | FPSO     | FPSO    |
| Tipo de Ancoragem                          |                 |                        |               |          |         |
| Tipo de Riser                              | Flexível        |                        | Rígido        | Flexível |         |
| Escoamento                                 | T / SS / SS / N | Duto                   | Duto          | T / N    |         |

**Tabela 5:** Resultado da recuperação dos casos similares ao caso consulta 27

|                                            | CASO CONSULTA | CASOS SIMILARES – RIGS |                   |          |            |
|--------------------------------------------|---------------|------------------------|-------------------|----------|------------|
| CASOS (Identificador)                      | 27            | 185                    | 51                | 235      | 170        |
| Similaridade                               |               | 0.7260                 | 0.3346            | 0.2582   | 0.2022     |
| Atributos de indexação                     | DESCRIÇÃO     |                        |                   |          |            |
| Área do Reservatório (Km <sup>2</sup> )    | 80            | 80                     | 40                | 61       | 111        |
| Profundidade Média (m)                     | 1760          | 1760                   | 876.3             | 3400     | 3200.1     |
| Número de Poços                            | 15            | 1                      | 21                | 11       |            |
| Tipo de Poço                               | Horizontal    | Horizontal             | Horizontal        | Vertical | Horizontal |
| Vazão Diária / Poço (m <sup>3</sup> / d)   | 1400          | 1400                   | 800               | 1900     | 2000       |
| Lâmina d'água (m)                          | 125           | 125                    | 104               | 90       | 815        |
| Condição Ambiental                         | Severa        | Severa                 | Severa            | Severa   | Amena      |
| Distância da Costa (Km)                    | 165           | 165                    | 134               | 195      | 125        |
| Infraestrutura Próxima                     | Sim           | Sim                    | Não               | Não      | Sim        |
| Capacidade Armazenamento (m <sup>3</sup> ) | 380           | 190                    | 554               | 630      | 1600       |
| Atributos da solução                       | SOLUÇÃO       |                        |                   |          |            |
| Arranjo de Poço                            | Satélite      | Satélite               | Satélite          | Satélite | Satélite   |
| Manifold                                   | Sim           | Sim                    | Sim               | Sim      | Não        |
| Tipo de UEP                                | FPSO          | FPSO                   | FPSO              | FPSO     | FSO        |
| Tipo de Ancoragem                          |               |                        |                   |          |            |
| Tipo de Riser                              | Flexível      | Flexível               | Flexível / Rígido | Flexível |            |
| Escoamento                                 | TS / D        | Duto                   | TS                | Navio    | Duto       |

**Tabela 6:** Resultado da recuperação dos casos similares ao caso consulta 42

|                               | CASO CONSULTA  | CASOS SIMILARES – RIGS |            |            |            |
|-------------------------------|----------------|------------------------|------------|------------|------------|
| CASOS (Identificador)         | 42             | 219                    | 178        | 175        | 170        |
| Similaridade                  |                | 1                      | 0.6297     | 0.5312     | 0.5094     |
| Atributos de indexação        | DESCRIÇÃO      |                        |            |            |            |
| Área do Reservatório (Km²)    | 111            | 111                    | 111        | 111        | 111        |
| Profundidade Média (m)        | 3200.1         | 3200.1                 | 3200.1     | 3200.1     | 3200.1     |
| Número de Poços               | 9              | 1                      | 17         | 19         |            |
| Tipo de Poço                  | Horizontal     | Horizontal             | Horizontal | Horizontal | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 2000           | 2000                   | 2000       | 2000       | 2000       |
| Lâmina d'água (m)             | 1290           | 1853                   | 1600       | 1400       | 815        |
| Condição Ambiental            | Amena          | Amena                  | Amena      | Amena      | Amena      |
| Distância da Costa (Km)       | 125            | 125                    | 125        | 125        | 125        |
| Infraestrutura Próxima        | Sim            | Sim                    | Sim        | Sim        | Sim        |
| Capacidade Armazenamento (m³) | 1700           | 306                    | 970        | 1500       | 1600       |
| Atributos da solução          | SOLUÇÃO        |                        |            |            |            |
| Arranjo de Poço               |                |                        |            |            |            |
| Manifold                      | Não            | Não                    | Não        | Não        | Não        |
| Tipo de UEP                   | FPSO           | FPSO                   | FPSO       | FPSO       | FSO        |
| Tipo de Ancoragem             |                |                        |            |            |            |
| Tipo de Riser                 |                | Rígido                 |            |            |            |
| Escoamento                    | T / TS / T / D | T / D                  | Duto       | TS / T / D | Duto       |

**Tabela 7:** Resultado da recuperação dos casos similares ao caso consulta 51

|                               | CASO CONSULTA     | CASOS SIMILARES – RIGS |            |            |            |
|-------------------------------|-------------------|------------------------|------------|------------|------------|
| CASOS (Identificador)         | 51                | 27                     | 185        | 276        | 193        |
| Similaridade                  |                   | 0.3346                 | 0.2558     | 0.2161     | 0.1990     |
| Atributos de indexação        | DESCRIÇÃO         |                        |            |            |            |
| Área do Reservatório (Km²)    | 40                | 80                     | 80         | 24         | 134.19     |
| Profundidade Média (m)        | 876.30            | 1760                   | 1760       | 3500       |            |
| Número de Poços               | 21                | 15                     | 1          | 8          | 21         |
| Tipo de Poço                  | Horizontal        | Horizontal             | Horizontal | Horizontal | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 800               | 1400                   | 1400       | 800        |            |
| Lâmina d'água (m)             | 104               | 125                    | 125        | 79         | 100        |
| Condição Ambiental            | Severa            | Severa                 | Severa     | Severa     | Amena      |
| Distância da Costa (Km)       | 134               | 165                    | 165        | 3200.1     | 100        |
| Infraestrutura Próxima        | Não               | Sim                    | Satélite   | Não        |            |
| Capacidade Armazenamento (m³) | 554               | 380                    | 190        | 8000       | 1600       |
| Atributos da solução          | SOLUÇÃO           |                        |            |            |            |
| Arranjo de Poço               | Satélite          | Satélite               | Satélite   | Satélite   |            |
| Manifold                      | Sim               | Sim                    | Sim        | Não        |            |
| Tipo de UEP                   | FPSO              | FPSO                   | FPSO       | Semi       | FPSO       |
| Tipo de Ancoragem             |                   |                        |            |            |            |
| Tipo de Riser                 | Flexível / Rígido | Flexível               | Flexível   | Rígido     |            |
| Escoamento                    | TS                | TS / D                 | Duto       | Navio      | Aliviador  |

**Tabela 8:** Resultado da recuperação dos casos similares ao caso consulta 98

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |         |                     |        |
|-------------------------------|---------------|------------------------|---------|---------------------|--------|
| CASOS (Identificador)         | 98            | 130                    | 225     | 173                 | 176    |
| Similaridade                  |               | 0.3241                 | 0.2135  | 0.1793              | 0.1756 |
| Atributos de indexação        | DESCRIÇÃO     |                        |         |                     |        |
| Área do Reservatório (Km²)    | 180           | 290                    | 263.6   | 215                 | 132.5  |
| Profundidade Média (m)        | 1499.77       | 1000                   |         | 2499.97             |        |
| Número de Poços               | 48            | 28                     | 6       | 32                  |        |
| Tipo de Poço                  | Vertical      | Vertical               |         | Horizontal          |        |
| Vazão Diária / Poço (m³ / d)  | 1000          | 1000                   | 1023.57 | 1400                |        |
| Lâmina d'água (m)             | 1400          | 373                    | 1560    | 1240                | 1250   |
| Condição Ambiental            | Amena         | Amena                  | Amena   | Amena               | Amena  |
| Distância da Costa (Km)       | 210           | 93                     | 37      | 120                 | 85     |
| Infraestrutura Próxima        | Não           | Não                    |         | Sim                 |        |
| Capacidade Armazenamento (m³) | 2000          | 1400                   | 300     | 1600                | 2000   |
| Atributos da solução          | SOLUÇÃO       |                        |         |                     |        |
| Arranjo de Poço               | Satélite      | Satélite               |         | Satélite            |        |
| Manifold                      | Sim           | Sim                    |         | 3                   |        |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO    | FPSO                | FPSO   |
| Tipo de Ancoragem             |               |                        |         |                     |        |
| Tipo de Riser                 | Tower         | Flexível               |         |                     |        |
| Escoamento                    | T / TB / N    | T / TS / N             |         | T / TS / T / TA / D |        |

**Tabela 9:** Resultado da recuperação dos casos similares ao caso consulta 130

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |                   |            |         |
|-------------------------------|---------------|------------------------|-------------------|------------|---------|
| CASOS (Identificador)         | 130           | 98                     | 51                | 193        | 225     |
| Similaridade                  |               | 0.3241                 | 0.1925            | 0.1776     | 0.1720  |
| Atributos de indexação        | DESCRIÇÃO     |                        |                   |            |         |
| Área do Reservatório (Km²)    | 290           | 180                    | 40                | 134.19     | 263.6   |
| Profundidade Média (m)        | 1000          | 1499.77                | 876.3             |            |         |
| Número de Poços               | 28            | 48                     | 21                | 21         | 6       |
| Tipo de Poço                  | Vertical      | Vertical               | Horizontal        | Horizontal |         |
| Vazão Diária / Poço (m³ / d)  | 1000          | 1000                   | 800               |            | 1023.57 |
| Lâmina d'água (m)             | 373           | 1400                   | 104               | 100        | 1560    |
| Condição Ambiental            | Amena         | Amena                  | Severa            | Amena      | Amena   |
| Distância da Costa (Km)       | 93            | 210                    | 134               | 100        | 37      |
| Infraestrutura Próxima        | Não           | Não                    | Não               |            |         |
| Capacidade Armazenamento (m³) | 1400          | 2000                   | 554               | 1600       | 300     |
| Atributos da solução          | SOLUÇÃO       |                        |                   |            |         |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite          |            |         |
| Manifold                      | Sim           | Sim                    | Sim               |            |         |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO              | FPSO       | FPSO    |
| Tipo de Ancoragem             |               |                        |                   |            |         |
| Tipo de Riser                 | Flexível      | Tower                  | Flexível / Rígido |            |         |
| Escoamento                    | T / TS / N    | T / TB / N             | TS                | Aliviador  |         |

**Tabela 10:** Resultado da recuperação dos casos similares ao caso consulta 162

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |            |          |          |
|-------------------------------|---------------|------------------------|------------|----------|----------|
| CASOS (Identificador)         | 162           | 196                    | 303        | 171      | 166      |
| Similaridade                  |               | 0.5439                 | 0.3444     | 0.3308   | 0.3220   |
| Atributos de indexação        | DESCRIÇÃO     |                        |            |          |          |
| Área do Reservatório (Km²)    | 455           | 455                    | 455        | 132      | 132.5    |
| Profundidade Média (m)        | 3631.08       | 3631.08                | 3631.08    | 3399.13  | 3399.13  |
| Número de Poços               | 36            | 12                     | 18         |          | 29       |
| Tipo de Poço                  | Horizontal    | Horizontal             | Horizontal | Vertical | Vertical |
| Vazão Diária / Poço (m³ / d)  | 300           | 300                    | 300        | 100      | 100      |
| Lâmina d'água (m)             | 330           | 224                    | 230        | 190      | 905      |
| Condição Ambiental            | Amena         | Amena                  | Amena      | Amena    | Amena    |
| Distância da Costa (Km)       | 110           | 110                    | 110        | 110      | 110      |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim        | Sim      | Sim      |
| Capacidade Armazenamento (m³) | 2000          | 340                    |            | 1600     | 2000     |
| Atributos da solução          | SOLUÇÃO       |                        |            |          |          |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite   | Satélite | Satélite |
| Manifold                      | Sim           | Sim                    | Sim        | Sim      | Sim      |
| Tipo de UEP                   | FPSO          | FPSO                   | Semi / SS  | FSO      | FPSO     |
| Tipo de Ancoragem             |               |                        |            |          |          |
| Tipo de Riser                 | Flexível      |                        | Flexível   | Flexível | Flexível |
| Escoamento                    | T / TS        | Duto                   | Duto       | TS       | T / TS   |

**Tabela 11:** Resultado da recuperação dos casos similares ao caso consulta 164

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |                 |            |          |
|-------------------------------|---------------|------------------------|-----------------|------------|----------|
| CASOS (Identificador)         | 164           | 171                    | 163             | 165        | 166      |
| Similaridade                  |               | 0.6252                 | 0.6173          | 0.4721     | 0.4369   |
| Atributos de indexação        | DESCRIÇÃO     |                        |                 |            |          |
| Área do Reservatório (Km²)    | 132           | 132                    | 132             | 132        | 132.5    |
| Profundidade Média (m)        | 3399.13       | 3399.13                | 3399.13         | 3399.13    | 3399.13  |
| Número de Poços               | 8             |                        |                 | 20         | 29       |
| Tipo de Poço                  | Vertical      | Vertical               | Vertical        | Vertical   | Vertical |
| Vazão Diária / Poço (m³ / d)  | 100           | 100                    | 100             | 100        | 100      |
| Lâmina d'água (m)             | 780           | 190                    | 160             | 860        | 905      |
| Condição Ambiental            | Amena         | Amena                  | Amena           | Amena      | Amena    |
| Distância da Costa (Km)       | 110           | 110                    | 110             | 110        | 110      |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim             | Sim        | Sim      |
| Capacidade Armazenamento (m³) | 2000          | 1600                   | 2000            | 1750       | 2000     |
| Atributos da solução          | SOLUÇÃO       |                        |                 |            |          |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite        | Satélite   | Satélite |
| Manifold                      | Sim           | Sim                    | Sim             | Sim        | Sim      |
| Tipo de UEP                   | FPSO          | FPSO                   | FSO / FPSO      | FPSO       | FPSO     |
| Tipo de Ancoragem             |               |                        |                 |            |          |
| Tipo de Riser                 | Flexível      | Flexível               | Flexível Lazy S | Flexível   | Flexível |
| Escoamento                    | T / TS        | TS                     | TS / D          | T / TS / D | T / TS   |

**Tabela 12:** Resultado da recuperação dos casos similares ao caso consulta 165

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |          |          |                 |
|-------------------------------|---------------|------------------------|----------|----------|-----------------|
| CASOS (Identificador)         | 165           | 166                    | 164      | 171      | 163             |
| Similaridade                  |               | 0.8953                 | 0.4721   | 0.4080   | 0.4043          |
| Atributos de indexação        | DESCRIÇÃO     |                        |          |          |                 |
| Área do Reservatório (Km²)    | 132           | 132                    | 132      | 132      | 132             |
| Profundidade Média (m)        | 3399.13       | 3399.13                | 3399.13  | 3399.13  | 3399.13         |
| Número de Poços               | 20            | 29                     | 8        |          |                 |
| Tipo de Poço                  | Vertical      | Vertical               | Vertical | Vertical | Vertical        |
| Vazão Diária / Poço (m³ / d)  | 100           | 100                    | 100      | 100      | 100             |
| Lâmina d'água (m)             | 860           | 905                    | 780      | 190      | 160             |
| Condição Ambiental            | Amena         | Amena                  | Amena    | Amena    | Amena           |
| Distância da Costa (Km)       | 110           | 110                    | 110      | 110      | 110             |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim      | Sim      | Sim             |
| Capacidade Armazenamento (m³) | 1750          | 2000                   | 2000     | 1600     | 2000            |
| Atributos da solução          | SOLUÇÃO       |                        |          |          |                 |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite | Satélite | Satélite        |
| Manifold                      | Sim           | Sim                    | Sim      | Sim      | Sim             |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO     | FSO      | FSO / FPSO      |
| Tipo de Ancoragem             |               |                        |          |          |                 |
| Tipo de Riser                 | Flexível      | Flexível               | Flexível | Flexível | Flexível Lazy S |
| Escoamento                    | T / TS / D    | T / TS                 | T / TS   | TS       | TS / D          |

**Tabela 13:** Resultado da recuperação dos casos similares ao caso consulta 166

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |          |          |                 |
|-------------------------------|---------------|------------------------|----------|----------|-----------------|
| CASOS (Identificador)         | 166           | 165                    | 164      | 171      | 163             |
| Similaridade                  |               | 0.8953                 | 0.4369   | 0.4033   | 0.3996          |
| Atributos de indexação        | DESCRIÇÃO     |                        |          |          |                 |
| Área do Reservatório (Km²)    | 132           | 132                    | 132      | 132      | 132.5           |
| Profundidade Média (m)        | 3399.13       | 3399.13                | 3399.13  | 3399.13  | 3399.13         |
| Número de Poços               | 29            | 20                     |          |          |                 |
| Tipo de Poço                  | Vertical      | Vertical               | Vertical | Vertical | Vertical        |
| Vazão Diária / Poço (m³ / d)  | 100           | 100                    | 100      | 100      | 100             |
| Lâmina d'água (m)             | 905           | 860                    | 780      | 190      | 160             |
| Condição Ambiental            | Amena         | Amena                  | Amena    | Amena    | Amena           |
| Distância da Costa (Km)       | 110           | 110                    | 110      | 110      | 110             |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim      | Sim      | Sim             |
| Capacidade Armazenamento (m³) | 2000          | 1750                   | 2000     | 1600     | 2000            |
| Atributos da solução          | SOLUÇÃO       |                        |          |          |                 |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite | Satélite | Satélite        |
| Manifold                      | Sim           | Sim                    | Sim      | Sim      | Sim             |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO     | FSO      | FSO / FPSO      |
| Tipo de Ancoragem             |               |                        |          |          |                 |
| Tipo de Riser                 | Flexível      | Flexível               | Flexível | Flexível | Flexível Lazy S |
| Escoamento                    | T / TS        | T / TS / D             | T / TS   | TS       | TS / D          |

**Tabela 14:** Resultado da recuperação dos casos similares ao caso consulta 168

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |                     |            |            |
|-------------------------------|---------------|------------------------|---------------------|------------|------------|
| CASOS (Identificador)         | 168           | 170                    | 173                 | 175        | 178        |
| Similaridade                  |               | 0.4311                 | 0.4094              | 0.3937     | 0.3846     |
| Atributos de indexação        | DESCRIÇÃO     |                        |                     |            |            |
| Área do Reservatório (Km²)    | 233           | 111                    | 215                 | 111        | 111        |
| Profundidade Média (m)        | 2799.89       | 3200.1                 | 2499.97             | 3200.1     | 3200.1     |
| Número de Poços               | 34            |                        | 32                  | 19         | 17         |
| Tipo de Poço                  | Horizontal    | Horizontal             | Horizontal          | Horizontal | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 23000         | 2000                   | 1400                | 2000       | 2000       |
| Lâmina d'água (m)             | 800           | 815                    | 1240                | 1400       | 1600       |
| Condição Ambiental            | Amena         | Amena                  | Amena               | Amena      | Amena      |
| Distância da Costa (Km)       | 95            | 125                    | 120                 | 125        | 125        |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim                 | Sim        | Sim        |
| Capacidade Armazenamento (m³) | 2000          | 1600                   | 1600                | 1500       | 970        |
| Atributos da solução          | SOLUÇÃO       |                        |                     |            |            |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite            | Satélite   | Satélite   |
| Manifold                      | Não           | Não                    | 3                   | Não        | Não        |
| Tipo de UEP                   | FPSO          | FSO                    | FPSO                | FPSO       | FPSO       |
| Tipo de Ancoragem             |               |                        |                     |            |            |
| Tipo de Riser                 | Flexível      |                        |                     |            |            |
| Escoamento                    | T / TS        | Duto                   | T / TS / T / TA / D | TS / T / D | Duto       |

**Tabela 15:** Resultado da recuperação dos casos similares ao caso consulta 219

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |            |            |            |
|-------------------------------|---------------|------------------------|------------|------------|------------|
| CASOS (Identificador)         | 219           | 42                     | 178        | 175        | 170        |
| Similaridade                  |               | 1                      | 0.6297     | 0.5312     | 0.5094     |
| Atributos de indexação        | DESCRIÇÃO     |                        |            |            |            |
| Área do Reservatório (Km²)    | 111           | 111                    | 111        | 111        | 111        |
| Profundidade Média (m)        | 3200.10       | 3200.1                 | 3200.1     | 3200.1     | 3200.1     |
| Número de Poços               | 1             | 9                      | 17         | 19         |            |
| Tipo de Poço                  | Horizontal    | Horizontal             | Horizontal | Horizontal | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 2000          | 2000                   | 2000       | 2000       | 2000       |
| Lâmina d'água (m)             | 1853          | 1290                   | 1600       | 1400       | 815        |
| Condição Ambiental            | Amena         | Amena                  | Amena      | Amena      | Amena      |
| Distância da Costa (Km)       | 125           | 125                    | 125        | 125        | 125        |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim        | Sim        | Sim        |
| Capacidade Armazenamento (m³) | 306           | 1700                   | 970        | 1500       | 1600       |
| Atributos da solução          | SOLUÇÃO       |                        |            |            |            |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite   | Satélite   | Satélite   |
| Manifold                      | Não           | Não                    | Não        | Não        | Não        |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO       | FPSO       | FSO        |
| Tipo de Ancoragem             |               |                        |            |            |            |
| Tipo de Riser                 |               |                        |            |            |            |
| Escoamento                    | Duto          | T / TS / T / D         | Duto       | TS / T / D | Duto       |

**Tabela 16:** Resultado da recuperação dos casos similares ao caso consulta 173

|                               | CASO CONSULTA       | CASOS SIMILARES – RIGS |            |            |            |
|-------------------------------|---------------------|------------------------|------------|------------|------------|
| CASOS (Identificador)         | 173                 | 174                    | 175        | 178        | 168        |
| Similaridade                  |                     | 0.5133                 | 0.5086     | 0.4955     | 0.4094     |
| Atributos de indexação        | DESCRIÇÃO           |                        |            |            |            |
| Área do Reservatório (Km²)    | 215                 | 457.69                 | 111        | 111        | 233        |
| Profundidade Média (m)        | 2499.97             | 2749.91                | 3200.1     | 3200.1     | 2799.89    |
| Número de Poços               | 32                  | 22                     | 19         | 17         | 37         |
| Tipo de Poço                  | Horizontal          | Horizontal             | Horizontal | Horizontal | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 1400                | 1576.92                | 2000       | 2000       | 23000      |
| Lâmina d'água (m)             | 1240                | 1080                   | 1400       | 1600       | 800        |
| Condição Ambiental            | Amena               | Amena                  | Amena      | Amena      | Amena      |
| Distância da Costa (Km)       | 120                 | 120                    | 125        | 125        | 95         |
| Infraestrutura Próxima        | Sim                 |                        | Sim        | Sim        | Sim        |
| Capacidade Armazenamento (m³) | 1600                | 2000                   | 1500       | 970        | 2000       |
| Atributos da solução          | SOLUÇÃO             |                        |            |            |            |
| Arranjo de Poço               | Satélite            |                        | Satélite   | Satélite   | Satélite   |
| Manifold                      | 3                   |                        | Não        | Não        | Não        |
| Tipo de UEP                   | FPSO                | FPU                    | FPSO       | FPSO       | FPSO       |
| Tipo de Ancoragem             |                     |                        |            |            |            |
| Tipo de Riser                 |                     |                        |            |            | Flexível   |
| Escoamento                    | T / TS / T / TA / D |                        | TS / T / D | Duto       | T / TS     |

**Tabela 17:** Resultado da recuperação dos casos similares ao caso consulta 175

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |                |            |                |
|-------------------------------|---------------|------------------------|----------------|------------|----------------|
| CASOS (Identificador)         | 175           | 178                    | 42             | 219        | 308            |
| Similaridade                  |               | 0.8864                 | 0.5312         | 0.5312     | 0.5138         |
| Atributos de indexação        | DESCRIÇÃO     |                        |                |            |                |
| Área do Reservatório (Km²)    | 111           | 111                    | 111            | 111        | 111            |
| Profundidade Média (m)        | 3200.10       | 3200.1                 | 3200.1         | 3200.1     | 3200.1         |
| Número de Poços               | 19            | 17                     | 9              | 1          | 28             |
| Tipo de Poço                  | Horizontal    | Horizontal             | Horizontal     | Horizontal | Horizontal     |
| Vazão Diária / Poço (m³ / d)  | 2000          | 2000                   | 2000           | 2000       | 2000           |
| Lâmina d'água (m)             | 1400          | 1600                   | 1290           | 1853       | 1360           |
| Condição Ambiental            | Amena         | Amena                  | Amena          | Amena      | Amena          |
| Distância da Costa (Km)       | 125           | 125                    | 125            | 125        | 125            |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim            | Sim        | Sim            |
| Capacidade Armazenamento (m³) | 1500          | 970                    | 1700           | 306        | 30             |
| Atributos da solução          | SOLUÇÃO       |                        |                |            |                |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite       | Satélite   | Satélite       |
| Manifold                      | Não           | Não                    | Não            | Não        | Não            |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO           | FPSO       | Semi / SS      |
| Tipo de Ancoragem             |               |                        |                |            |                |
| Tipo de Riser                 |               |                        |                | Rígido     | Flexível / SCR |
| Escoamento                    | TS / T / D    | Duto                   | T / TS / T / D | T / D      | T / D          |

**Tabela 18:** Resultado da recuperação dos casos similares ao caso consulta 178

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |                |            |                     |
|-------------------------------|---------------|------------------------|----------------|------------|---------------------|
| CASOS (Identificador)         | 178           | 175                    | 42             | 219        | 173                 |
| Similaridade                  |               | 0.8864                 | 0.6297         | 0.6297     | 0.4955              |
| Atributos de indexação        | DESCRIÇÃO     |                        |                |            |                     |
| Área do Reservatório (Km²)    | 111           | 111                    | 111            | 111        | 215                 |
| Profundidade Média (m)        | 3200.1        | 3200.1                 | 3200.1         | 3200.1     | 2499.97             |
| Número de Poços               | 17            | 19                     | 9              | 1          | 32                  |
| Tipo de Poço                  | Horizontal    | Horizontal             | Horizontal     | Horizontal | Horizontal          |
| Vazão Diária / Poço (m³ / d)  | 2000          | 2000                   | 2000           | 2000       | 1400                |
| Lâmina d'água (m)             | 1600          | 1400                   | 1290           | 1853       | 1240                |
| Condição Ambiental            | Amena         | Amena                  | Amena          | Amena      | Amena               |
| Distância da Costa (Km)       | 125           | 125                    | 125            | 125        | 120                 |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim            | Sim        | Sim                 |
| Capacidade Armazenamento (m³) | 970           | 1500                   | 1700           | 306        | 1600                |
| Atributos da solução          | SOLUÇÃO       |                        |                |            |                     |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite       | Satélite   | Satélite            |
| Manifold                      | Não           | Não                    | Não            | Não        | 3                   |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO           | FPSO       | FPSO                |
| Tipo de Ancoragem             |               |                        |                |            |                     |
| Tipo de Riser                 |               |                        |                | Rígido     |                     |
| Escoamento                    | Duto          | TS / T / D             | T / TS / T / D | T / D      | T / TS / T / TA / D |

**Tabela 19:** Resultado da recuperação dos casos similares ao caso consulta 185

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |          |                   |            |
|-------------------------------|---------------|------------------------|----------|-------------------|------------|
| CASOS (Identificador)         | 185           | 27                     | 235      | 51                | 170        |
| Similaridade                  |               | 0.7260                 | 0.2790   | 0.2558            | 0.2256     |
| Atributos de indexação        | DESCRIÇÃO     |                        |          |                   |            |
| Área do Reservatório (Km²)    | 80            | 80                     | 61       | 40                | 111        |
| Profundidade Média (m)        | 1760          | 1760                   | 3400     | 876.3             | 3200.1     |
| Número de Poços               | 1             | 15                     | 11       | 21                |            |
| Tipo de Poço                  | Horizontal    | Horizontal             | Vertical | Horizontal        | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 1400          | 1400                   | 1900     | 800               | 2000       |
| Lâmina d'água (m)             | 125           | 125                    | 90       | 104               | 815        |
| Condição Ambiental            | Severa        | Severa                 | Severa   | Severa            | Amena      |
| Distância da Costa (Km)       | 165           | 165                    | 195      | 134               | 125        |
| Infraestrutura Próxima        | Sim           | Sim                    | Não      | Não               | Sim        |
| Capacidade Armazenamento (m³) | 190           | 380                    | 630      | 554               | 1600       |
| Atributos da solução          | SOLUÇÃO       |                        |          |                   |            |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite | Satélite          | Satélite   |
| Manifold                      | Sim           | Sim                    | Sim      | Sim               | Não        |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO     | FPSO              | FSO        |
| Tipo de Ancoragem             |               |                        |          |                   |            |
| Tipo de Riser                 | Flexível      | Flexível               | Flexível | Flexível / Rígido |            |
| Escoamento                    | Duto          | TS / D                 | Navio    | TS                | Duto       |

**Tabela 20:** Resultado da recuperação dos casos similares ao caso consulta 196

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |            |                 |          |
|-------------------------------|---------------|------------------------|------------|-----------------|----------|
| CASOS (Identificador)         | 196           | 171                    | 162        | 163             | 164      |
| Similaridade                  |               | 0.5694                 | 0.5439     | 0.5084          | 0.4027   |
| Atributos de indexação        | DESCRIÇÃO     |                        |            |                 |          |
| Área do Reservatório (Km²)    | 455           | 132                    | 455        | 132             | 132.5    |
| Profundidade Média (m)        | 3631.08       | 3399.13                | 3631.08    | 3399.13         | 3399.13  |
| Número de Poços               | 12            |                        | 36         |                 | 8        |
| Tipo de Poço                  | Horizontal    | Vertical               | Horizontal | Vertical        | Vertical |
| Vazão Diária / Poço (m³ / d)  | 300           | 100                    | 300        | 100             | 100      |
| Lâmina d'água (m)             | 224           | 190                    | 330        | 160             | 780      |
| Condição Ambiental            | Amena         | Amena                  | Amena      | Amena           | Amena    |
| Distância da Costa (Km)       | 110           | 110                    | 110        | 110             | 110      |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim        | Sim             | Sim      |
| Capacidade Armazenamento (m³) | 340           | 1600                   | 2000       | 2000            | 2000     |
| Atributos da solução          | SOLUÇÃO       |                        |            |                 |          |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite   | Satélite        | Satélite |
| Manifold                      | Sim           | Sim                    | Sim        | Sim             | Sim      |
| Tipo de UEP                   | FPSO          | FSO                    | FPSO       | FSO / FPSO      | FPSO     |
| Tipo de Ancoragem             |               |                        |            |                 |          |
| Tipo de Riser                 |               | Flexível               | Flexível   | Flexível Lazy S | Flexível |
| Escoamento                    | Duto          | TS                     | T / TS     | TS / D          | T / TS   |

**Tabela 21:** Resultado da recuperação dos casos similares ao caso consulta 222

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |            |            |         |
|-------------------------------|---------------|------------------------|------------|------------|---------|
| CASOS (Identificador)         | 222           | 6                      | 168        | 82         | 81      |
| Similaridade                  |               | 0.2362                 | 0.2158     | 0.1762     | 0.1712  |
| Atributos de indexação        | DESCRIÇÃO     |                        |            |            |         |
| Área do Reservatório (Km²)    | 41            | 13                     | 233        |            |         |
| Profundidade Média (m)        | 2590.8        | 3500                   | 2799.89    | 3359.66    | 2280.92 |
| Número de Poços               | 28            | 2                      | 34         | 18         | 2799.89 |
| Tipo de Poço                  | Vertical      | Horizontal             | Horizontal |            |         |
| Vazão Diária / Poço (m³ / d)  | 2000          | 2900                   | 23000      |            |         |
| Lâmina d'água (m)             | 800           | 800                    | 800        | 800        | 38      |
| Condição Ambiental            | Amena         | Amena                  | Amena      | Amena      | Amena   |
| Distância da Costa (Km)       | 41            | 50                     | 95         |            |         |
| Infraestrutura Próxima        | Não           | Não                    | Sim        |            |         |
| Capacidade Armazenamento (m³) | 2000          | 550                    | 2000       | 1900       | 1900    |
| Atributos da solução          | SOLUÇÃO       |                        |            |            |         |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite   |            |         |
| Manifold                      | Sim           | Sim                    | Não        |            |         |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO       | FPSO       | FPSO    |
| Tipo de Ancoragem             |               |                        |            |            |         |
| Tipo de Riser                 | Flexível      | Flexível               | Flexível   | Flexível   |         |
| Escoamento                    | T / N         | T / SS / SS / N        | T / TS     | T / TS / D |         |

**Tabela 22:** Resultado da recuperação dos casos similares ao caso consulta 235

|                               | CASO CONSULTA | CASOS SIMILARES - RIGS |            |              |            |
|-------------------------------|---------------|------------------------|------------|--------------|------------|
| CASOS (Identificador)         | 235           | 185                    | 27         | 355          | 276        |
| Similaridade                  |               | 0.2790                 | 0.2582     | 0.2090       | 0.1836     |
| Atributos de indexação        | DESCRIÇÃO     |                        |            |              |            |
| Área do Reservatório (Km²)    | 61            | 80                     | 80         | 53           | 24         |
| Profundidade Média (m)        | 3400          | 1760                   | 1760       | 4500         | 3500       |
| Número de Poços               | 11            | 1                      | 15         | 10           | 8          |
| Tipo de Poço                  | Vertical      | Horizontal             | Horizontal | Horizontal   | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 1900          | 1400                   | 1400       | 5000         | 800        |
| Lâmina d'água (m)             | 90            | 125                    | 125        | 910          | 79         |
| Condição Ambiental            | Severa        | Severa                 | Severa     | Severa       | Severa     |
| Distância da Costa (Km)       | 195           | 165                    | 165        | 306          | 320        |
| Infraestrutura Próxima        | Não           | Sim                    | Sim        | Sim          | Não        |
| Capacidade Armazenamento (m³) | 630           | 190                    | 380        | 15000        | 8000       |
| Atributos da solução          | SOLUÇÃO       |                        |            |              |            |
| Arranjo de Poço               | Satélite      | Satélite               | Satélite   | Agrupado     | Satélite   |
| Manifold                      | Sim           | Sim                    | Sim        | Sim          | Não        |
| Tipo de UEP                   | FPSO          | FPSO                   | FPSO       | TLP/4 CC TLP | Semi       |
| Tipo de Ancoragem             |               |                        |            |              |            |
| Tipo de Riser                 | Flexível      | Flexível               | Flexível   | Rígido       | Rígido     |
| Escoamento                    | Navio         | Duto                   | TS / D     | Duto         | Navio      |

**Tabela 23:** Resultado da recuperação dos casos similares ao caso consulta 276

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |              |          |                 |
|-------------------------------|---------------|------------------------|--------------|----------|-----------------|
| CASOS (Identificador)         | 276           | 51                     | 355          | 235      | 6               |
| Similaridade                  |               | 0.2161                 | 0.2104       | 0.1836   | 0.1717          |
| Atributos de indexação        | DESCRIÇÃO     |                        |              |          |                 |
| Área do Reservatório (Km²)    | 24            | 40                     | 53           | 61       | 132.5           |
| Profundidade Média (m)        | 3500          | 876.3                  | 4500         | 3400     | 3500            |
| Número de Poços               | 8             | 21                     | 10           | 11       | 2               |
| Tipo de Poço                  | Horizontal    | Horizontal             | Horizontal   | Vertical | Horizontal      |
| Vazão Diária / Poço (m³ / d)  | 800           | 800                    | 5000         | 1900     | 2900            |
| Lâmina d'água (m)             | 79            | 104                    | 910          | 90       | 800             |
| Condição Ambiental            | Severa        | Severa                 | Severa       | Severa   | Amena           |
| Distância da Costa (Km)       | 320           | 134                    | 306          | 195      | 50              |
| Infraestrutura Próxima        | Não           | Não                    | Sim          | Não      | Não             |
| Capacidade Armazenamento (m³) | 8000          | 554                    | 15000        | 630      | 550             |
| Atributos da solução          | SOLUÇÃO       |                        |              |          |                 |
| Arranjo de Poço               | Satélite      | Satélite               | Agrupado     | Satélite | Satélite        |
| Manifold                      | Não           | Sim                    | Sim          | Sim      | Sim             |
| Tipo de UEP                   | Semi          | FPSO                   | TLP/4 CC TLP | FPSO     | FPSO            |
| Tipo de Ancoragem             |               |                        |              |          |                 |
| Tipo de Riser                 | Rígido        | Flexível / Rígido      | Rígido       | Flexível | Flexível        |
| Escoamento                    | Navio         | TS                     | Duto         | Navio    | T / SS / SS / N |

**Tabela 24:** Resultado da recuperação dos casos similares ao caso consulta 308

|                               | CASO CONSULTA  | CASOS SIMILARES – RIGS |            |            |            |
|-------------------------------|----------------|------------------------|------------|------------|------------|
| CASOS (Identificador)         | 308            | 311                    | 175        | 312        | 178        |
| Similaridade                  |                | 0.5292                 | 0.5138     | 0.5017     | 0.4864     |
| Atributos de indexação        | DESCRIÇÃO      |                        |            |            |            |
| Área do Reservatório (Km²)    | 111            | 111                    | 111        | 111        | 111        |
| Profundidade Média (m)        | 3200.1         | 3200.1                 | 3200.1     | 3200.1     | 3200.1     |
| Número de Poços               | 28             | 27                     | 19         | 18         | 17         |
| Tipo de Poço                  | Horizontal     | Horizontal             | Horizontal | Horizontal | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 2000           | 2000                   | 2000       | 2000       | 2000       |
| Lâmina d'água (m)             | 1360           | 1800                   | 1400       | 1790       | 1600       |
| Condição Ambiental            | Amena          | Amena                  | Amena      | Amena      | Amena      |
| Distância da Costa (Km)       | 125            | 125                    | 125        | 125        | 125        |
| Infraestrutura Próxima        | Sim            | Sim                    | Sim        | Sim        | Sim        |
| Capacidade Armazenamento (m³) | 30             |                        | 1500       |            | 970        |
| Atributos da solução          | SOLUÇÃO        |                        |            |            |            |
| Arranjo de Poço               | Satélite       | Satélite               | Satélite   | Satélite   | Satélite   |
| Manifold                      | Não            | Não                    | Não        | Não        | Não        |
| Tipo de UEP                   | Semi / SS      | Semi / SS              | FPSO       | Semi / FPU | FPSO       |
| Tipo de Ancoragem             |                | Taunt Leg              |            |            |            |
| Tipo de Riser                 | Flexível / SCR |                        |            | SCR        |            |
| Escoamento                    | T / D          | Duto                   | TS / T / D | Duto       | Duto       |

**Tabela 25:** Resultado da recuperação dos casos similares ao caso consulta 328

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |                     |                     |                     |
|-------------------------------|---------------|------------------------|---------------------|---------------------|---------------------|
| CASOS (Identificador)         | 328           | 327                    | 323                 | 374                 | 329                 |
| Similaridade                  |               | 0.8127                 | 0.3062              | 0.2486              | 0.2329              |
| Atributos de indexação        | DESCRIÇÃO     |                        |                     |                     |                     |
| Área do Reservatório (Km²)    | 95            | 95                     | 62                  | 62                  | 25                  |
| Profundidade Média (m)        | 1547          | 1547                   | 2712.72             | 2697.48             | 3078.48             |
| Número de Poços               | 32            | 23                     | 26                  | 46                  | 18                  |
| Tipo de Poço                  | Horizontal    | Horizontal             | Horizontal          | Horizontal          | Horizontal          |
| Vazão Diária / Poço (m³ / d)  | 1000          | 1000                   | 1400                | 1400                | 1300                |
| Lâmina d'água (m)             | 340           | 320                    | 350                 | 335                 | 175                 |
| Condição Ambiental            | Severa        | Severa                 | Severa              | Severa              | Severa              |
| Distância da Costa (Km)       | 80            | 80                     | 210                 | 210                 | 145                 |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim                 | Sim                 | Sim                 |
| Capacidade Armazenamento (m³) | 3.8           |                        |                     |                     |                     |
| Atributos da solução          | SOLUÇÃO       |                        |                     |                     |                     |
| Arranjo de Poço               | Satélite      | Satélite               | Agrupado + Satélite | Agrupado + Satélite | Agrupado            |
| Manifold                      | Sim           | Sim                    | Não                 | Não                 | Não                 |
| Tipo de UEP                   | Semi / SS     | Semi / SS              | Semi / SS           | TLP/4 CC TLP        | Semi / SS           |
| Tipo de Ancoragem             | Wire Chain    | Catenary               | Spread / Wire Chain |                     | Spread / Wire Chain |
| Tipo de Riser                 | Flexível      | Flexível               | Flexível            | Rígido              | Flexível            |
| Escoamento                    | Duto          | Duto                   | Duto                | Duto                | Duto                |

**Tabela 26:** Resultado da recuperação dos casos similares ao caso consulta 355

|                               | CASO CONSULTA | CASOS SIMILARES – RIGS |            |                |            |
|-------------------------------|---------------|------------------------|------------|----------------|------------|
| CASOS (Identificador)         | 355           | 364                    | 170        | 42             | 219        |
| Similaridade                  |               | 0.4130                 | 0.2785     | 0.2760         | 0.2760     |
| Atributos de indexação        | DESCRIÇÃO     |                        |            |                |            |
| Área do Reservatório (Km²)    | 53            | 59                     | 111        | 111            | 111        |
| Profundidade Média (m)        | 4500          | 4297                   | 3200.1     | 3200.1         | 3200.1     |
| Número de Poços               | 10            | 14                     |            | 9              | 1023.57    |
| Tipo de Poço                  | Horizontal    | Horizontal             | Horizontal | Horizontal     | Horizontal |
| Vazão Diária / Poço (m³ / d)  | 5000          | 2000                   | 2000       | 2000           | 2000       |
| Lâmina d'água (m)             | 910           | 894                    | 815        | 1290           | 1853       |
| Condição Ambiental            | Severa        | Severa                 | Amena      | Amena          | Amena      |
| Distância da Costa (Km)       | 306           | 241                    | 125        | 125            | 125        |
| Infraestrutura Próxima        | Sim           | Sim                    | Sim        | Sim            | Sim        |
| Capacidade Armazenamento (m³) | 15000         |                        | 1600       | 1700           | 306        |
| Atributos da solução          | SOLUÇÃO       |                        |            |                |            |
| Arranjo de Poço               | Agrupado      | Agrupado               | Satélite   | Satélite       | Satélite   |
| Manifold                      | Sim           | Sim                    | Não        | Não            | Não        |
| Tipo de UEP                   | TLP/4 CC TLP  | TLP/4 CC TLP           | FSO / FPSO | FPSO           | FPSO       |
| Tipo de Ancoragem             |               |                        |            |                |            |
| Tipo de Riser                 | Rígido        | Rígido                 |            |                | Rígido     |
| Escoamento                    | Duto          | Duto                   | Duto       | T / TS / T / D | T / D      |

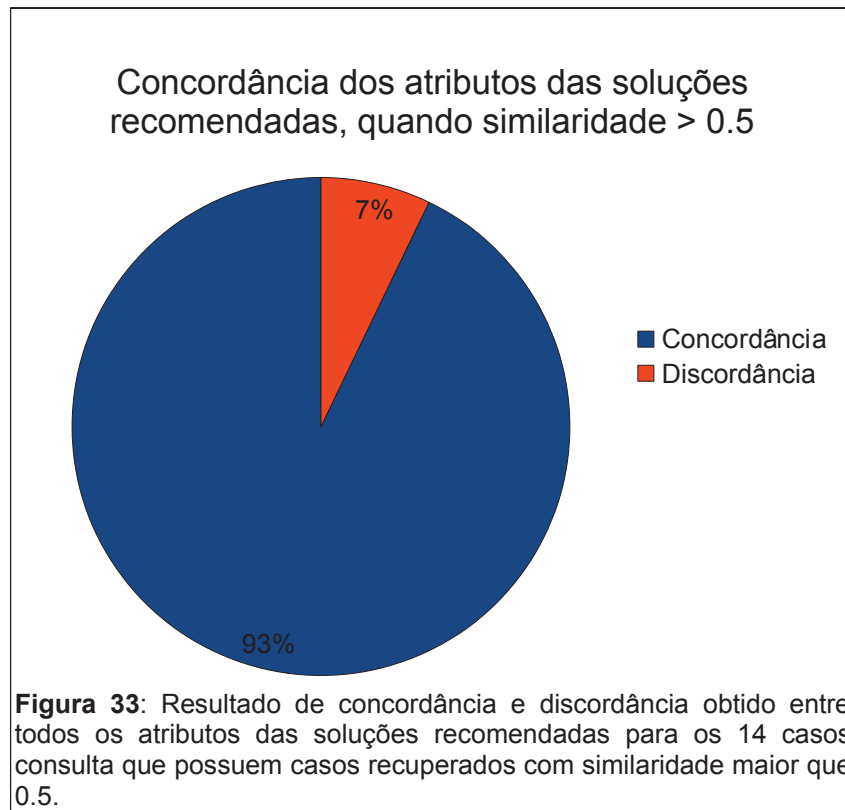
#### 4.2.5.2 Resultados

De acordo com a avaliação dos testes realizados, pode-se concluir que o sistema, de modo geral, apresenta resultados compatíveis com os casos reais.

Quando é considerada a concordância de cada atributo da solução do sistema, isto é, a comparação entre cada um dos atributos da solução recomendada com os atributos da solução real, de todos os 23 casos consulta, o resultado é bastante satisfatório, atingindo 89% de concordância, como ilustrado na Figura 32.

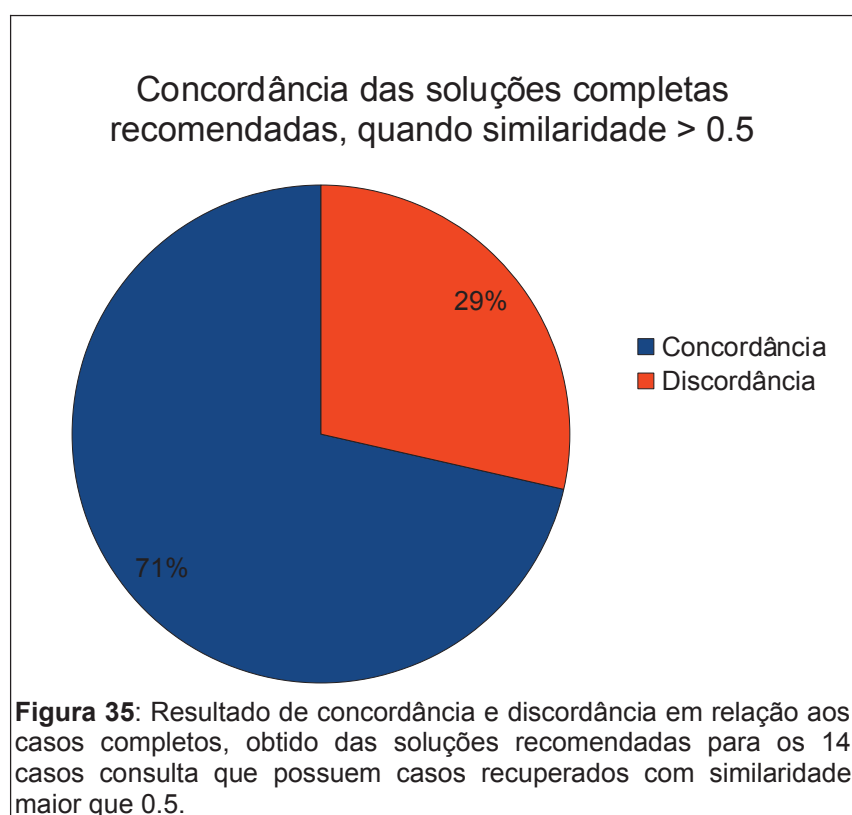
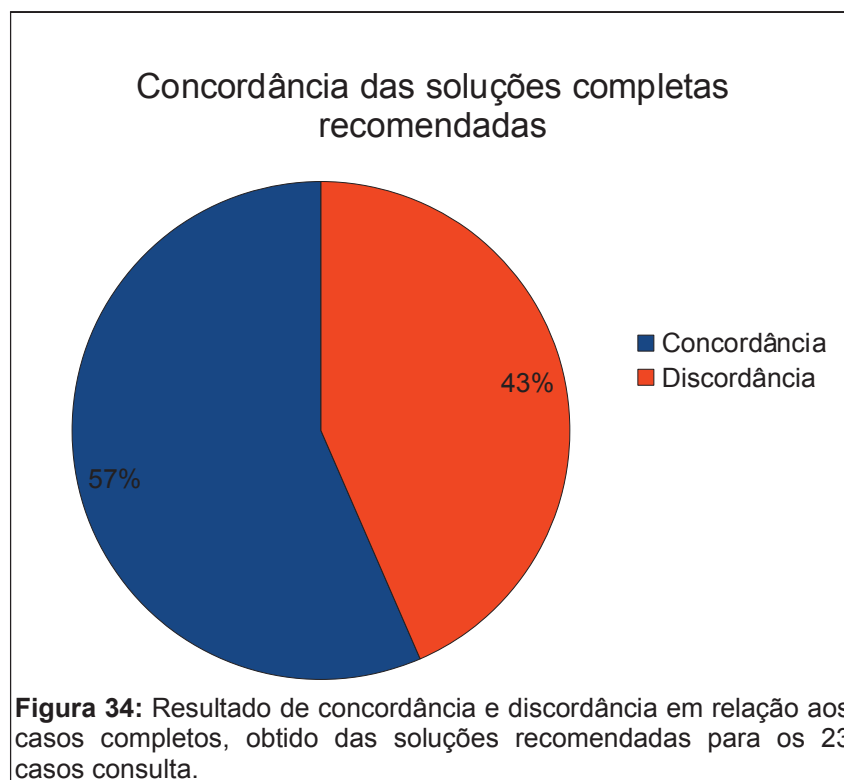
Se definirmos um limiar de similaridade de 0.5, que é o valor mínimo para um caso ser considerado similar ao outro, teremos 14 casos consulta que possuem casos recuperados com similaridade maior que 0.5. Neste cenário, o sistema apresenta um melhor resultado, atingindo 93% de concordância, conforme é mostrado na Figura 33.





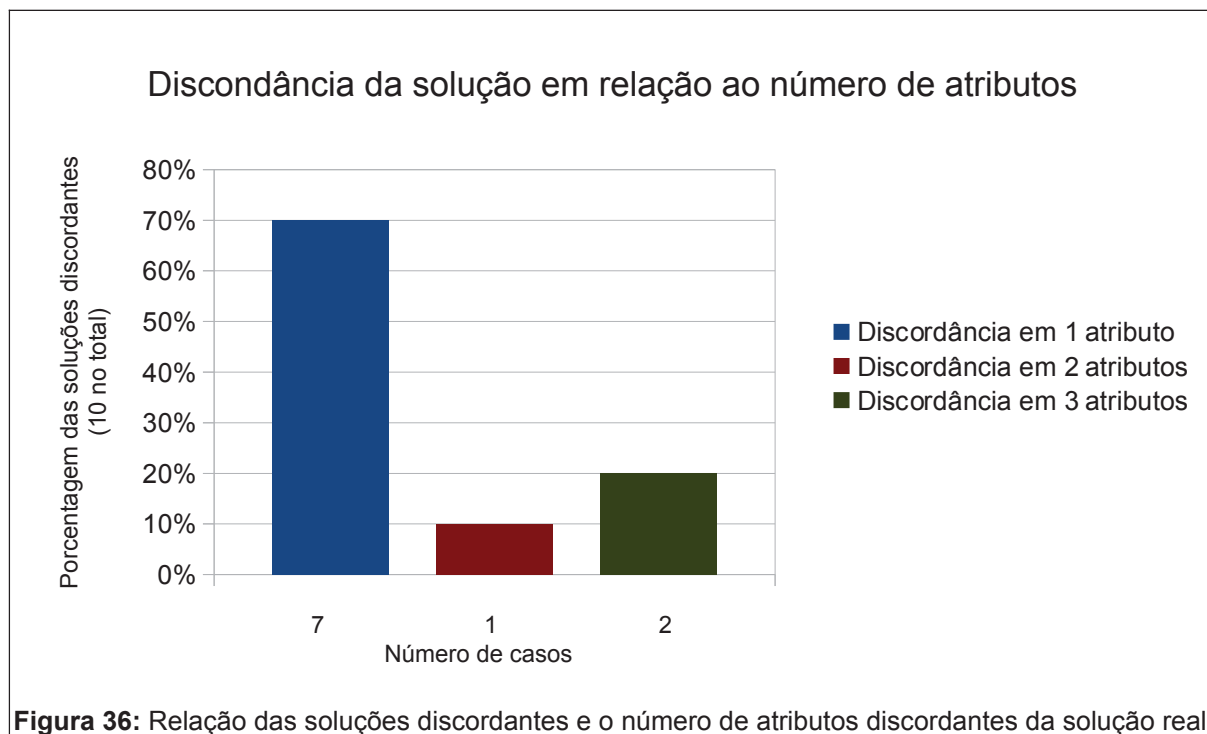
Quando avaliamos a concordância do caso completo, ou seja, quando todos os atributos da solução recomendada são concordantes com os atributos do caso real, o sistema apresenta um desempenho inferior, atingindo 57% de concordância, equivalendo a 13 casos com completa concordância, como ilustra a Figura 34.

Se também definirmos o limiar de similaridade 0.5 e considerarmos os 14 casos consulta que possuem casos recuperados com similaridade maior que 0.5, o sistema também apresenta um resultado mais satisfatório, atingindo 71% de concordância, conforme mostra a Figura 35.



É importante notar que quando a solução proposta não apresenta concordância completa, não significa que todos os seus atributos não são concordantes com a solução real. Muitas vezes, somente um único atributo da solução proposta pelo sistema não é concordante com o atributo correspondente do caso real.

Como ilustrado na Figura 36, 70% das soluções propostas pelo sistema que não apresentam concordância completa, possuem somente um atributo discordante da solução real. Em outras palavras, 7 das 10 solução discordantes somente diferem da solução real em um atributo.



Algumas considerações devem ser feitas em relação às informações da base de casos e o desempenho do sistema.

Como apresentado nos fundamentos do Raciocínio Baseado em Casos, o seu princípio de solução de problemas é por analogia a problemas antigos, experiência adquiridas. Logo, seu desempenho é totalmente relacionado à aquisição e representação dessas experiências na forma de casos.

O conhecimento representado na base de casos utilizada neste trabalho está incompleto, principalmente no que se refere às informações dos atributos de indexação, ou atributos de descrição dos casos. Este fato tem forte impacto diretamente no cálculo de similaridade, que utiliza esses atributos para recuperar os casos mais similares.

Somente 23 casos, dos 377 representados na base de casos, possuem todas as informações

referentes aos atributos de indexação. Portanto, estes foram os casos escolhidos para serem os casos consulta, justamente para não distorcer muito o desempenho do sistema.

Porém, somente 1 único caso desse 23 possui a solução completa, com todas as informações referentes aos atributos da solução. Os demais casos faltam com um ou outro atributo, geralmente o atributo “Tipo de Ancoragem”. Isso prejudica a avaliação das soluções recomendadas, um vez que não se pode afirmar se o “Tipo de Ancoragem” proposto pelo sistema, por exemplo, está em concordância com a solução real, pois esta informação não é fornecida.

Outra questão importante e que influencia também no desempenho do sistema, é a padronização dos possíveis valores dos atributos dos casos. Há uma falta de padronização dos valores da base de casos utilizada, prejudicando o cálculo de similaridade do sistema.

Como exemplo da falta de padronização, o atributo “Condições Ambientais” possui valores como “Winds/storms = 80 mph; Wave height = 32 ft”, quando deveria conter somente os valores “amena”, “moderada”, “severa”, como definido nos trabalhos de Franco (2003) e Morooka & Carvalho (2011). Outro atributo que foge à padronização, é o “Infraestrutura Próxima”, que muitas vezes apresenta o nome da infraestrutura localizada próxima ao campo, como “P-MXL” ou “P-43”, quando deveria conter somente os valores “Sim” ou “Não”.

No caso do atributo “Condições Ambientais”, o sistema fica impossibilitado de saber se “Winds/storms = 80 mph” é considerado como sendo uma condição ambiental “amena”, “moderada” ou “severa”, prejudicando o cálculo de similaridade. Também para o caso do atributo “Infraestrutura Próxima”, o sistema não considera que o valor “P-MXL” é equivalente à presença da infraestrutura.

Uma observação interessante na avaliação dos testes é que muitas vezes quando a solução recomendada não apresenta concordância em todos os seus atributos com a solução real, ou então não apresenta valores para alguns atributos, os outros casos similares recuperados apresentam tal concordância ou possuem as informações que faltam. Este fato pode auxiliar a tomada de decisão correta pelo usuário, uma vez que todos os casos recuperados são apresentados ao usuário. O sistema poderia implementar um algoritmo para a adaptação dos casos, considerando essa observação, para melhorar o desempenho do sistema.

De maneira geral, o sistema apresentou resultados satisfatórios, principalmente quando considerados somente os casos mais similares, premissa do bom desempenho do raciocínio baseado em casos. Ficou claro que o seu desempenho é bastante dependente do conhecimento adquirido e representado na base de casos. Conforme o conhecimento for crescendo com novos casos inseridos na base de casos, e as informações sendo melhoradas, padronizadas e mais completas, o sistema irá apresentando melhores resultados. Também mostrou-se necessário o desenvolvimento de algoritmos de adaptação dos casos e realização de testes com outras medidas de similaridade para tentar melhorar o desempenho do sistema.

## 5 Conclusões

O trabalho apresentado consiste na iniciativa de elaboração e desenvolvimento de uma arquitetura de raciocínio baseado em casos para a construção de sistemas de recomendação. Procurou-se realizar o estudo das técnicas envolvidas no Raciocínio Baseado em Casos, formalizar os conceitos envolvidos e apresentar algumas das abordagens possíveis em cada uma das etapas do seu ciclo.

Uma das principais características do Raciocínio Baseado em Casos que foi determinante para a sua utilização, é o comportamento de refletir o raciocínio humano de solucionar novos problemas lembrando problemas e soluções antigos. Devido a esse comportamento, o processo de convencimento dos implementadores, usuários e gestores da validade desta abordagem é facilitada. Outro fator determinante para a escolha da abordagem RBC, é a manutenção e preservação do conhecimento especialista, adquirido através de experiências e da prática de muitos anos, possibilitando sua disseminação e a integração a todos os interessados.

Para a validação da arquitetura proposta, foi desenvolvido um sistema de raciocínio baseado em casos nebuloso para recomendação, específico para o domínio de sistemas marítimos de produção de petróleo, denominado RIGS. A restrição do desenvolvimento do sistema para abranger um domínio específico, tem como finalidade reduzir a compreensão e conhecimento multidisciplinar necessários, facilitando o desenvolvimento e acelerando a obtenção dos resultados para a validação arquitetural.

Considerando que em alguns domínios do conhecimento os atributos são melhor descritos por valores qualitativos, como é o caso dos sistemas marítimos de produção de petróleo, optou-se pela utilização da abordagem dos conjuntos nebulosos para representá-los e realizar o cálculo de similaridade na etapa de recuperação dos casos.

A escolha do domínio de conhecimento dos sistemas marítimos de produção de petróleo foi feita devido à existência de diversos trabalhos relacionados já realizados disponíveis na literatura, proximidade a especialistas da área, ao conhecimento adquirido e apresentado nos trabalhos de Morooka & Carvalho (2011) e Franco (2003), e à importância dessa área do conhecimento nos cenários econômico, político, ambiental e acadêmico.

De acordo com os resultados obtidos, pode-se concluir que a arquitetura proposta bem como o sistema desenvolvido tiveram um desempenho satisfatório. Do ponto de vista arquitetural, através do modelo estabelecido, foi possível realizar todas as atividades previstas, principalmente a aquisição do conhecimento e as definições do modelo de representação, tratamento dos índices e a método de recuperação, sem muitas complicações. As atividades previstas de estratégia de

adaptação dos casos e método de retenção foram realizadas de forma mais simples, como previsto, por não serem objetos principais de estudo deste trabalho.

Já do ponto de vista do sistema, o desempenho foi satisfatório por apresentar resultados condizentes com a realidade, tendo em vista as limitações e problemas apresentados, principalmente relacionados à base de conhecimento.

Fica evidente nos resultados que o desempenho de sistemas RBC é extremamente relacionado ao conhecimento adquirido e representado na base de casos. Tendo em vista que somente 23 dos 377 casos continham todas as informações, talvez o desempenho do sistema pudesse ter sido consideravelmente alterado com uma base de casos mais completa.

Sobre as dificuldades encontradas no desenvolvimento deste trabalho, pode-se destacar a grande quantidade de métodos e técnicas que podem ser utilizados na abordagem RBC, tornando dispendioso o estudo e compreensão de todos os seus aspectos e ainda dificultando a escolha do mais adequado à proposta do trabalho. Também houve dificuldades na representação e armazenamento do conhecimento, uma vez que a sua aquisição foi feita sem a definição prévia do modelo dos casos.

As principais contribuições deste trabalho que podem ser destacadas são:

- Levantamento e estudo dos principais aspectos que devem ser considerados na elaboração e construção de sistemas baseado em casos.
- Estudo e apresentação de técnicas utilizadas nas diversas etapas da abordagem de Raciocínio Baseado em Casos.
- Especificação de uma arquitetura nebulosa de raciocínio baseado em casos para sistemas de recomendação, independente do domínio do conhecimento.
- Desenvolvimento de um sistema RBC nebuloso para recomendação de sistemas marítimos de produção de petróleo, denominado RIGS, que pode auxiliar e agilizar a definição de sistemas marítimos de produção na elaboração de projetos de desenvolvimento de campos petrolíferos.
- Especificação e desenvolvimento de um *framework* para desenvolvimento de sistemas RBC, utilizado na construção do sistema RIGS.
- Possibilidade de contínua evolução e melhoria no sistema desenvolvido, de acordo as necessidades, devido à arquitetura e *framework* construído.

- Construção de ferramentas para acesso, manutenção e preservação dos conhecimento especialista representado no RIGS, acessíveis aos usuários independente do local e plataforma utilizada.

Como sugestões para trabalhos futuros, podemos ressaltar algumas das muitas possíveis:

- Na etapa de recuperação dos casos, o cálculo de similaridade foi realizado tendo como base a distância Euclidiana ponderada. Como abordado, existem outras distâncias que poderiam ser utilizadas para efeito de comparação de resultados, bem como outras medidas de similaridade.
- Também na etapa de recuperação, a abordagem nebulosa foi utilizada no cálculo de similaridade. Seria válido obter a similaridade utilizando outras abordagens, que não a dos conjuntos nebulosos, para que a comparação dos resultados.
- As funções de pertinência dos conjuntos nebulosos foram especificadas no trabalho de Franco (2003). Não houve uma pesquisa detalhada para saber se existem valores mais atuais para essas funções. Seria oportuna uma pesquisa e atualização desses valores, caso haja, para que os resultados obtidos pelo sistema sejam mais condizentes com a realidade.
- Estudo e implementação de algoritmos de adaptação dos casos recuperados, a fim de melhorar o desempenho do sistema, uma vez que este trabalho abordou de maneira superficial essa etapa.
- Para a etapa de retenção dos casos e aprendizagem, seria interessante o estudo e implementação de outras estratégias para comparação, pois foi realizada somente a estratégia de adição de todo novo caso e sua solução à base de casos.
- Implementação de técnicas de manutenção qualitativa da base de casos, com a finalidade de assegurar a efetividade do sistema RBC.
- Para efeito comparativo, seria valioso confrontar os resultados obtidos pelo sistema desenvolvido com outros obtidos por sistemas que utilizam diferentes abordagens, como por exemplo redes neurais, ou que utilizam a mesma abordagem de RBC porém com diferentes técnicas e métodos.

- Apesar da arquitetura e *framework* desenvolvidos permitirem, diversas ferramentas que facilitariam a utilização do sistema e auxiliariam na análise dos resultados não foram implementadas. Entre elas, podemos destacar: possibilidade de selecionar quais atributos devem ser considerados no cálculo de similaridade; selecionar qual medida de similaridade utilizar, dentre as diversas que poderiam ser implementadas; opção de qual medida de distância utilizar no cálculo de similaridade; possibilidade de escolher a utilização de conjuntos nebulosos ou não, entre outras.
- Quanto às informações dos casos armazenados, seria importante para o desempenho do sistema que houvesse um refinamento para garantir a padronização dos dados. Também seria extremamente valioso que fosse feito um levantamento e aquisição das informações que faltam na descrições e soluções dos casos armazenados, para efetivamente melhorar os resultados obtidos.
- Como o sistema foi construído especificamente para o domínio de sistemas marítimos de produção de petróleo, para validar a arquitetura proposta, uma extensão para este trabalho poderia ser o desenvolvimento da sua generalização para outros domínios de conhecimentos. Desse modo, o sistema permitiria ao usuário percorrer todas as etapas e atividades da arquitetura especificada de modo a obter, como resultado, um sistema RBC para recomendação específica ao domínio desejado. A arquitetura e o *framework* foram elaborados e desenvolvidos considerando tendo como objetivo esta generalização.
- Por fim, diversas outras funcionalidades e ferramentas poderiam ser desenvolvidas e adicionadas ao sistema de modo a permitir maior flexibilidade e facilidade para análises das soluções recomendadas, para que possam ser mais efetivas no auxílio ao processo decisório do usuário.

## Referências Bibliográficas

AAMODT, A. & PLAZA, E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. Published in AICom – Artificial Intelligence Communications, IOS Press, Vol. 7: 1, pp. 39-59, 1994.

DELPIZZO, V. L. F. Prescrição de atividades físicas através do uso da inteligência artificial. Dissertação de Mestrado. ENGENHARIA DE PRODUÇÃO DA UNIVERSIDADE FEDERAL DE SANTA CATARINA 1997.

FRANCO, K. P. M. Desenvolvimento de um Sistema Inteligente para Auxiliar a Escolha de Sistema para Produção no Mar. Campinas: Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 2003. 117p. Tese de Mestrado.

FOSTER, E. “AnArch: uma arquitetura orientada a serviços para aplicações ricas para a internet”, Rio Claro, Trabalho de conclusão de curso – Universidade Estadual Paulista Júlio de Mesquita Filho 2007.

GOMES, P. J. S. A Case-Based Approach to Software Design. Universidade de Coimbra, Faculdade de Ciências e Tecnologia, Departamento de Engenharia Informática. ISBN 972-9039-69-0, 2003.

HAMMOND, K.J. Case-based planning. Academic Press, 1989.

HINRICHS, T.R. Problem solving in open worlds. Lawrence Erlbaum Associates, 1992.

JCOLIBRI. COLIBRI CBR Plataforma <http://gaia.fdi.ucm.es/projects/jcolibri/>, 2011

KAUFMANN, A.; GUPTA, M. M. Fuzzy Mathematical Models in Engineering and Management Science, North Holland, 337p., 1988.

KOLODNER, J. Case-Based Reasoning. Morgan Kaufmann Publishers, Inc. San Mateo, CA, 1993.

KOLODNER, J. Reconstructive Memory: A Computer Model. Cognitive Science, 7, p.281-328, 1983.

KOSKO, B. Neural Networks and Fuzzy Systems; A Dynamical Systems Approach to Machine Intelligence, Englewood Cliffs, NJ: Prentice Hall, 1992.

KOTON, P. Using experience in learning and problem solving. Massachusetts Institute of Technology, Laboratory of Computer Science (Ph.D. diss, October 1988). MIT/LCS/TR-441. 1989.

LEAKE, D. Case-Based Reasoning: Experiences, Lessons & Future Directions, California: AAAI Press/The MIT Press, 1996.

MÁNTARAS, R.L. de, McSHERRY, D., BRIDGE, D., LEAKE, D., SMYTH, B., CRAW, S., FALTINGS, B., MAHER, M.L., COX, M.T., FORBUS, K., KEANE, M., AAMODT, A., WATSON, I.: Retrieval, reuse, revision and retention in case-based reasoning. *Knowl. Eng. Rev.* **20(3)** (2005) 215–240

MENDES, J. R. P. Raciocínio Baseado em Casos Aplicado ao Projeto de Poços de Petróleo. Campinas, 2001, 99p. Dissertação (Doutorado em Ciências e Engenharia de Petróleo) – Faculdade de Engenharia Mecânica, UNICAMP, 2001.

MENDES, J. R. P; GUILHERME, I. R. ; MOROOKA, C. K. Case Based System: Indexing and Retrieval with Fuzzy Hypercube. In: International, 2001, Vancouver. The 9th IFSA World Congress -2001, 2001. p. 818-823.

MOROOKA, C.K.; CARVALHO, M. D. M. Evaluation of Alternatives for Offshore Petroleum Production System in Deep and Ultradeep Water Depth. In: International Conference on Ocean, Offshore and Arctic Engineering, 2011, Rotterdam, The Netherlands. Proceedings fo the 30th International Conference on Ocean, Offshore and Arctic Engineering. Estados Unidos : ASME - American Society of Mechanical Engineers, 2011.

PAL, S. K.; SHIU, S. C. K. Foundations of Soft Case-Based Reasoning. New Jersey: Wiley and Sons 2004.

SÁ, F. P. Avaliação da Recuperação no Raciocínio Baseado em Casos Estrutural e Textual em um Sistema Help-Desk. Dissertação de Mestrado. Universidade Católica de Santos, 2007

SADEGH-ZADEH<sup>(a)</sup>, K. Advances in fuzzy theory. *Artificial Intelligence in Medicine*, 1999, v.15, p.309-323.

SADEGH-ZADEH<sup>(b)</sup>, K. Fundamentals of clinical methodology: 3. Nosology. *Artificial Intelligence in Medicine*, 1999, v.17, p.87-108.

SCHANK, R. Dynamic memory; a theory of reminding and learning in computers and people. Cambridge University Press, 1982.

SIMPSON, R.L. A computer model of case-based reasoning in problem solving: An investigation in the domain of dispute mediation. Technical Report GIT-ICS-85/18, Georgia Institute of Technology, 1985.

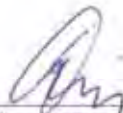
SYCARA, K. The Persuader, The Encyclopedia of Artificial Intelligence, D. Shapiro, ed., John Wiley and Sons, 1992

ZADEH, L. A. Fuzzy sets, *Information and Control*, vol. 8, pp. 338–353, 1965.

ZADEH, L. A. Soft computing and fuzzy logic, *IEEE Software*, vol. 11, no. 6, pp. 48–56, 1994.

Autorizo a reprodução xerográfica para fins de pesquisa.

São José do Rio Preto, 11/05/2012



---

Assinatura