

Universidade Regional de Blumenau
Centro de Ciências Exatas e Naturais
Departamento de Sistemas e Computação

Business Intelligence aplicando Data Mining utilizando a Teoria dos Conjuntos Aproximativos na Gestão da Saúde

Acadêmico: Oraci Izidoro Corrêa
Prof. Orientador: Dr. Oscar Dalfovo



Roteiro da apresentação

- Introdução
- Objetivos
- Fundamentação teórica
- Especificação
- Desenvolvimento do sistema
- Operacionalidade
- Resultados e discussões
- Conclusão
- Extensões
- Referências



Introdução

- Torna-se de extrema importância a necessidade de administrar as informações.
- Informações apresentadas muitas vezes são complexas e sem organização.
- As técnicas Data Mining visam extrair dos dados as informações e conhecimento pertinentes.



Objetivos

- Geral

Desenvolver uma ferramenta aplicando *Data Mining* utilizando a teoria dos conjuntos aproximativos para a extração de informações que auxiliarão os gestores no setor hospitalar nas tomadas de decisões, através de relatórios e gráficos

Objetivos

- Específicos

- aplicar a técnica de mineração de dados utilizando a TCA;
- possibilitar o acesso dinamicamente as informações nas bases de dados hospitalar;
- apresentar as informações extraídas, através de filtros e apresentá-las em forma de relatórios e gráficos;
- disponibilizar as informações gerenciais através de relatórios, telas e gráficos, para auxiliar os gestores do setor hospitalar na tomada de decisão.



Fundamentação teórica

- Business Intelligence
- Data Mining
- Teoria dos Conjuntos Aproximativos (TCA)



Fundamentação teórica

- Business Intelligence
 - Realizar consultas ao banco de dados e gerar relatórios sem a necessidade de conhecimento de programação.
 - Transformar os registros obtidos em informação útil para o conhecimento empresarial.
 - Utilização de variadas fontes de informação.

Fundamentação teórica

- Business Intelligence
 - Ferramentas:
 - Data Warehouse e Data Marts (BD – SAD);
 - Database Marketing (BD clientes, construção de perfis);
 - Customer Relationship Management (foco no cliente);
 - Data Mining (garimpar dados, análise automática, padrões);
 - Balanced Scorecard (monitorar evolução de decisões).



Fundamentação teórica

- Data Mining
 - Ferramenta capaz de analisar automaticamente base de dados para obter o conhecimento;
 - Composto por técnicas e algoritmos especializados que buscam transformar os dados armazenados em informações úteis.

Fundamentação teórica

- Teoria dos Conjuntos Aproximativos (TCA)
 - Surgiu na década de 80, ao qual foi desenvolvida por Zdzislaw Pawlak;
 - Ferramenta matemática de redução de dimensionalidade;
 - Busca encontrar todos os objetos que produzem um mesmo tipo de informação;
 - Comparada a outras técnicas é mais fácil de aplicar.

Fundamentação teórica

- Exemplo TCA

Tabela

- Considerando os dados da tabela, tem-se as seguintes condições:

$U = \{1,2,3,4,5,6\}$ é o conjunto universo de objetos (pacientes);

$C = \{C,M,T\}$ é o conjunto de atributos de condição;

$D = \{G\}$ é o conjunto de atributos de decisão.

Fundamentação teórica

- Exemplo TCA

Tabela

Tabela

- Considerando os atributos {M, T}, têm-se os seguintes conjuntos P-Elementares:

$$\text{Des}_p(\{1\}) = \{(M, \text{Sim}), (T, \text{Alta})\};$$

$$\text{Des}_p(\{2,5\}) = \{(M, \text{Não}), (T, \text{Alta})\};$$

$$\text{Des}_p(\{3,6\}) = \{(M, \text{Sim}), (T, \text{Muito Alta})\};$$

$$\text{Des}_p(\{4\}) = \{(M, \text{Sim}), (T, \text{Normal})\}.$$

Fundamentação teórica

- Exemplo TCA

Tabela

- $\{1,2,3,6\}$, conjunto de pacientes que apresentam a gripe;
- $\{4,5\}$, conjunto de pacientes que não apresentam a gripe.

- Sendo Y^1 Sim e Y^2 Não, temos:

$$P\text{-Elementar} = \{(1), (2,5), (3,6), (4)\}$$

$$P_{Y^1} = \{1\} \cup \{3, 6\} = \{1, 3, 6\};$$

$$P^{Y^1} = \{1\} \cup \{2, 5\} \cup \{3, 6\} = \{1, 2, 3, 5, 6\};$$

$$P_{Y^2} = \{4\};$$

$$P^{Y^2} = \{4\} \cup \{2, 5\} = \{2, 4, 5\};$$

$$\gamma^P(Y) = \frac{\text{card}(P_{Y^1}) + \text{card}(P_{Y^2})}{\text{card}(U)} = \frac{3 + 1}{6} = 0,667$$

Fundamentação teórica

- Exemplo TCA
 - Calculando todas as aproximações obtêm-se os seguintes resultados:

AtributosP	Qualidade Aproximação	Conj. P-elementares
{C,M,T}	0,667	{1},{2,5},{3},{4},{6}
{M,T}	0,667	{1},{2,5},{3,6},{4}
{C,T}	0,667	{1},{2,5},{3},{4},{6}
{C,M}	0,167	{1,4,6},{2,5},{3}
{T}	0,500	{1,2,5},{3,6},{4}
{M}	0,000	{1,3,4,6},{2,5}
{C}	0,000	{1,4,6},{2,3,5}

Fundamentação teórica

- Exemplo TCA

- as Y-Reduções de $P = \{C, M, T\}$ são $\{M, T\}$ e $\{C, T\}$;

- o núcleo de P é:

$$\text{Core}_\gamma(\{C, M, T\}) = \{C, T\} \cap \{M, T\} = T$$

- T é o atributo mais significativo.

Fundamentação teórica

- Exemplo TCA
 - Caso adote-se a Y-Redução $\{M, T\}$, teremos as seguintes informações:

Conjunto	Decisão	Regra	Determinística
{1}	{1,2,3,6}	$\{(M,Sim),(T,Alta)\} \rightarrow \{(G,Sim)\}$	Sim
{2,5}	{1,2,3,6}	$\{(M,Não),(T,Alta)\} \rightarrow \{(G,Sim)\}$	Não
{3,6}	{1,2,3,6}	$\{(M,Sim),(T,Muito Alta)\} \rightarrow \{(G,Sim)\}$	Sim
{2,5}	{4,5}	$\{(M,Não),(T,Alta)\} \rightarrow \{(G,Não)\}$	Não
{4}	{4,5}	$\{(M,Sim),(T,Normal)\} \rightarrow \{(G,Não)\}$	Sim

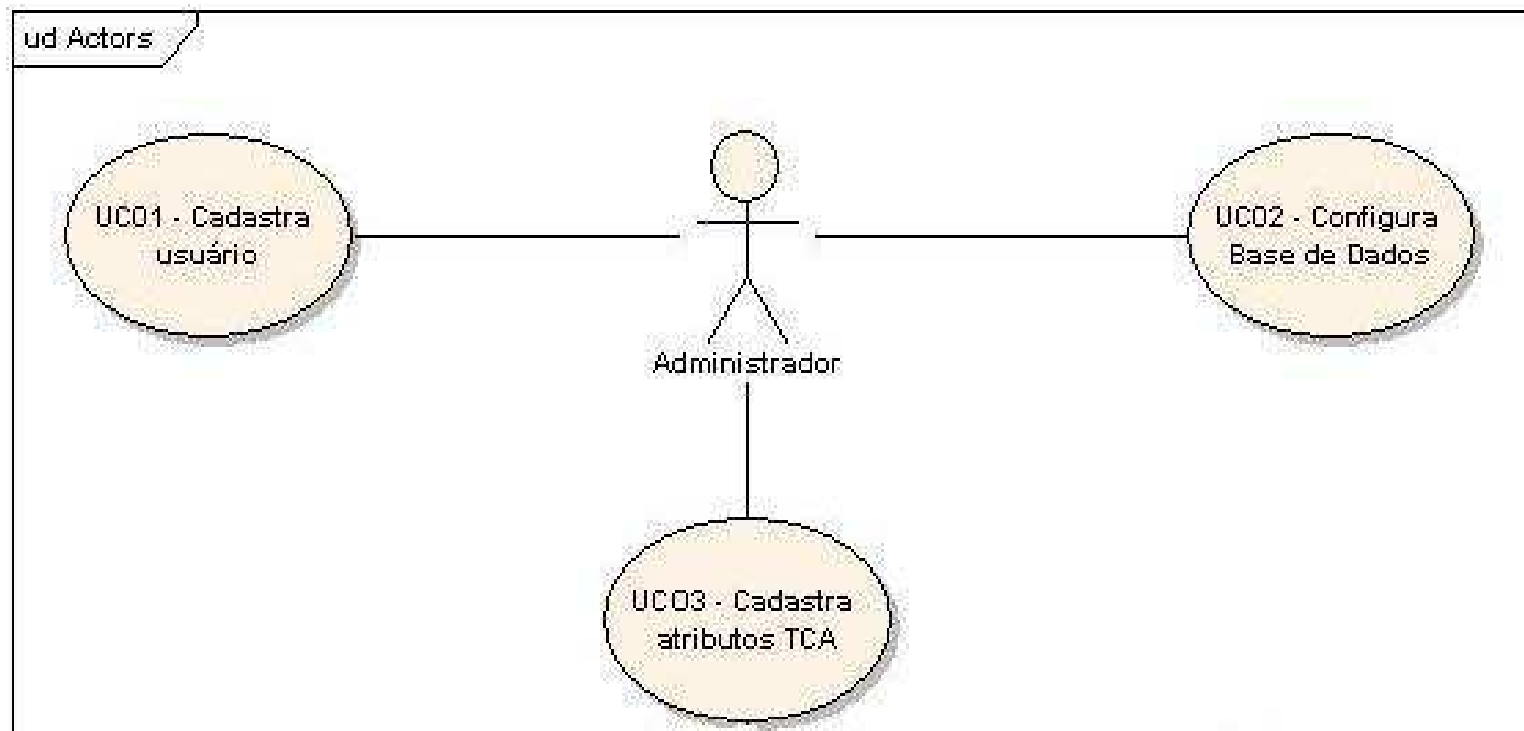
Especificação

- Levantamento de requisitos

Requisitos Funcionais	Caso de Uso
RF01: O sistema deverá permitir o cadastro de usuários.	UC01
RF02: O sistema deverá permitir a configuração de acesso a mais de uma base de dados.	UC02
RF03: O sistema deverá permitir o cadastro de atributos TCA.	UC03
RF04: O sistema deverá aplicar a técnica de mineração de dados utilizando a TCA.	UC04
RF05: O sistema deverá permitir a visualização do cubo de decisão com os resultados da TCA.	UC05
RF06: O sistema deverá permitir a visualização dos históricos dos resultados resumidos da TCA.	UC06
RF07: O sistema deverá permitir a visualização do relatório de acordo com os resultados da TCA.	UC07
RF08: O sistema deverá permitir a visualização dos resultados da TCA em forma de gráfico.	UC08

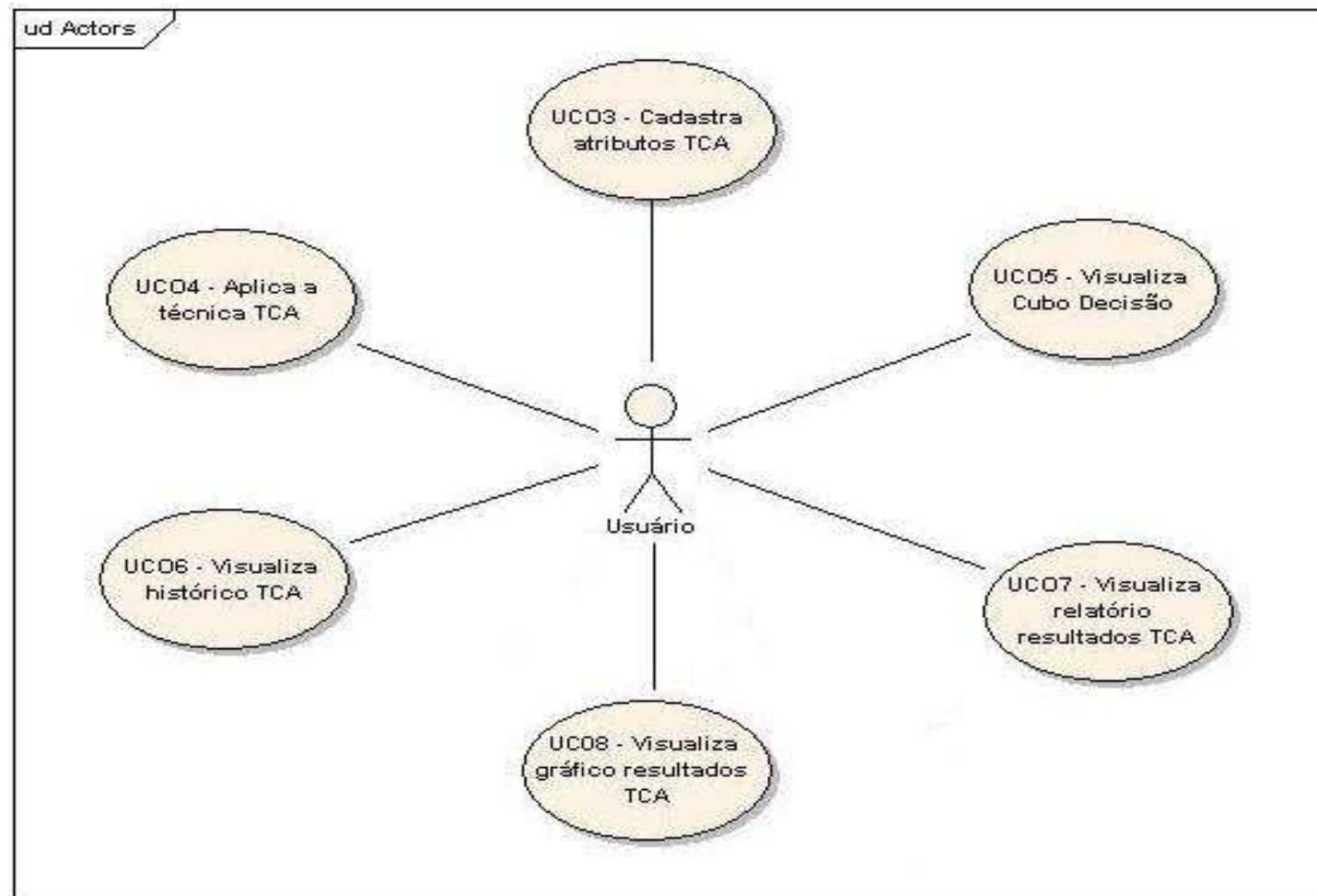
Especificação

- Diagrama de caso de uso Administrador:



Especificação

- Diagrama de caso de uso Usuário:

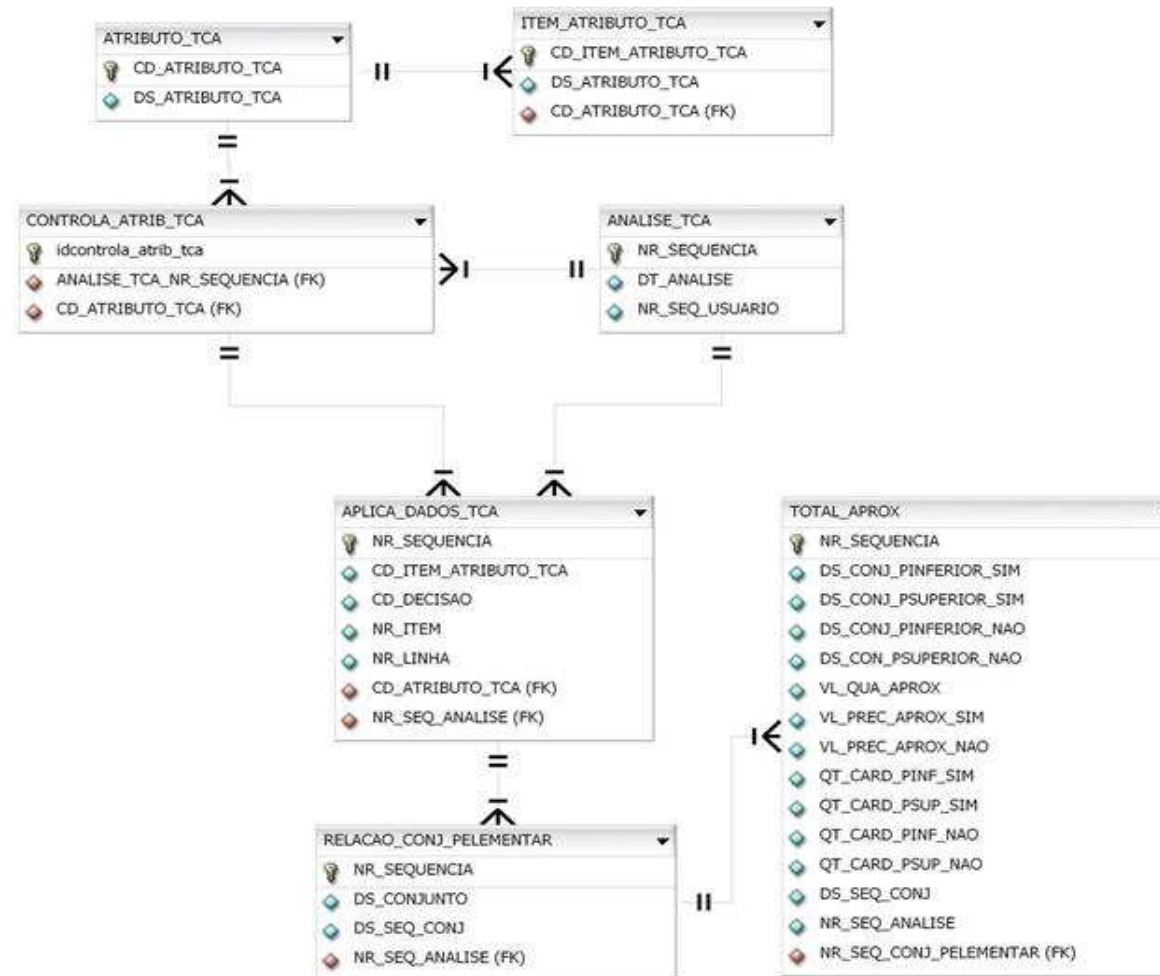


Especificação

- Diagrama de atividade do processo de análise TCA:
- Diagrama de atividades

Especificação

- Modelo de Entidade de Relacionamento (MER)





Desenvolvimento do sistema

- Técnicas e ferramentas utilizadas:
 - Enterprise Architect – UML:
 - Criação dos diagramas e casos de uso.
 - Oracle 10g:
 - Base de dados de informação do sistema.
 - PL/Sql - Procedural Language / Structured Query Language:
 - Criação das procedures e functions para o cálculo TCA.
 - Delphi 7.0:
 - Ambiente de desenvolvimento para a implementação do trabalho.

Operacionalidade

- Tela de login:



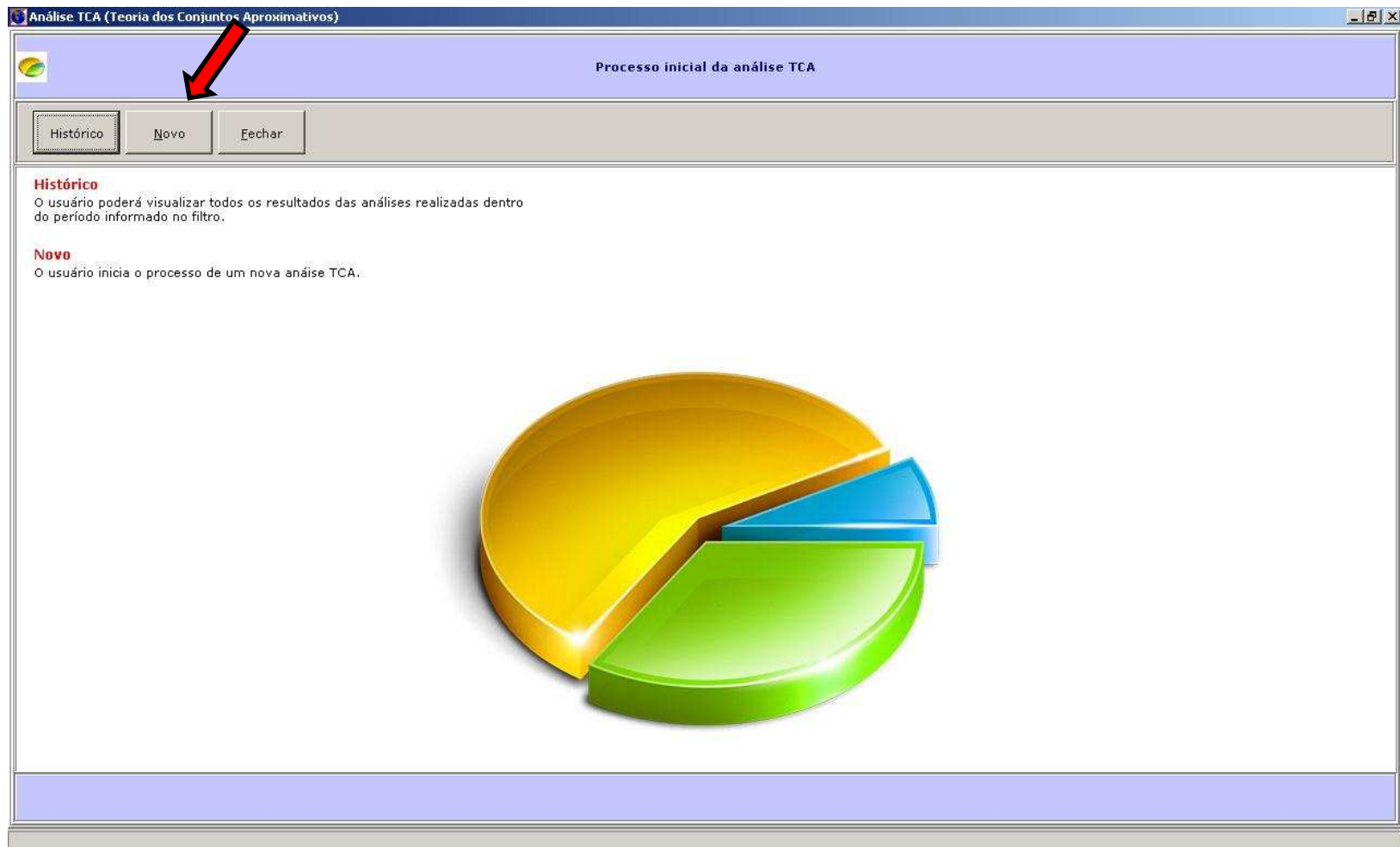
Operacionalidade

- Tela principal do sistema:



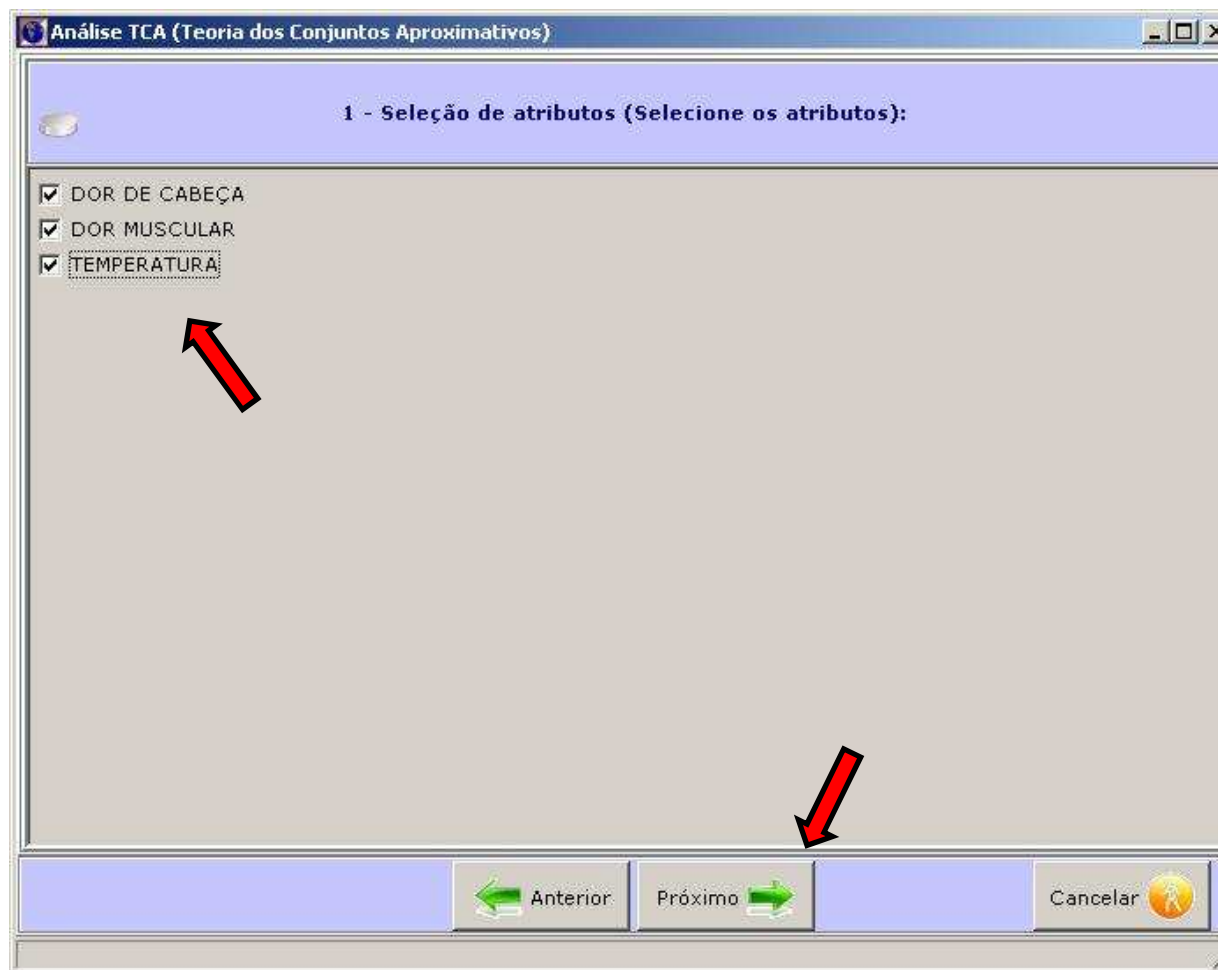
Operacionalidade

- Tela principal da análise TCA:



Operacionalidade

- Tela de seleção de atributos:



Operacionalidade

- Tela de seleção dos valores dos atributos:

The screenshot shows a software window titled "Análise TCA (Teoria dos Conjuntos Aproximativos)". The main area is titled "2 - Universo de objetos (Selecione os valores dos atributos selecionados):". Below this title are two buttons: "Incluir" and "Excluir".

The "Atributos" section contains four dropdown menus with the following values selected: "DOR DE CABEÇA" (NÃO), "DOR MUSCULAR" (SIM), "TEMPERATURA" (MUITO ALTA), and "DECISÃO" (SIM).

The "Atributos, valores e decisões informadas" section contains a table with the following data:

Universo informado	Decisão
(Linha 1) - DOR DE CABEÇA: NÃO, DOR MUSCULAR: SIM, TEMPERATURA: ALTA	SIM
(Linha 2) - DOR DE CABEÇA: SIM, DOR MUSCULAR: NÃO, TEMPERATURA: ALTA	SIM
(Linha 3) - DOR DE CABEÇA: SIM, DOR MUSCULAR: SIM, TEMPERATURA: MUITO ALTA	SIM
(Linha 4) - DOR DE CABEÇA: NÃO, DOR MUSCULAR: SIM, TEMPERATURA: NORMAL	NÃO
(Linha 5) - DOR DE CABEÇA: SIM, DOR MUSCULAR: NÃO, TEMPERATURA: ALTA	NÃO
(Linha 6) - DOR DE CABEÇA: NÃO, DOR MUSCULAR: SIM, TEMPERATURA: MUITO ALTA	SIM

At the bottom of the window are three buttons: "Anterior" (with a left arrow), "Próximo" (with a right arrow), and "Cancelar" (with a yellow circle icon). A red arrow points to the "Próximo" button.

Operacionalidade

- Resultados da análise TCA:

Resultados TCA

3 - Resultado:

Regras geradas com decisão = SIM:

TEMPERATURA: ALTA
DOR MUSCULAR: SIM
TEMPERATURA: MUITO ALTA
DOR MUSCULAR: SIM
TEMPERATURA: NORMAL
DOR MUSCULAR: SIM

Qualidade de aproximação

Conjuntos de atributos relacionados	Qualidade de aproximação	Conjuntos P-Elementares
▶ (TEMPERATURA, DOR MUSCULAR, DOR DE CABEÇA,)	0,0000	(1),(2,5),(3),(4),(6)
(TEMPERATURA, DOR MUSCULAR,)	0,0000	(1),(2,5),(3,6),(4)
(TEMPERATURA, DOR DE CABEÇA,)	0,0000	(1),(2,5),(3),(4),(6)
(TEMPERATURA,)	0,0000	(1,2,5),(3,6),(4)
(DOR MUSCULAR, DOR DE CABEÇA,)	0,0000	(1,4,6),(2,5),(3)
(DOR MUSCULAR,)	0,0000	(1,3,4,6),(2,5)
(DOR DE CABEÇA,)	0,0000	(1,4,6),(2,3,5)

Universo de dados informados

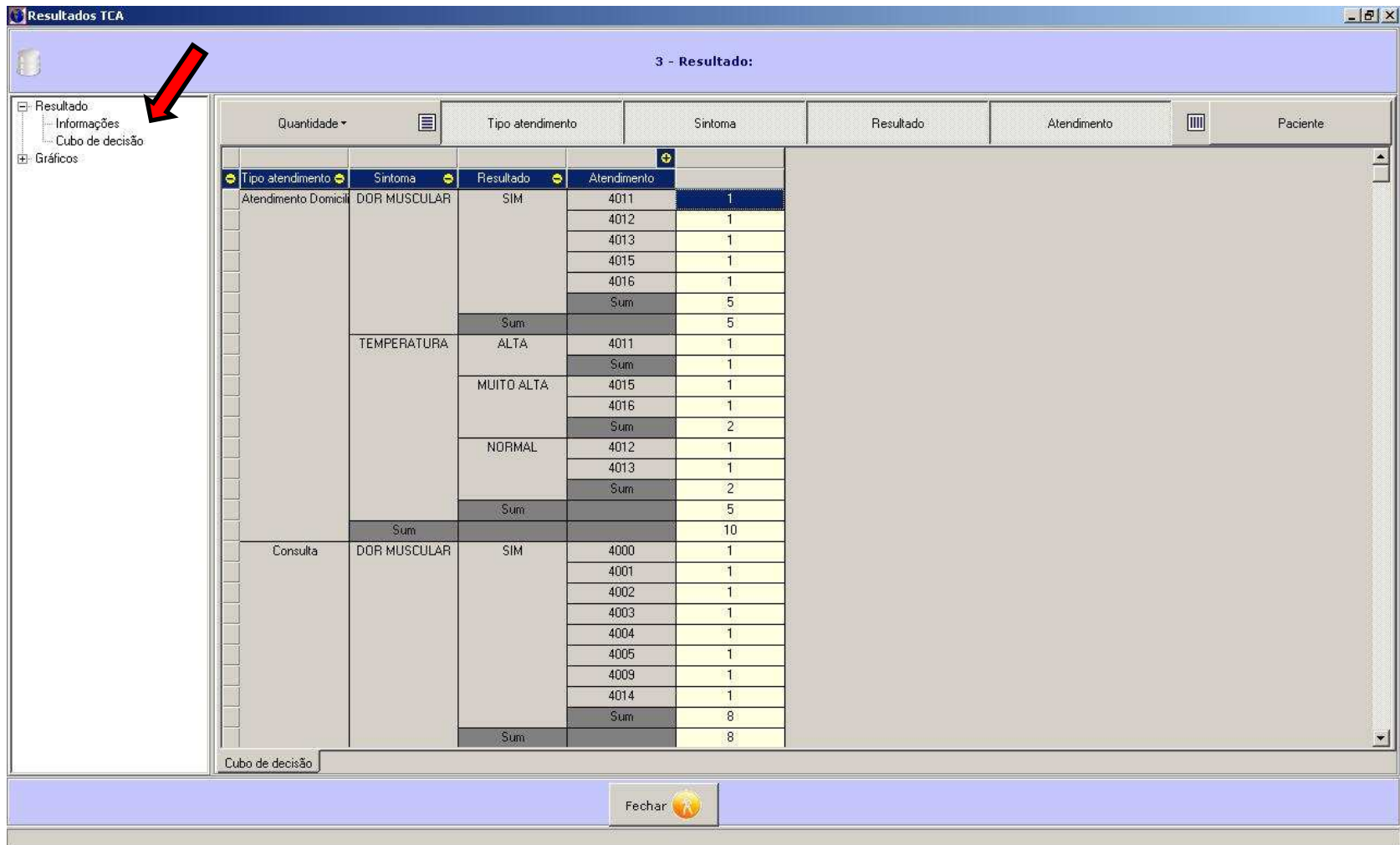
Atributos	Decisão	Linha
▶ (Linha 1) - DOR DE CABEÇA: NÃO, DOR MUSCULAR: SIM, TEMPE SIM		1
(Linha 2) - DOR DE CABEÇA: SIM, DOR MUSCULAR: NÃO, TEMPE SIM		2
(Linha 3) - DOR DE CABEÇA: SIM, DOR MUSCULAR: SIM, TEMPE SIM		3
(Linha 4) - DOR DE CABEÇA: NÃO, DOR MUSCULAR: SIM, TEMPE NÃO		4
(Linha 5) - DOR DE CABEÇA: SIM, DOR MUSCULAR: NÃO, TEMPE NÃO		5
(Linha 6) - DOR DE CABEÇA: NÃO, DOR MUSCULAR: SIM, TEMPE SIM		6

Informações

Fechar

Operacionalidade

- Cubo de decisão:



Resultados TCA

3 - Resultado:

Resultado
Informações
Cubo de decisão
Gráficos

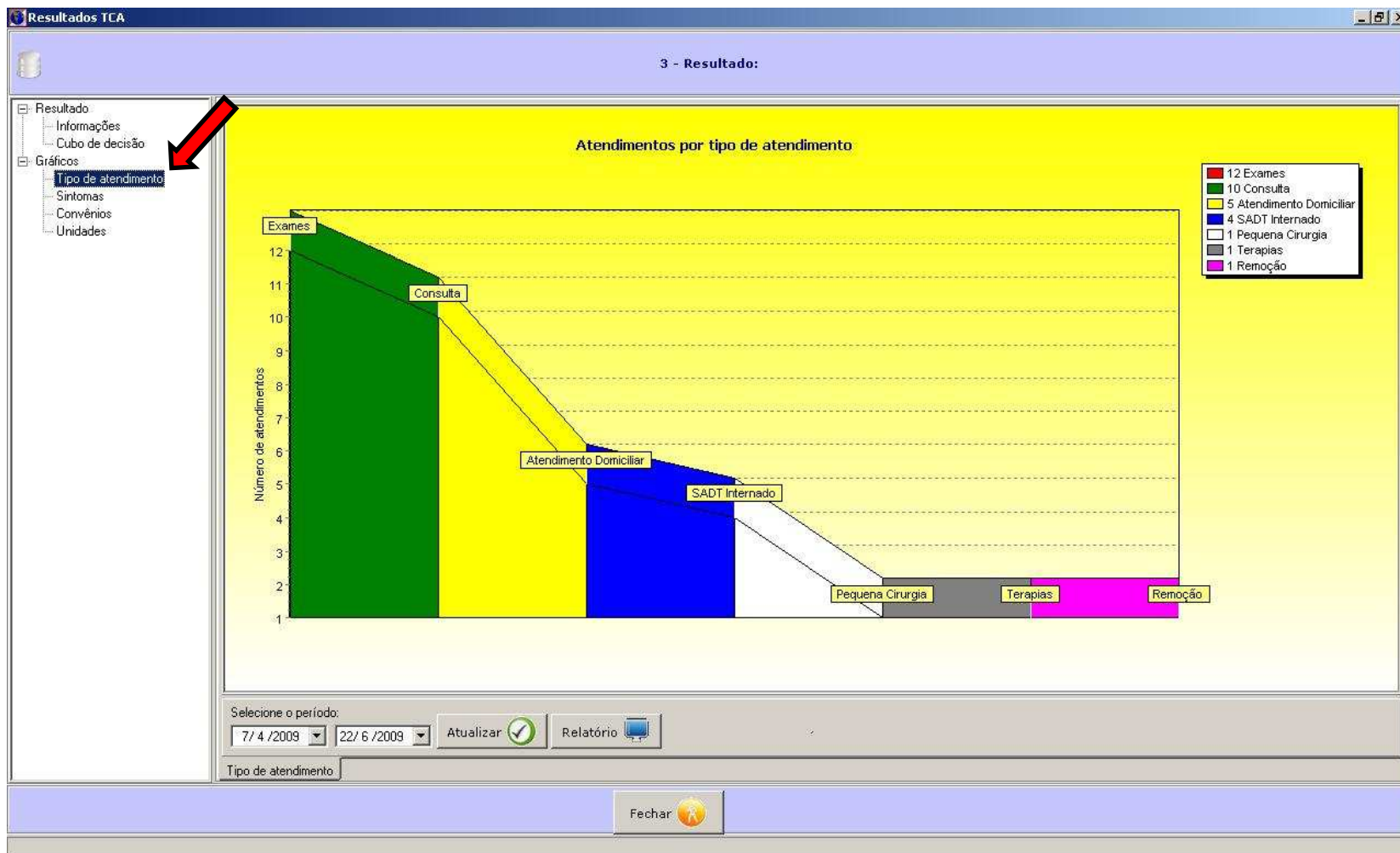
Quantidade	Tipo atendimento	Sintoma	Resultado	Atendimento	Paciente
	Tipo atendimento	Sintoma	Resultado	Atendimento	
	Atendimento Domicil	DOR MUSCULAR	SIM	4011	1
				4012	1
				4013	1
				4015	1
				4016	1
				Sum	5
			Sum		5
		TEMPERATURA	ALTA	4011	1
				Sum	1
			MUITO ALTA	4015	1
				4016	1
				Sum	2
			NORMAL	4012	1
				4013	1
				Sum	2
			Sum		5
		Sum			10
	Consulta	DOR MUSCULAR	SIM	4000	1
				4001	1
				4002	1
				4003	1
				4004	1
				4005	1
				4009	1
				4014	1
				Sum	8
			Sum		8

Cubo de decisão

Fechar

Operacionalidade

- Resultados em gráfico:



Operacionalidade

- Resultados em relatório:

Relatório de atendimentos por tipo de atendimento

Data inicial: 7/4/2009 Data final: 22/6/2009

Tipo atendimento	Atendimento	Paciente	Prontuário	Data entrada	Data nascimento
Atendimento Domiciliar					
	4011	Pessoa_2011	2012	15/5/2009	29/6/2009
	4012	Pessoa_2012	2013	15/5/2009	29/6/2009
	4013	Pessoa_2013	2014	15/5/2009	29/6/2009
	4015	Pessoa_2015	2016	15/5/2009	29/6/2009
	4016	Pessoa_2016	2017	15/5/2009	29/6/2009
Consulta					
	4000	Pessoa_2000	2001	11/5/2009	29/6/2009
	4001	Pessoa_2001	2002	11/5/2009	29/6/2009
	4002	Pessoa_2002	2003	11/5/2009	29/6/2009
	4003	Pessoa_2003	2004	11/5/2009	29/6/2009
	4004	Pessoa_2004	2005	11/5/2009	29/6/2009
	4005	Pessoa_2005	2006	11/5/2009	29/6/2009
	4009	Pessoa_2009	2010	11/5/2009	29/6/2009
	4014	Pessoa_2014	2015	15/5/2009	29/6/2009
	4023	Pessoa_2023	2024	15/5/2009	29/6/2009

Page 1 of 1

Resultados e Discussões

- A técnica TCA se mostrou eficaz na redução do dimensionamento das informações;
- Informações foram apresentadas em forma de cubo de decisão, relatórios e gráficos.
- Gerando auxílio na tomada de decisão.



Conclusão

- Informação força vital das empresas.
- Os processos e ferramentas Data Mining são indispensáveis nos processos de busca pelo conhecimento.
- Todos os objetivos foram alcançados.
- A metodologia aplicada mostrou-se adequada.
- Aprendizado.



Extensão

- Para trabalhos futuros sugere-se que este trabalho seja aplicado em outros nichos de mercado, para uma melhor análise de resultados. Aplicar outras técnicas de DM para que seja possível realizar um comparativo dos resultados obtidos.

Referências

- BARBIERI, Carlos. **BI – Business Intelligence: modelagem & tecnologia**. Rio de Janeiro:Axel Books do Brasil, 2001.
- PAWLAK, Zdzislaw. **Rough Sets. Internacional Journal of Information & Computer Sciences**. [S.1.], v.11, p 341-356, 1982.
- PATRÍCIO, Cristian M. M. M.; PINTO, João O. P.; SOUZA, Celso C. **Rough Sets – Técnica de Redução de Atributos e Geração de Regras para Classificação de Dados**. Campo Grande, 2005. Disponível em:<
http://www.sbmac.org.br/eventos/cnmac/cd_xxviii_cnmac/resumos%20estendidos/cristian_patricio_ST18.pdf>. Acesso em: 5 março 2009.
- PESSOA, Alex S. A.; SIMÕES, José D. S.; **Estudo do comportamento climático utilizando uma abordagem neuro-aproximativa**. 2004. Disponível em: <
http://hermes2.dpi.inpe.br:1905/col/lac.inpe.br/worcap/2004/10.06.13.09/doc/worcap_alex2004.pdf>. Acesso em: 07 abr. 2009.
- QUONIAM, Luc et al. **Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil**. Brasília. 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n2/6208.pdf>>. Acesso em: 22 mar. 2009.
- SERRA, Laércio. **A essência do Business Intelligence**. 1ª Edição, São Paulo: Editora Berkely Brasil, 2002.



Obrigado!

- “Quanto maior a dificuldade, tanto maior o mérito em superá-la”

Henry Ward Beecher

Teólogo estadunidense (1813-1887)