

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE SISTEMAS DE INFORMAÇÃO – BACHARELADO

BUSINESS INTELLIGENCE APLICANDO DATA MINING
UTILIZANDO A TEORIA DOS CONJUNTOS
APROXIMATIVOS NO DIAGNÓSTICO DE DOENÇAS

ORACI IZIDORO CORRÊA

BLUMENAU
2009

2009/1-11

ORACI IZIDORO CORRÊA

BUSINESS INTELLIGENCE APLICANDO DATA MINING

UTILIZANDO A TEORIA DOS CONJUNTOS

APROXIMATIVOS NO DIAGNÓSTICO DE DOENÇAS

Trabalho de Conclusão de Curso submetido à Universidade Regional de Blumenau para a obtenção dos créditos na disciplina Trabalho de Conclusão de Curso II do curso de Sistemas de Informação— Bacharelado.

Prof. Oscar Dalfovo, Dr. - Orientador

**BLUMENAU
2009**

2009/1-11

BUSSINES INTELLIGENCE APLICANDO DATA MINING
UTILIZANDO A TEORIA DOS CONJUNTOS
APROXIMATIVOS NA GESTÃO DA SAÚDE

Por

ORACI IZIDORO CORRÊA

Trabalho aprovado para obtenção dos créditos na disciplina de Trabalho de Conclusão de Curso II, pela banca examinadora formada por:

Presidente: _____
Prof. Oscar Dalfovo, Dr. – Orientador, FURB

Membro: _____
Prof. Everaldo Artur Grahl, Mestre – FURB

Membro: _____
Prof. Alexander Roberto Valdameri, Mestre – FURB

Blumenau, 09 de julho de 2009

Dedico este trabalho, especialmente a minha família e amigos que acreditaram na conquista de mais este desafio e que me ajudaram diretamente na realização deste.

AGRADECIMENTOS

A Deus, pelo seu imenso amor e graça.

Aos meus pais, por todo o seu amor, atenção e dedicação, e por todo o momento que me incentivaram. Por estarem sempre presentes nos momentos complicados e por acreditarem em mim.

A minha namorada, Anna pelo seu amor e pela sua paciência enquanto eu desenvolvia este trabalho, me cobrando e incentivando.

Aos meus amigos que fazem parte de um grupo denominado Cúpula pelos empurrões e cobranças, e por agüentar minhas reclamações a cada dia que envolveu o desenvolvimento deste trabalho.

Ao meu orientador, Dr. Oscar Dalfovo, por ter acreditado e me dado a força e o apoio necessários para a conclusão deste trabalho.

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

Henry Ward Beecher

RESUMO

Muitas instituições sofrem com a falta de qualidade das informações geradas causando um grave problema de crescimento dentro das instituições. Diante disto, surge a necessidade dessas informações serem geradas de uma maneira organizada fazendo com que realmente tragam um valor significativo no auxílio à tomada de decisão das pessoas envolvida. Visando auxiliar na qualidade de geração do conhecimento das informações o presente trabalho apresenta o uso de Business Intelligence aplicando Data Mining aliado à teoria dos conjuntos aproximativos na de área de diagnóstico de doenças, mais especificamente no auxílio ao diagnóstico médico de acordo com os sintomas informados pelo usuário. Apresentando um sistema desktop para o gerenciamento das informações, dentro deste processo, o sistema lista os atendimentos dos pacientes de acordo com os parâmetros informados e através do cálculo da TCA, demonstrando os resultados através do cubo de decisão auxiliando no atendimento do paciente.

Palavras-chave: Business Intelligence; Teoria dos conjuntos aproximativos; Mineração de dados; Cubo de decisão.

ABSTRACT

Many institutions suffer from lack of quality of information generated causing a serious problem for growth within the institutions. Considering this, it is necessary that information be generated in an organized manner so that really bring significant value to aid in decision-making of people involved. Aiming to help a generation of knowledge of this information paper presents the of Business Intelligence using Data Mining combined with Rough Sets theory area of diagnosis of diseases, especially in aid to medical diagnosis according to symptoms reported by the user. Introducing a system for desktop information management within this process, the system lists the care of patients in accordance with the parameters and informed by calculating the TCA, showing the results through the hub of decision cube in the care of the patient.

Key-words: Business Intelligence; Rough Sets Theory; Data Mining, Decision cube.

LISTA DE ILUSTRAÇÕES

Figura 1 - Agregando valor ao dado para a tomada de decisão.....	17
Figura 2 - Tecnologia da informação nas organizações.	18
Figura 3 - Do dado ao conhecimento.	18
Figura 4 - Etapas do processo de Data Mining.....	19
Figura 5 - Representação da redução da base de dados.....	21
Quadro 1 - Exemplo de regra de decisão.....	21
Quadro 2 - Definição do conjuntos dentro do universo	22
Quadro 3 - Relação de equivalência.	22
Quadro 4 - Classes $B = \{\text{Tipo}\}$	23
Quadro 5 - Aproximação Inferior.....	24
Quadro 6 - Aproximação Superior.	24
Quadro 7 - Aproximações	25
Quadro 8 - Coeficiente de imprecisão.	25
Quadro 9 - Coeficiente da Qualidade da Aproximação Inferior.	26
Quadro 10 - Coeficiente da Qualidade da Aproximação Inferior.	26
Figura 6 - Cubo de decisão apresentado no trabalho de Zimmermann	28
Quadro 11 - Requisitos Funcionais	29
Figura 7 - Diagrama de Caso de Uso Administrador e Usuário.....	31
Quadro 12 - Requisitos Não-Funcionais	30
Figura 8 - Diagrama de Atividades de processo de análise TCA.....	33
Quadro 13 - Detalhamento do caso de uso - UC04 - Aplica a técnica TCA.....	32
Figura 9 - Modelo de entidade e relacionamento	34
Quadro 14 - Definição dos conjuntos.....	35
Quadro 15 - Conjuntos do sistema de informação	35
Quadro 16 - Relação de indiscernibilidade.....	36
Quadro 17 - Relação IP	36
Quadro 18 - Conjuntos P-elementares.....	36
Quadro 19 – Definições para os conjuntos de pacientes que apresentam e não apresentam gripe	37
Quadro 20 - Fórmula qualidade de aproximação	37
Quadro 21 - Resultados da qualidade de aproximação dos conjuntos P-Elementares	37

Figura 10 - Fragmento da rotina que realiza o relacionamento entre atributos.....	39
Figura 11 – Fragmento da rotina que define os conjuntos P-elementares.....	40
Figura 12 – Fragmento da rotina que define os valores P-Superior, P-Inferior, Fronteira e qualidade de aproximação dos conjuntos P-elementares.....	41
Figura 13 - Fragmento da rotina responsável por gerar as regras para a mineração na base de dados.....	42
Figura 14 - Tela de login	43
Figura 15 - Tela principal do sistema.....	44
Figura 16 - Cadastro de usuário.....	44
Figura 17 - Cadastro de atributo TCA.....	44
Figura 18 - Tela inicial análise TCA	45
Figura 19 - Seleção de atributos de análise	46
Figura 20 - Universo de objetos	46
Figura 21 - Informações cálculo TCA.....	48
Figura 22 - Cubo de decisão	48
Quadro 22 - Resultados obtidos	47
Quadro 23 - Resultados obtidos no exemplo de operacionalidade.....	47
Figura 23 - Fragmento da rotina de mineração.....	49
Figura 24 - Gráfico atendimentos por tipo de convênio.....	50
Figura 25 - Atendimentos por unidade internação	50
Figura 26 - Relatório de pacientes e sintomas.....	51
Figura 27 - Relatório de atendimentos por convênio	51
Quadro 24- Tabela ATRIBUTO_TCA.....	58
Quadro 25 - Tabela ITEM_ATRIBUTO_TCA.....	58
Quadro 26 - Tabela CONTROLA_ATRIB_TCA	58
Quadro 27 - Tabela CONTROLA_ITEM_ATRIB_TCA	58
Quadro 28 - Tabela ANALISE_TCA.....	58
Quadro 29 - Tabela APLICA_DADOS_TCA.....	59
Quadro 30 - Tabela RELACAO_CONJ_PELEMENTAR.....	59
Quadro 31 - Tabela TOTAL_APROX	59

LISTA DE TABELAS

Tabela 1 - Representação de um Sistema de Informação.....	22
Tabela 2 - Tabela de informação de diagnóstico de gripe.....	35
Tabela 3 - Resultados de aproximações determinísticas {C,M,T}.....	38
Tabela 4 - Resultados das aproximações determinísticas de {M,T}.....	38
Tabela 5 - Regras geradas.....	49

LISTA DE SIGLAS

BI – *Business Intelligence*

DM – *Data mining*

MER – Modelo de Entidade Relacional

UML – Linguagem de Modelagem Unificada

TCA – Teoria dos Conjuntos Aproximativos

SGBD – Sistema de Gerenciamento de Banco de Dados

ERP – *Enterprise Resource Planning*

LISTA DE SÍMBOLOS

\emptyset - conjunto vazio

\subset - esta contido

\Rightarrow - implica que

\cap - intersecção

\in - pertence

\cup - união

\forall - para todo ou qualquer que seja

\neq - diferente

\subseteq - subconjunto

SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 OBJETIVOS DO TRABALHO	14
1.2 ESTRUTURA DO TRABALHO	14
2 FUNDAMENTAÇÃO TEÓRICA	16
2.1 BUSINESS INTELLIGENCE.....	16
2.2 O PROCESSO DE <i>DATA MINING</i>	19
2.3 TEORIA DOS CONJUNTOS APROXIMATIVOS	20
2.3.1 Indiscernibilidade	22
2.3.2 Aproximação dos Conjuntos	23
2.3.3 Qualidade das Aproximações.....	25
2.3.4 Redução do Sistema de informação	26
2.4 CUBO DE DECISÃO	27
2.5 TRABALHOS CORRELATOS	27
3 DESENVOLVIMENTO	29
3.1 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO.....	29
3.2 ESPECIFICAÇÃO	30
3.2.1 Diagrama de Caso de Uso	30
3.2.2 Diagrama de Atividades	32
3.2.3 Modelo de entidade de Relacionamento (MER).....	34
3.2.4 Exemplo de cálculo TCA	35
3.3 IMPLEMENTAÇÃO	38
3.3.1 Técnicas e ferramentas utilizadas.....	39
3.3.2 Operacionalidade da implementação	42
3.4 RESULTADOS E DISCUSSÃO	52
4 CONCLUSÕES.....	54
4.1 EXTENSÕES	55
REFERÊNCIAS BIBLIOGRÁFICAS	56
DICIONÁRIO DE DADOS	58

INTRODUÇÃO

Entende-se que *Business Intelligence* (BI) é um conjunto de soluções referentes a análise e interpretação de dados. Torna-se de extrema necessidade para as organizações a missão de administrar as informações, pois elas apresentam-se como as principais armas nas tomadas de decisões gerenciais. Uma vez que essas informações sejam expostas em forma de relatório ou telas gráficas, de maneira organizada e coesa, tornam a tomada de decisão mais facilitada (DALFOVO; AMORIM, 2000).

O problema enfrentado pelas organizações está na falta da informação e na falta de organização na apresentação dessas informações, ou em muitos casos as informações que não tem um valor real, são entregues em momentos inoportunos para os tomadores de decisões. Sendo expostas de uma maneira a torná-las muito complexas para o entendimento, estas informações são consideradas de pouco valor para a organização.

Informações pertinentes podem sofrer variações de acordo com cada atributo de qualidade que a mesma apresenta. “Por exemplo, dados de inteligência de mercado até certo ponto imprecisos e incompletos são admissíveis, mas a apresentação no momento exato é essencial” (STAIR; REYNOLDS, 2006, p. 6)

A tecnologia *Data Mining* (DM) esta inserida no BI. O grande volume de dados disponíveis cresce a cada dia e desafia a capacidade de armazenamento, seleção e uso. Essa tecnologia com suas ferramentas permitem a “mineração” desses dados, a fim de gerar um real valor do dado, transformando-o em informação e conhecimento (REZENDE; ABREU, 2003, p. 215).

Conforme Rezende e Abreu (2003, p. 216), o DM é uma tecnologia que tem a capacidade de fazer a seleção de dados relevantes, no intuito de gerar informações e conhecimentos para as empresas. Este tem a capacidade “de aprender com base nos dados, extrair deduções, gerar informações com hipóteses, correlacionar coisas aparentemente desvinculadas, fazer previsões, revelar os atributos importantes, gerar cenários, relatar e descobrir conhecimentos” importantes para a gestão empresarial.

Diante deste contexto, do problema da falta de informação e da organização da informação, neste trabalho será utilizada a técnica de mineração. Embasada na Teoria dos Conjuntos Aproximativos (TCA), em uma base de dados clínica/ hospitalar com o objetivo de levantar as informações necessárias para os tomadores de decisão. De acordo com a parametrização do usuário, as informações poderão ser demonstradas de uma maneira

organizada e de fácil entendimento, através de gráficos e relatórios. A extensão da teoria dos conjuntos é a TCA, com o enfoque no tratamento de vagueza e incerteza em dados. “Foi inicialmente desenvolvida por *Zdzislaw Pawlak* no início da década de 80, década esta cujo preço e o desempenho dos computadores propiciaram o crescimento e surgimento de novas extensões para a TCA” (PESSOA; SIMÕES, 2004).

Este trabalho é voltado para a área da saúde por esta razão decidiu-se auxiliar os gestores nos diagnósticos das doenças, ou seja, os gestores irão informar uma série de sintomas, formando uma tabela, um universo de conjunto de atributos, dos quais o sistema irá aplicar o cálculo TCA gerando regras nas quais serão a base para a mineração de dados na base de dados, demonstrando os resultados através do cubo de decisão.

1.1 OBJETIVOS DO TRABALHO

O objetivo deste trabalho é o desenvolvimento de 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.

Os objetivos específicos do trabalho são:

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

1.2 ESTRUTURA DO TRABALHO

No capítulo um apresenta-se a introdução, onde é colocado o problema a ser resolvido e os objetivos que devem ser atingidos.

No capítulo dois apresenta-se a fundamentação do sistema, os embasamentos teóricos a respeito da Informação, Business Intelligence, o processo de Data Mining, Teoria dos Conjunto Aproximativos e o Cubo de decisão. Após o usuário informar as informações que desejar e o sistema realizar o cálculo da TCA, o usuário verificar estas informações no cubo de decisão, bem com nos relatórios e gráficos.

No capítulo três apresenta-se o desenvolvimento do trabalho, tais como a especificação, o sistema anterior, o sistema desenvolvido, a operacionalidade e os resultados.

No capítulo quatro apresentam-se as conclusões do trabalho desenvolvido e as sugestões para trabalhos futuros.

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo apresenta-se a fundamentação teórica referentes aos conceitos de *Business Intelligence* (BI), *Data Mining* (DM), Teoria dos Conjuntos Aproximativos (TCA), Cubo de decisão e trabalhos correlatos.

1.3 BUSINESS INTELLIGENCE

O termo *Business Intelligence* surgiu na década de 1970, contudo, este modelo vem sendo aplicado e desenvolvido empiricamente por civilizações antepassadas a milhares de anos atrás, entre estas civilizações encontram-se os egípcios, chineses, romanos, persas e fenícios (LEME FILHO, 2004, p. 3).

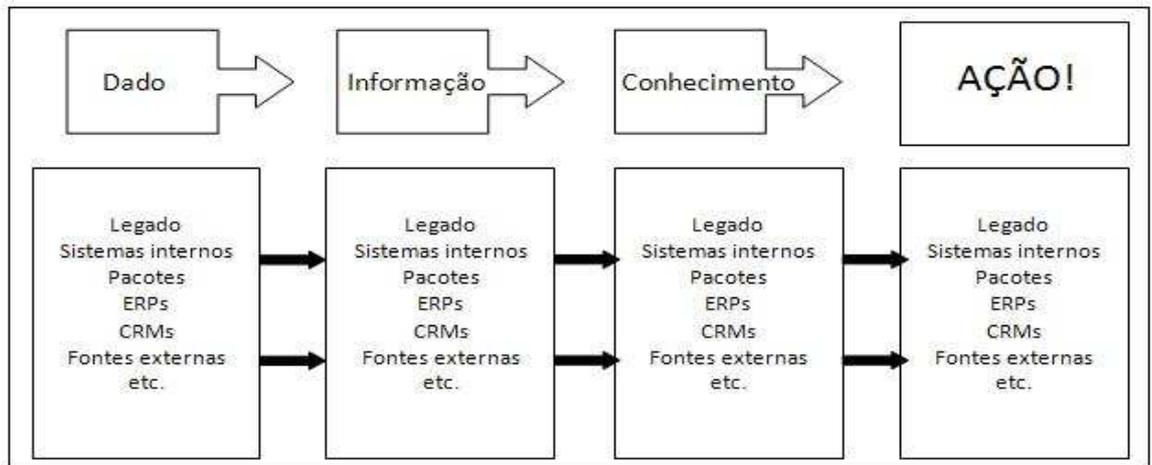
Estas civilizações cruzavam as informações provenientes da natureza, para tomar decisões que permitissem a melhoria na qualidade da vida de suas comunidades. Através do armazenamento de informações referentes ao ambiente em que estavam inseridos, tais como o estado da lua, a temperatura, o período da seca e de chuvas, o nível que o rio atingia, a posição dos astros, entre outras (LEME FILHO, 2004, p. 4).

Ainda no começo de sua história, onde os primeiros produtos de BI foram apresentados aos analistas de negócios apresentavam graves problemas relativos ao grande tempo exigido na programação e as informações não eram apresentadas em um tempo hábil, tampouco de uma forma flexível, além disso havia um alto custo de implantação (ROSINI; PALMISANO, 2003).

Mais tarde com o surgimento de inúmeras ferramentas, como os bancos de dados relacionais, e as interfaces gráficas, juntamente com o aumento da complexidade dos negócios, começaram a surgir produtos que mostravam um tempo de resposta mais satisfatório e uma maior flexibilidade na análise (ROSINI; PALMISANO, 2003).

Conforme Barbieri (2001, p.34), o conceito de BI é definido como “a utilização de variadas fontes de informação para se definir estratégias de competitividade nos negócios da empresa”. As organizações apresentam grandes volumes de dados o que ocasiona uma desordem dificultando o processo de tomada de decisão dos executivos, os quais ficam

perdidos em meio a esse amontoado sem nenhuma conclusão, a figura 1 mostra esse volume de dados até o conhecimento necessário para as tomadas de decisões (BARBIERI, 2001).



Fonte: Leme Filho (2004).

Figura 1 - Agregando valor ao dado para a tomada de decisão.

Leme Filho (2004, p. 2), acrescenta que BI é um modelo que auxilia os executivos, que tem influência direta nos negócios da organização, usado para extrair, transformar e analisar todo esse imenso volume de dados, apresentando ao final da sua aplicação o conhecimento necessário para o auxílio dos gestores para a tomada de decisão.

Abaixo são descritas as características dos sistemas BI (SERRA, 2002):

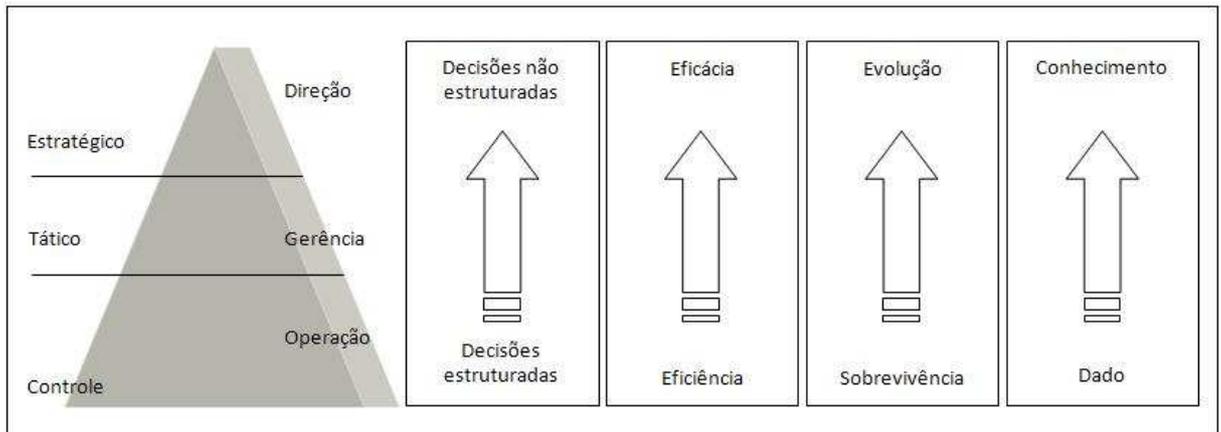
- a) extrair e integrar dados de múltiplas fontes;
- b) fazer uso da experiência;
- c) analisar dados contextualizados;
- d) trabalhar com hipóteses;
- e) procurar relações de causa e efeito;
- f) transformar os registros obtidos em informação útil para o conhecimento empresarial.

Ainda segundo Serra (2002) as principais ferramentas do BI são:

- a) Data Warehouse e Data Marts;
- b) Database Marketing;
- c) Customer Relationship Management;
- d) Data Mining;
- e) Balanced Scorecard.

A evolução do negócio esta ligada ao topo da hierarquia, a figura 2 ilustra como uma organização está dividida, onde a direção tem o controle das funções estratégicas, tendo como principal função a criação dos processos, estes serão controlados pela gerência e por fim serão

executados pela operação. “Com a cadeia produtiva em perfeito funcionamento, a direção não trabalha mais com dados, tampouco com informações, mas sim com conhecimento” (LEME FILHO, 2004, p. 10). Com isso, as ações passam a visar a eficácia, onde a surge a questão chave: estamos fazendo a coisa certa?. “O BI deve atender as necessidades de mercado, sustentando a evolução do negócio, possibilitando responder questões como “estamos fazendo a coisa certa?” (LEME FILHO, 2004, p. 10).



Fonte: Leme (2004).

Figura 2 - Tecnologia da informação nas organizações.

A grande importância no uso do BI esta na utilização dos dados como matéria-prima permitindo a transformação deste em informação, permitindo que os envolvidos no processo passem a lapidar esta informação transformando-a em conhecimento, como mostra a figura 3 (LEME FILHO, 2004, p. 10).



Fonte: Leme(2004)

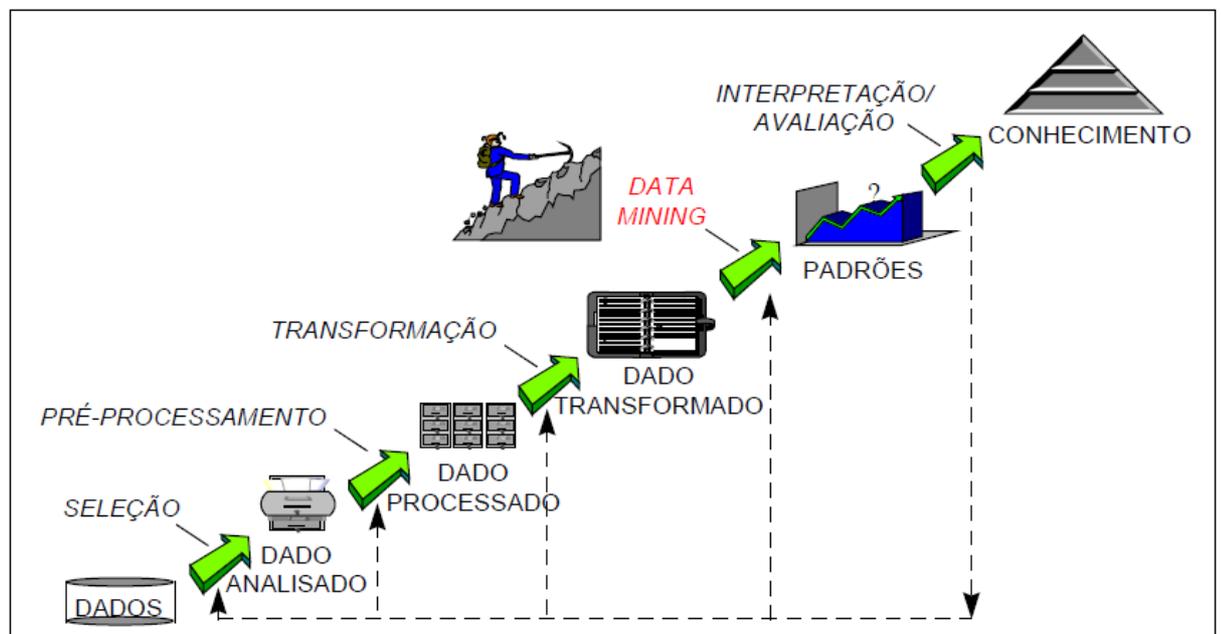
Figura 3 - Do dado ao conhecimento.

1.4 O PROCESSO DE *DATA MINING*

Essa tecnologia é formada por um conjunto de ferramentas que, por meio do uso de algoritmos de aprendizado ou baseados em redes neurais e estatística, são capazes de explorar um grande conjunto de dados transformando-os em conhecimento (REZENDE; ABREU, 2003, p. 215).

Todos os dias as empresas inserem inúmeros dados em seus bancos de dados, fazendo com que esses se tornem verdadeiras minas de ouro de informações sobre os mais diversificados processos utilizados dentro das organizações. “Todos esses dados podem contribuir com a empresa, sugerindo tendências e particularidades pertinentes a ela e seu meio ambiente interno e externo, visando a uma rápida ação de seus gestores” (REZENDE; ABREU, 2003, p. 215).

Para que seja possível a extração de conhecimento de uma base de dados é necessário um conjunto de processos focados em um determinado domínio, a partir de uma base de dados bruta. Esses processos estão constituídos em: seleção, pré-processamento, transformação, mineração de dados (*Data Mining*) e interpretação/ avaliação, abaixo pode-se observar através da figura 4 todo o processo de DM (ALMEIDA et al., 2004).



Fonte: Almeida et AL (2004).

Figura 4 - Etapas do processo de Data Mining.

Seleção é a principal característica da mineração e sua finalidade é a de definir o conjunto de dados a ser analisado, deixando de fora os dados que não serão relevantes para o

processo de extração do conhecimento exclusivamente do domínio proposto (RAMOS; SANTOS, 2003).

O pré-processamento, conforme Ramos e Santos (2003), “Passa essencialmente pela redução do espaço de pesquisa, isto é, pela diminuição de linhas/colunas a analisar”. Almeida et al. (2004), acrescenta “que outra tarefa importante é a verificação de predominância de classes, sendo que nestes casos, deve-se eliminar alguns dos registros da classe predominante ou acrescentar registros das outras classes”.

A transformação consiste na preparação da base de dados, onde por sua vez sofrerão formatações para que seja possível a aplicação dos algoritmos de extração do conhecimento (ALMEIDA et al., 2004).

O *Data Mining*, por sua vez, tem como objetivo construir hipótese. Para a construção dessas hipóteses são utilizados sistemas especialistas. Conforme Almeida et al. (2004), “Um sistema de aprendizado é um programa de computador que toma decisões baseadas em experiências acumuladas contidas em exemplos resolvidos com sucesso”.

Na interpretação/avaliação todo conhecimento extraído será analisado e caso o conhecimento apresentado não estiver de acordo com o esperado o processo pode ser reiniciado, alterando a base de dados ou modificando os parâmetros informados inicialmente, porém é de suma importância que a análise deva ser feita em conjunto com os especialistas deste determinado domínio (ALMEIDA et al., 2004).

A aplicação do *Data Mining* torna possível comprovar o pressuposto da transformação de dados em informação e posteriormente em conhecimento. Esta possibilidade torna a técnica imprescindível para o processo de tomada de decisão. Para chegar-se a este resultado, é preciso investigar o uso efetivo do conhecimento obtido pelo *Data Mining* no processo de tomada de decisão, bem como os impactos que teve na solução efetiva de problemas e ações propostas e concretizadas. (QUONIAM et al., 2001, p. 20).

1.5 TEORIA DOS CONJUNTOS APROXIMATIVOS

“No início da década de 80, surgiu uma teoria, caracterizada pela simplicidade e bom formalismo matemático, o que facilita a manipulação de informações, em especial, incertas, conhecida como Teoria dos Conjuntos Aproximativos (TCA), do inglês *Rough Set Theory*.” (PESSOA; SIMÕES, 2004).

A TCA é uma extensão da teoria dos conjuntos, onde tem como principal objetivo o tratamento das incertezas dos dados utilizando a relação de indiscernibilidade “que diz que dois elementos são ditos indiscerníveis, se possuem as mesmas propriedades, segundo Leibniz” (PESSOA; SIMÕES, 2004). A outra relação é em torno das reduções, que nada mais são que a retirada dos atributos que não serão utilizados, que não agregam valor para a base de conhecimento analisada, a figura 5 representa esta redução (PESSOA; SIMÕES, 2004).



Fonte: Patrício, Pinto e Souza (2005).

Figura 5 - Representação da redução da base de dados.

O espaço aproximativo é representado por um par ordenado $S = (U, A)$, onde U é um conjunto finito não-vazio de objetos chamado de universo e A é um conjunto finito não vazio de atributos, tal que $U \rightarrow Va$ para todo $a \in A$. O conjunto Va é chamado de conjuntos de valores de Va (DOMINGOS, UCHÔA, 2006).

Uma forma comum de representação da abordagem da TCA é através de um sistema de informação que nada mais é que um conjunto de objetos, sendo que cada objeto tem uma quantidade de atributos. E esses atributos são os mesmos para cada um dos objetos, contudo, nominalmente estes valores podem diferir. “ (PATRÍCIO; PINTO; SOUZA, 2005).

Muitos sistemas utilizam ainda um atributo denominado atributo de decisão. Podendo ser um sistema qualquer de informação na forma de $S = (U; A \cup \{d\})$, onde $d \notin A$ é o atributo de decisão. Os elementos de A são chamados atributos condicionais ou simplesmente condições (PESSOA; SIMÕES, 2004).

Pode-se resumir um sistema de decisão com regras, como por exemplo mostra o quadro 1.

Se a = '1' e b = '2' então d = 'Ótimo'
--

Quadro 1 - Exemplo de regra de decisão

A seguir, através da tabela 1, pode-se visualizar um sistema de informação, no qual apresenta dados referentes a pacientes e algumas características de seus ferimentos, onde o objetivo é que o sistema de informação seja capaz de fornecer a atitude das enfermeiras com

relação aos ferimentos.

Tabela 1 - Representação de um Sistema de Informação.

U	Atributos					Atributo Decisão
Pessoa	Tipo	Tamanho	Inflamação	Cicatrização	Classificação	Atitude
1	Oclusivo	Grande	Fase reparadora	Primeira intenção	Ferimento aberto	Negativa
2	Úmido	Médio	Fase exsudativa	Segunda intenção	Ferimento acidental	Neutra
3	Seco	Pequeno	Fase vascular	Terceira intenção	Ferimento fechado	Positiva
4	Oclusivo	Médio	Fase exsudativa	Terceira intenção	Ferimento aberto	Negativa
5	Seco	Pequeno	Fase vascular	Primeira intenção	Ferimento aberto	Neutra
6	Compressivo	Grande	Fase reparadora	Segunda intenção	Ferimento acidental	Positiva
7	Seco	Pequeno	Fase reparadora	Primeira intenção	Ferimento intencional	Positiva
8	Seco	Pequeno	Fase reparadora	Primeira intenção	Ferimento aberto	Positiva
9	Compressivo	Grande	Fase reparadora	Segunda intenção	Ferimento acidental	Neutra
10	Compressivo	Médio	Fase exsudativa	Segunda intenção	Ferimento aberto	Neutra

Fonte: Adaptado de Patrício, Pinto e Souza (2005).

Para o sistema de informação apresentado na tabela 1, tem-se o quadro 2.

- $U = \{1,2,3,4,5,6,7,8,9,10\}$, representa o conjunto de objetos denominado universo;
- $A = \{Tipo, Tamanho, Inflamação, Cicatrização, Classificação\}$, representa o conjunto de objetos denominado atributos;
- $D = \{Atitude\}$, representa os atributos de decisão.

Quadro 2 - Definição do conjuntos dentro do universo

1.5.1 Indiscernibilidade

Em uma base de dados pode-se encontrar muitas informações “iguais” ou ainda muitos atributos desnecessários. No que diz respeito à repetição das informações, estes objetos muitas vezes tem uma relação de equivalência, esta relação acaba possibilitando a capacidade de tratar este problema, pois, apenas um objeto pode representar toda uma classe, essa relação pode ser verificada a seguir (PATRÍCIO; PINTO; SOUZA, 2005).

Tendo um sistema de informação representado através de $S = (U; A)$, então com qualquer $B \subseteq A$ existe uma relação de equivalência $IND_A(B)$, indicado pelo quadro 3.

$$IND_A(B) = \{(x, x') \in U \mid \forall a \in B, a(x) = a(x')\}$$

Quadro 3 - Relação de equivalência.

$IND_A(B)$ é chamada de relação de indiscernibilidade. Se $(x, x') \in IND_A(B)$, então objetos x e x' são indiscerníveis, são iguais relativamente a qualquer atributo de B . A classe de equivalência da relação determinada por x pertencente a X é denotado $[x]_B$.

Alguns exemplos desta relação de equivalência podem ser verificados, utilizando o universo de dados da tabela 1, através de um subconjunto $B = \{Tipo\}$, onde os objetos 1 e 4 estão na mesma classe de equivalência e são indiscerníveis, assim como os objetos 3, 5, 7, e 8, e os objetos, 6, 9, 10 (PESSOA; SIMOES, 2004).

Cada um desses grupos origina uma classe de equivalência, como mostra o quadro 4.

	Atributos
Classes	Tipo
C11	Oclusivo
C12	Seco
C13	Compressivo

Fonte: Adaptado de Patrício, Pinto e Souza (2005).

Quadro 4 - Classes $B = \{Tipo\}$

1.5.2 Aproximação dos Conjuntos

A relação de equivalência propicia a partição do universo. Estas partições podem ser usada para a criação de novos subconjuntos do universo. “Subconjuntos que são freqüentemente de interesse têm o mesmo valor do atributo de resultado”. Contudo, alguns elementos podem não ser definidos claramente devido a serem indiscerníveis, ocasionando decisões contraditórias. Desta maneira, acaba surgindo à noção de conjunto aproximativo (*Rough Set*). Por sua vez, estes elementos dividem-se em três classes: os que podem ser classificados pertencentes à classe desejada, os que não pertencem à classe desejada e os que não podem ser classificados. A seguir são expressas formalmente estas noções (PATRÍCIO; PINTO; SOUZA, 2005).

Seja o sistema de informação $S = (U; A)$, tem-se $B \subseteq A$ e $X \subseteq U$. Onde X é o conjunto de objetos ou registros com respeito à B , isto é, utilizando somente as informações dos atributos contidos em B é possível obter X . Assim define-se Aproximação Inferior de X em relação à B , indicado por \underline{BX} , conforme pode ser visto no quando 5 e Aproximação

Superior de X em relação a B , indicado por \overline{BX} , apresentado no quadro 6 (PESSOA; SIMÕES, 2004).

$$\underline{BX} = \{x | [x]_B \subseteq X\}$$

Quadro 5 - Aproximação Inferior.

$$\overline{BX} = \{x | [x]_B \cap X \neq \emptyset\}$$

Quadro 6 - Aproximação Superior.

Na base de conhecimento de B , os objetos da Aproximação Inferior \underline{BX} podem ser classificados claramente como membros de X , enquanto que os objetos da Aproximação Superior \overline{BX} podem apenas serem classificados como possíveis membros de X . Portanto, obtém-se o conjunto $BN_B(x) = \overline{BX} - \underline{BX}$ que é denominado região de fronteira de X , estes objetos não podem ser classificados pertencentes a X na base de conhecimento de B . Ainda é possível definir como Fora da Região X o conjunto $E_B(x) = U - \overline{BX}$, consiste de objetos que não pertencem a X , considerando a mesma base de conhecimento B (PESSOA; SIMÕES, 2004).

Um conjunto de X é determinado como *rough* (impreciso) caso sua região de fronteira for diferente do conjunto vazio ($RF(X) \neq \emptyset$). Entretanto, se o conjunto for vazio ($RF(X) = \emptyset$) é determinado como *crisp* (preciso) (DOMINGOS; UCHÔA, 2006).

Utilizando-se do sistema de informação da tabela 1, para exemplificar a aproximação dos conjuntos, tem-se a seguinte questão: quais as características dos atributos que definem as atitudes das enfermeiras com relação aos ferimentos como sendo Negativa, Neutra ou Positiva? Nota-se que não há uma única resposta a este questionamento, pois os pacientes 6 e 9 apresentam os mesmos atributos, diferenciando-se apenas no atributo de decisão. Diferentemente dos pacientes 1 ou 4 que terão atitude Negativa, assim como os pacientes 2, 5 ou 10 que apresentam atitude Neutra. Nestes casos, surge a necessidade da utilização das noções da *Rough Sets* (PATRÍCIO; PINTO; SOUZA, 2005).

O conjunto de pacientes que correspondem ao atributo de decisão a Atitude = Positiva, dado $S = \{Tipo, Tamanho, Inflamação, Cicatrização, Classificação\}$, $U / IND_A(B) = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}, \{10\}\}$ e $X = \{3, 6, 7, 8\}$, tem-se as aproximações demonstradas no quadro 7.

- a) Aproximação Inferior: $\underline{BX} = \{\{3\},\{7\},\{8\}\}$;
- b) Aproximação Superior: $\overline{BX} = \{\{3\},\{6,9\},\{7\},\{8\}\}$;
- c) Região de Fronteira: $RF(X) = \{6,9\}$;
- d) Fora da Região: $U - \overline{BX} = \{\{1\},\{2\},\{4\},\{5\},\{10\}\}$.

Quadro 7 - Aproximações

1.5.3 Qualidade das Aproximações

“A qualidade das aproximações obtidas pelas definições dadas previamente pode ser caracterizada numericamente a partir dos próprios elementos que a definem” (PATRÍCIO; PINTO; SOUZA, 2005). O coeficiente utilizado para medir essas qualidades é indicado por $\alpha_B(X)$, sendo X o conjunto de registros ou objetos referentes à B , e podem ser realizadas de três formas (PESSOA; SIMÕES, 2004).

- a) Coeficiente de imprecisão $\alpha_B(X)$, que pode ser entendido como a qualidade da aproximação de X , indicado pelo quadro 8.

$$\alpha_B(X) = \frac{|\underline{BX}|}{|\overline{BX}|}$$

Quadro 8 - Coeficiente de imprecisão.

Onde as cardinalidades das aproximações Inferior e Superior são indicados respectivamente por $|\underline{BX}|$ e $|\overline{BX}|$ e são conjuntos não-vazios. Se $\alpha_B(X) < 1$, X é dito como *rough* (impreciso) em relação ao conjunto de atributos de B . Caso $\alpha_B(X) = 1$, X é dito como *crisp* (preciso), em relação ao conjunto de atributos de B .

- b) Coeficiente da Qualidade da Aproximação Inferior $\alpha_B(\underline{BX})$, indica o percentual de todos os objetos que são classificados como pertencentes em X , indicado pelo quadro 9.

$$\alpha_B(\underline{BX}) = \frac{|\underline{BX}|}{|U|}$$

Quadro 9 - Coeficiente da Qualidade da Aproximação Inferior.

- c) Coeficiente da Qualidade da Aproximação Superior $\alpha_B(\overline{BX})$, indica o percentual de todos os objetos possivelmente classificados como pertencentes a X , como indicado pelo quadro 10.

$$\alpha_B(\overline{BX}) = \frac{|\overline{BX}|}{|U|}$$

Quadro 10 - Coeficiente da Qualidade da Aproximação Inferior.

1.5.4 Redução do Sistema de informação

A redundância da informação dentro do sistema de informação acaba gerando uma complexidade muito alta, de tal forma que a extração de regras torna-se uma tarefa cansativa e árdua. Anteriormente foi visto que um artifício natural para a redução de dados é a identificação das classes de equivalência, pois utiliza-se apenas um objeto para a representação de uma classe inteira, o que acaba constituindo uma otimização desse sistema (PATRÍCIO; PINTO; SOUZA, 2005).

“Um outro artifício para redução é manter somente os atributos que preservam a relação de indiscernibilidade”, de tal modo que um conjunto reduzido de atributos forneça a mesma qualidade de classificação em relação a um conjunto original de atributos. Habitualmente existem vários subconjuntos de atributos e os que são mínimos são chamados de reduções (PESSOA; SIMOES, 2004).

Dado o sistema de informação $S = (U, A)$, para que se tenha uma redução de S em um conjunto mínimo de atributos, através $B \subseteq A$ tal que $IND_s(B) = IND_s(A)$, ou seja, uma redução ($RED(B)$) é o conjunto mínimo de atributos de A que pode ser utilizado preservando o conjunto de atributos do universo completo (PESSOA; SIMOES, 2004).

Um sistema de informação pode ter mais de uma redução. A intersecção de todas as

reduções é chamada de núcleo, que é dada por $N(B) = \bigcap_{i=1..n} RED_i(B)$. (PATRÍCIO; PINTO; SOUZA, 2005).

1.6 CUBO DE DECISÃO

Cubo de decisão é o grupo de componentes que auxiliam nas decisões, onde podem ser utilizados para relacionar tabelas de um banco de dados, que por sua vez são apresentados em forma de planilhas e gráficos. “Nesta etapa é envolvido cálculo de dados que o usuário virá a solicitar, mas que podem ser derivados de outros dados. Quando o usuário solicitar os dados, estes já estarão calculados e agregados em um Cubo de Decisão” (GHODDOSI; DALFOVO; MAIA, 2004).

1.7 TRABALHOS CORRELATOS

A seguir são apresentados três trabalhos que se relacionam de forma direta com o sistema desenvolvido, entre os quais, Barreto (2003), Schmitt (2007) e Zimmermann (2006).

No trabalho de Schmitt (2007) desenvolveu-se um aplicativo utilizando *Data Mining* baseado na Teoria dos Conjuntos Aproximativos. Basicamente consiste em aplicar a TCA em uma base de dados de recursos humanos, para a identificação de perfis dos profissionais cadastrados que mais se adequavam para comporem as equipes de projetos de software, com isso auxiliando o executivo no processo de seleção.

O objetivo principal do trabalho de Zimmermann (2006) foi o desenvolvimento de uma aplicação de Sistema de Apoio à Decisão baseado em *Business Intelligence* em que o sistema realizava a busca das informações no banco apresentando-as através do cubo de decisão, possibilitando o usuário de realizar o cruzamento das informações. Como mostra a figura 6 a seguir.

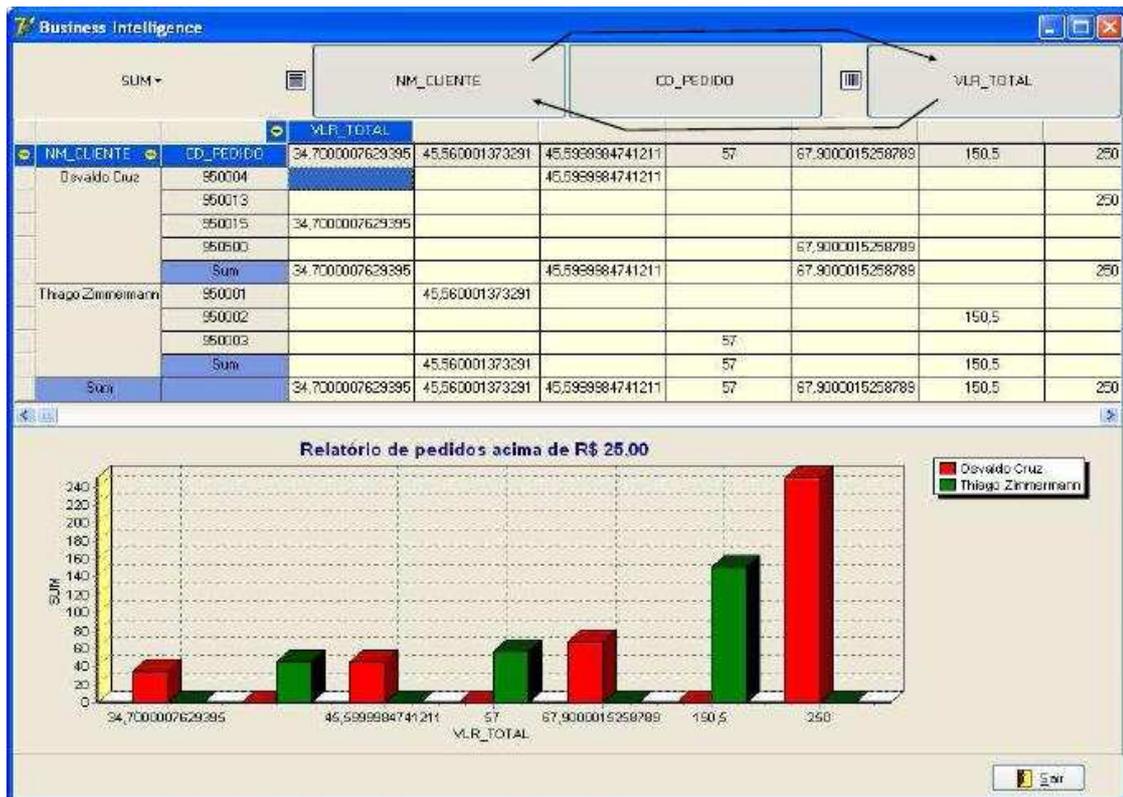


Figura 6 - Cubo de decisão apresentado no trabalho de Zimmermann

A dissertação de Barreto (2003) teve como objetivo trazer a definição e a comparação da utilização do conjunto de ferramentas em Business Intelligence entre duas companhias, Microsoft e Oracle dentro das organizações, apresentando-as como uma ferramenta de suma importância para a obtenção das informações necessárias no escopo da gerência.

Correlacionando os trabalhos citados anteriormente, o trabalho desenvolvido visa à utilização da técnica TCA em uma base de dados clínica/ hospitalar. Diferentemente de Schmitt que a utilizou em uma base de dados de Recursos Humanos. Através de uma ferramenta integrada no BI, à técnica de mineração de dados *Data Mining*, para a seleção e coleta das informações. Para a apresentação das informações para o usuário será utilizado o cubo de decisão, além de relatórios e gráficos, assim como Zimmermann (2006) e Barreto (2003) utilizaram em seus trabalhos. Através do uso do cubo de decisão o usuário terá a possibilidade de cruzar as informações da mineração de dados.

DESENVOLVIMENTO

O sistema desenvolvido neste trabalho realiza a técnica de mineração de dados com base na TCA. Através do estudo do funcionamento da TCA a fim de gerar o conhecimento a partir das informações fornecidas pelos usuários da organização. Este capítulo apresenta a realização e análise dos requisitos que definem do sistema desenvolvido. A seguir são descritas as suas especificações.

1.8 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

Os requisitos descrevem o que o sistema deve e o que não deve fazer. Os quais são classificados como Requisitos Funcionais (RF) e os Requisitos Não Funcionais (RNF). Os RF correspondem às funcionalidades e o comportamento que o sistema deve possuir e o comportamento do sistema em determinadas situações, podendo também apresentar o que o sistema não deve realizar. Os RNF são vistos como restrições que o sistema terá sobre alguns serviços ou funções oferecidas, como hardware, segurança e usabilidade.

No quadro 11 apresentam-se os requisitos funcionais do sistema e seus respectivos vínculos com os casos de uso.

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

Quadro 11 - Requisitos Funcionais

No quadro 12 apresentam-se os Requisitos Não-Funcionais do sistema, estes requisitos poderiam ser ou não implementados.

Requisitos Não Funcionais
RNF01: O sistema deverá ser desenvolvido para o banco de dados Oracle.
RNF02: O sistema deverá ser construído sobre a plataforma de desenvolvimento Delphi.
RNF03: O sistema deverá possuir um controle de acesso somente em nível de usuário.

Quadro 12 - Requisitos Não-Funcionais

1.9 ESPECIFICAÇÃO

Neste capítulo são apresentadas as especificações como Diagrama de Casos de Uso, Diagrama de Atividades e Diagrama de Entidade de Relacionamento (MER). Para a criação dos Diagramas de Casos de Uso e de Atividades foi utilizada a ferramenta *Enterprise Architect*, enquanto que o diagrama de entidade de relacionamento foi utilizada a ferramenta *DBDesigner*.

1.9.1 Diagrama de Caso de Uso

Os casos de uso têm como função representar as principais funcionalidades que se pode observar em um sistema e dos elementos externos que interagem com o mesmo (BEZERRA, 2002).

Na figura 7 é apresentado o diagrama de caso de uso administrador e usuário, onde o administrador controla quais bases de dados podem ser acessadas e realiza a manutenção do cadastro de usuários. Por sua vez, o usuário realiza análise TCA, pode visualizar o cubo de decisão, os gráficos e os relatórios.

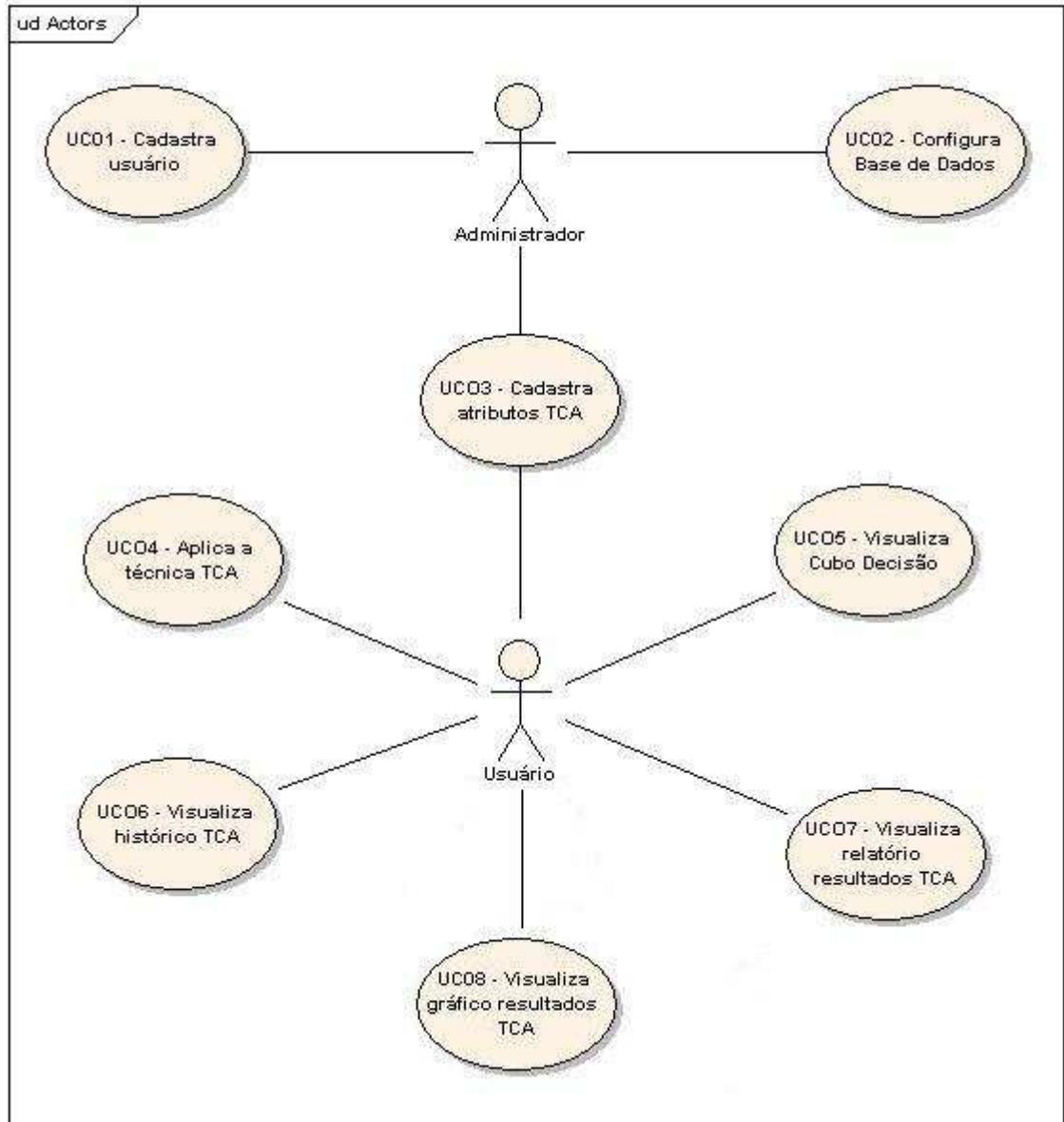


Figura 7 - Diagrama de Caso de Uso Administrador e Usuário

A seguir é realizado o detalhamento dos principais casos de uso correspondentes ao diagrama de casos de uso. No quadro 13 é apresentado o detalhamento do caso de uso UC04 – Aplica a técnica TCA, o qual está relacionado aos usuários.

Descrição Caso de Uso	UC04 – Aplicar a técnica TCA
Objetivo	Aplicar o cálculo TCA
Ator	Usuário
Pré-condições	- O usuário deverá possuir um cadastro ativo no sistema. - O usuário deverá estar logado.
Cenário Principal	<ol style="list-style-type: none"> 1. O usuário clica no botão novo. 2. O sistema apresenta a tela de seleção de atributos para que o usuário possa selecionar entre os atributos cadastrados previamente no sistema. 3. Após o usuário selecionar os atributos clica no botão próximo.

	<ol style="list-style-type: none"> 4. O sistema apresenta a tela de universo de atributos, onde o usuário deve informar os valores para cada atributo escolhido anteriormente. 5. Após o usuário informar o universo de atributos clica no botão próximo. 6. O sistema apresenta as informações do cálculo TCA.
Cenário Alternativo	No passo 1, caso o usuário queira, pode acessar a análise TCA através do botão da barra de botões rápida na tela principal do sistema.
Cenário Exceção	<p>No passo 5, caso o usuário clique no botão anterior o universo de dados será perdido, tendo que informar novamente o universo.</p> <p>Nos passos 3, 4, 5 o usuário pode cancelar a operação de aplicação do cálculo TCA clicando no botão cancelar.</p> <p>No passo 6, o usuário poderá visualizar as informações geradas através do cubo de decisão, de gráficos ou relatórios.</p>
Pós-condições	1. Cálculo TCA encerrado.

Quadro 13 - Detalhamento do caso de uso - UC04 - Aplica a técnica TCA

1.9.2 Diagrama de Atividades

Um diagrama de atividades tem como característica centrada apresentar o fluxo de uma atividade para outra em um sistema (BEZERRA, 2002). Na figura 8 a seguir apresenta-se o diagrama de atividades do usuário realizando o processo de entrada das informações de seu conhecimento para que seja realizado o cálculo da TCA.

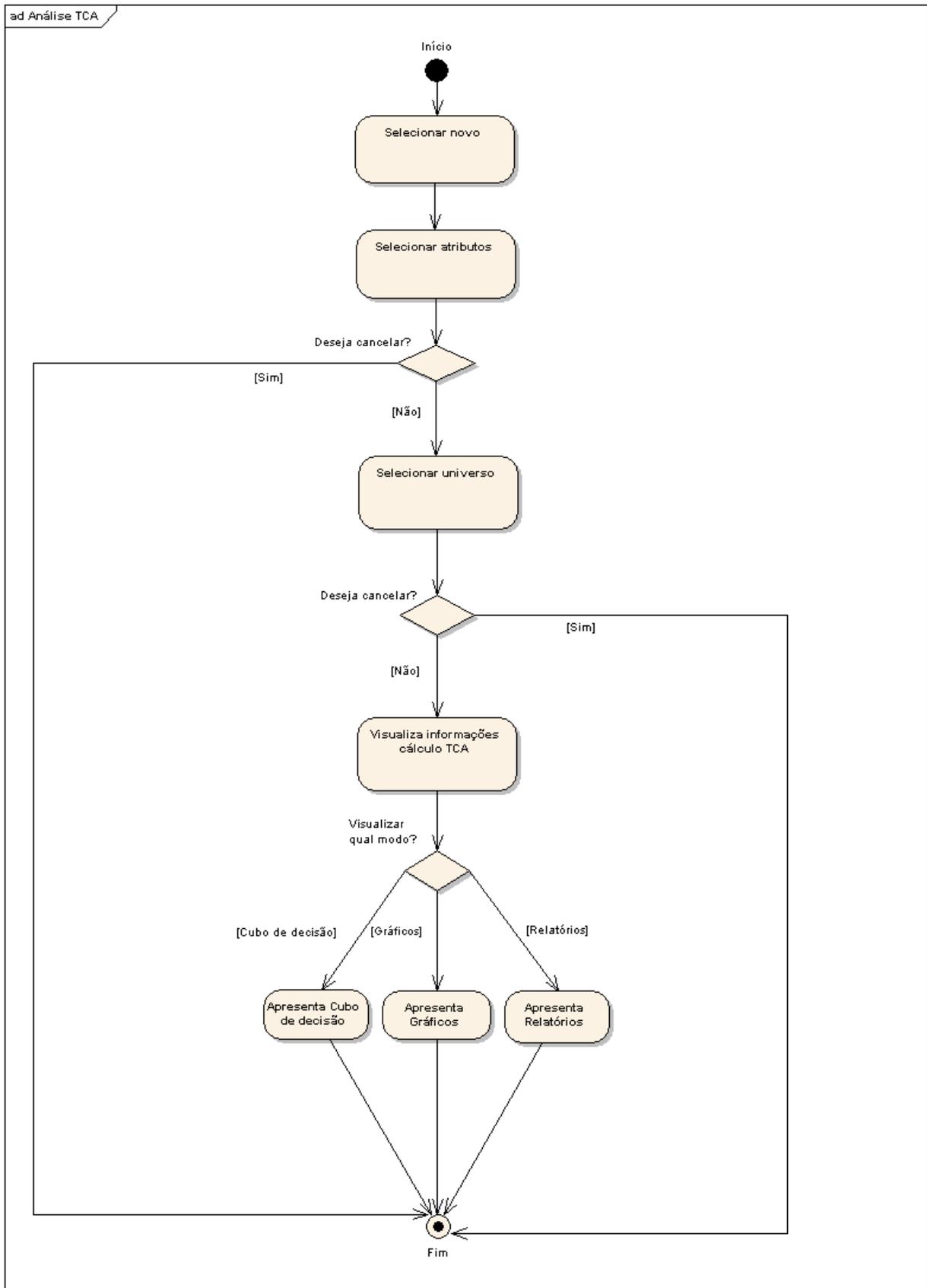


Figura 8 - Diagrama de Atividades de processo de análise TCA

1.9.3 Modelo de entidade de Relacionamento (MER)

Este modelo não se preocupa com os aspectos de implementação, como por exemplo, estruturas físicas e formas de acesso de um SGBD. O modelo conceitual é aquele que apresenta os objetos, suas características e relacionamento. Usufruindo deste modelo é possível criar uma descrição da realidade fácil de entender e de se interpretar (BEZERRA, 2002).

A figura 9 apresenta um modelo físico entidade-relacionamento da base de dados utilizada para realizar o cálculo TCA.

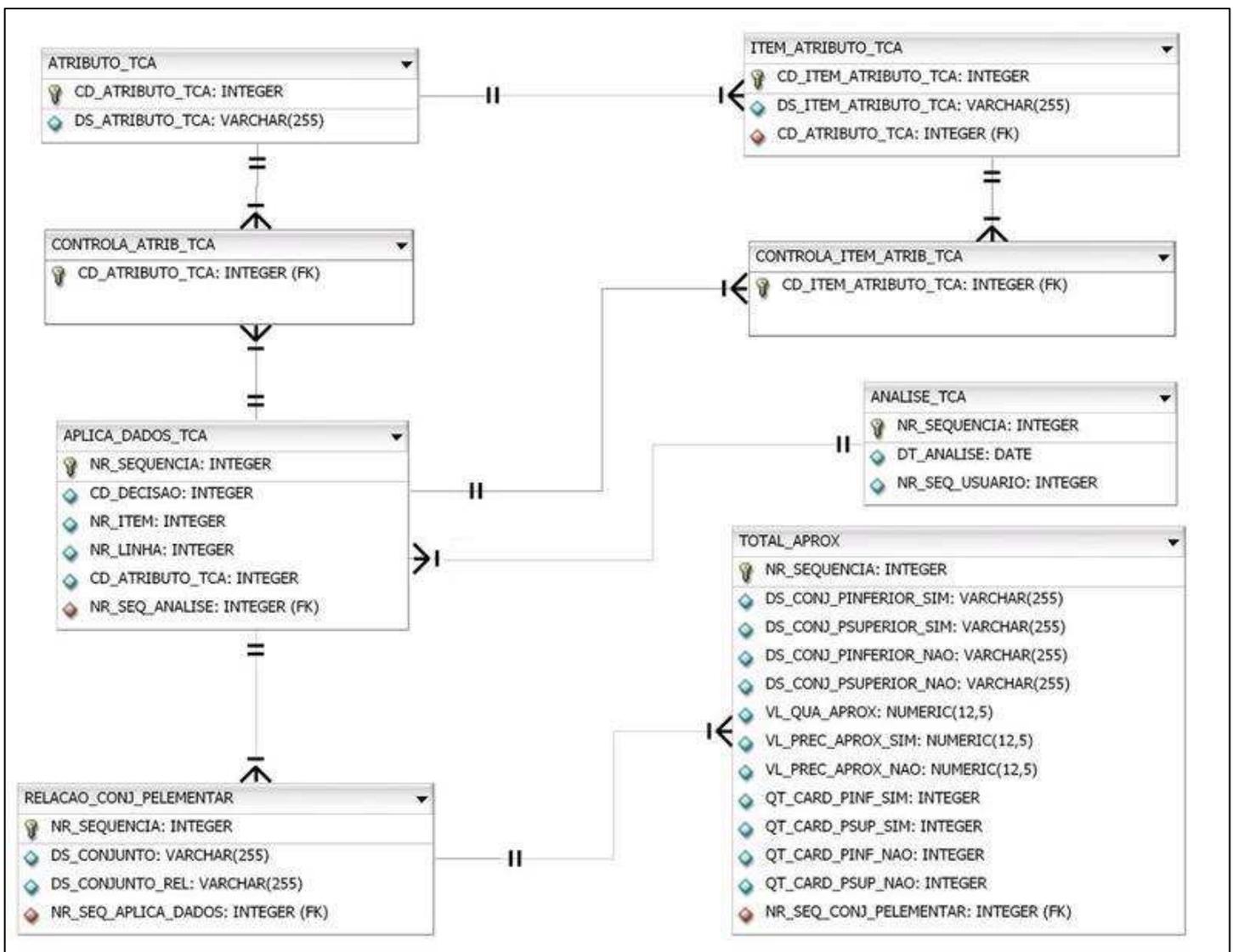


Figura 9 - Modelo de entidade e relacionamento

1.9.4 Exemplo de cálculo TCA

Para um melhor entendimento do funcionamento do cálculo da TCA, a seguir é apresentado um exemplo utilizando os conceitos e fórmulas apresentados na seção 2.3.

Inicialmente, considerando o seguinte sistema de informação, apresentado na tabela 2.

Tabela 2 - Tabela de informação de diagnóstico de gripe

Paciente	Dor de cabeça(C)	Dor muscular (M)	Temperatura(T)	Gripe (G)
1	Não	Sim	Alta	Sim
2	Sim	Não	Alta	Sim
3	Sim	Sim	Muito Alta	Sim
4	Não	Sim	Normal	Não
5	Sim	Não	Alta	Não
6	Não	Sim	Muito Alta	Sim

Fonte: Pawlak(1982).

A partir do sistema de informação considerado, tem-se os conjuntos U, C D e Q descritos conforme quadro 14.

$U = \{1,2,3,4,5,6\}$ é o conjunto de objetos de universo;
 $C = \{C, M, T\}$ é o conjunto de atributos de condição;
 $D = \{G\}$ é o conjunto de atributos de decisão;
 $Q = C \cup D = \{C, M, T, G\}$ é o conjunto de todos os atributos.

Fonte: Pawlak(1982).

Quadro 14 - Definição dos conjuntos

Com isso, tem-se representados no quadro 15 os valores assumidos de cada função do sistema de informação proposto.

$f(1, C) = \text{Não}; f(1, M) = \text{Sim}; f(1, T) = \text{Alta}; f(1, G) = \text{Sim};$
 $f(2, C) = \text{Sim}; f(2, M) = \text{Não}; f(2, T) = \text{Alta}; f(2, G) = \text{Sim};$
 $f(3, C) = \text{Sim}; f(3, M) = \text{Sim}; f(3, T) = \text{MuitoAlta}; f(3, G) = \text{Sim};$
 $f(4, C) = \text{Não}; f(4, M) = \text{Sim}; f(4, T) = \text{Normal}; f(4, G) = \text{Não};$
 $f(5, C) = \text{Sim}; f(5, M) = \text{Não}; f(5, T) = \text{Alta}; f(5, G) = \text{Não};$
 $f(6, C) = \text{Não}; f(6, M) = \text{Sim}; f(6, T) = \text{MuitoAlta}; f(6, G) = \text{Sim};$

Fonte: Pawlak(1982).

Quadro 15 - Conjuntos do sistema de informação

De forma a diagnosticar a gripe, particiona-se o U do sistema de informação proposto, de acordo com a relação de indiscernibilidade da Lina ID sobre o atributo de decisão G,

obtendo-se o quadro 16.

$Y = U / ID = \{Y_1, Y_2\}$, onde
 $Y_1 = \{1,2,3,6\}$, conjunto de pacientes que apresentam a gripe;
 $Y_2 = \{4,5\}$, conjunto de pacientes que não apresentam a gripe.

Fonte: Pawlak(1982).

Quadro 16 - Relação de indiscernibilidade

Seja o conjunto de atributos $P = \{M, T\}$ a relação de indiscernibilidade que são os apresentados no quadro 17.

$IP = \{(1,1), (2,2), (2,5), (3,3), (3,6), (4,4), (5,2), (5,5), (6,3), (6,6)\}$
 $U / I_p = \{\{1\}, \{2,5\}, \{3,6\}, \{4\}\}$

Fonte: Pawlak(1982).

Quadro 17 - Relação IP

Desta maneira, tem-se descritos os conjuntos P-elementares em U / I_p demonstrados no quadro 18.

$Des_p(\{1\}) = \{(M, Sim), (T, Alta)\};$
 $Des_p(\{2,5\}) = \{(M, Não), (T, Alta)\};$
 $Des_p(\{3,6\}) = \{(M, Sim), (T, MuitoAlta)\};$
 $Des_p(\{4\}) = \{(M, Sim), (T, Normal)\}.$

Fonte: Pawlak(1982).

Quadro 18 - Conjuntos P-elementares

Considerando que o conjunto P_{Y_1} de pacientes que apresentam gripe e P_{Y_2} dos pacientes que não apresentam gripe o quadro 19 demonstra as definições de aproximações P-Superior P^Y_1 , P-Inferior P_{Y_1} , P-Fronteira $Frp(Y_1)$ e a precisão de aproximação $\alpha_p(Y_1)$, bem como P-Superior P^Y_2 , P-Inferior P_{Y_2} , P-Fronteira $Frp(Y_2)$ e a precisão de aproximação $\alpha_p(Y_2)$.

$$\begin{aligned}
P_{Y_1} &= \{1\} \cup \{3,6\} = \{1,3,6\}; \\
P^Y_1 &= \{1\} \cup \{2,5\} \cup \{3,6\} = \{1,2,3,5,6\}; \\
Frp(Y_1) &= P_{Y_1} - P^Y_1 = \{2,5\}; \\
\alpha(Y_1) &= \frac{\text{card}(P_{Y_1})}{\text{card}(P^Y_1)} = \frac{3}{5} = 0,6. \\
\\
P_{Y_2} &= \{4\}; \\
P^Y_2 &= \{4\} \cup \{2,5\} = \{2,4,5\}; \\
Frp(Y_2) &= P_{Y_2} - P^Y_2 = \{2,5\}; \\
\alpha(Y_2) &= \frac{\text{card}(P_{Y_2})}{\text{card}(P^Y_2)} = \frac{1}{3} = 0,3333.
\end{aligned}$$

Fonte: Pawlak(1982).

Quadro 19 – Definições para os conjuntos de pacientes que apresentam e não apresentam gripe

Para se obter a qualidade de aproximação da partição Y pelo conjunto de atributos P é demonstrado no quadro 20.

$$\gamma_P(Y) = \frac{\text{card}(P_{Y_1}) + \text{card}(P_{Y_2})}{\text{card}(U)} = \frac{3+1}{6} = 0,667$$

Fonte: Pawlak(1982).

Quadro 20 - Fórmula qualidade de aproximação

Primeiramente deve-se encontrar a qualidade de aproximação de todos os conjuntos P-elementares de atributos de condição, desta forma será possível descobrir as dependências e obter as reduções desejadas do sistema de informação em questão, quadro 21.

Atributos P	Qualidade de Aproximação $\gamma_P(Y)$	Conjuntos P-Elementares U / I_P
{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}

Fonte: Pawlak(1982).

Quadro 21 - Resultados da qualidade de aproximação dos conjuntos P-Elementares

Pode-se ainda verificar, na tabela 21, que as Y-reduções de $P = \{CMT\}$ são $\{M,T\}$ e $\{C,T\}$ e o Y-núcleo de P é $Core_Y(\{C, M, T\}) = \{C, T\} \cap \{M, T\} = \{T\}$. Isso quer dizer que T é o atributo mais significativo, o qual não pode ser desconsiderado, pois sua eliminação acarretaria na baixa de qualidade das aproximações. Em relação aos conjuntos $\{C,T\}$ e $\{M,T\}$ são mutuamente intercambiáveis, ou seja, os dois conjuntos produzem o mesmo resultado da

qualidade de informação em relação a $\{C,M,T\}$.

Da família $U/IC = \{\{1\},\{2,5\},\{3\},\{4\},\{6\}\}$ de classes de condição e da família $U/Ic = \{\{1,2,3,6\},\{2,5\}\}$ de classes de decisão surgem as regras descritas na tabela 3.

$X_i \in U/I_C$	$Y_j \in U/I_D$	Regra $Desc_C(X_i) \Rightarrow Desc_D(Y_j)$	Determinística?
{1}	{1,2,3,6}	$\{(C,Não), (M,Sim), (T,Alta)\} \Rightarrow \{(G,Sim)\}$	Sim
{2,5}	{1,2,3,6}	$\{(C,Sim), (M,Não), (T,Alta)\} \Rightarrow \{(G,Sim)\}$	Não
{3}	{1,2,3,6}	$\{(C,Sim), (M,Sim), (T,Muito Alta)\} \Rightarrow \{(G,Sim)\}$	Sim
{6}	{1,2,3,6}	$\{(C,Não), (M,Sim), (T,Muito Alta)\} \Rightarrow \{(G,Sim)\}$	Sim
{2,5}	{4,5}	$\{(C,Sim), (M,Não), (T,Alta)\} \Rightarrow \{(G,Não)\}$	Não
{4}	{4,5}	$\{(C,Não), (M,Sim), (T,Normal)\} \Rightarrow \{(G,Não)\}$	Sim

Fonte: Pawlak(1982).

Tabela 3 - Resultados de aproximações determinísticas $\{C,M,T\}$

Caso se adote a Y-Redução $\{M,T\}$ de C, estas regras simplificam as determinísticas, sem perda de qualidade de aproximação da partição. Neste caso obtém-se a situação demonstrada na tabela 4.

$X_i \in U/I_C$	$Y_j \in U/I_D$	Regra $Desc_C(X_i) \Rightarrow Desc_D(Y_j)$	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

Fonte: Pawlak(1982).

Tabela 4 - Resultados das aproximações determinísticas de $\{M,T\}$

1.10 IMPLEMENTAÇÃO

Neste capítulo são apresentadas informações teóricas sobre técnicas e ferramentas utilizadas para a implementação do sistema desenvolvido, tais como Borland Delphi, PL/Sql, Oracle e Enterprise Architect. Em seguida é apresentado um exemplo da operacionalidade da implementação.

1.10.1 Técnicas e ferramentas utilizadas

O ambiente de programação utilizado para o desenvolvimento do sistema é o Delphi. O Delphi é largamente utilizado para desenvolvimento de aplicações *desktop* e aplicações multicamadas (cliente/servidor), compatível com os principais bancos de dados que estão no mercado. O Delphi é uma ferramenta genérica comportando diversos tipos de projetos de desenvolvimento que vão de serviços a aplicações Web (MANZANO; MENDES, 2003).

Para o desenvolvimento das rotinas de cálculo da TCA foi utilizada a linguagem *Procedural Language / Structured Query Language* (PL/Sql), que é utilizada nos bancos de dados para a criação de functions, trigger, procedures.

Na figura 10 apresenta-se a rotina implementada utilizando PL/Sql, responsável por relacionar os atributos selecionados entre si.

```

40  end loop;
41  close c01;
42
43  /* Verifica o número de possibilidades que os atributos informados pelo usuário geram */
44  qt_possibilidade_w := power(2, y) - 1;
45
46  /* Realiza o relacionamento de todos os atributos utilizando a transformação binária
47  insere os conjuntos na tabela w_relacao_conj ela irá conter os relacionamentos dos atributos
48  da análise em questão.
49  */
50  while (qt_possibilidade_w > 0) loop
51  begin
52      bin_w := obter_decimal_binario(qt_possibilidade_w);
53
54      j := 0;
55      while (j < length(bin_w)) loop
56      begin
57          k := length(bin_w)-j-1;
58          if (nvl(substr(bin_w, j+1, 1),0) = 1) then
59              begin
60                  ds_conjunto_w := ds_conjunto_w || vetorTabela(k).cd_atributo_tca_w || ',';
61              end;
62          end if;
63          j := j + 1;
64      end;
65      end loop;
66      ds_conjunto_w := substr(ds_conjunto_w, 1, length(ds_conjunto_w) - 1);
67
68      select (nvl(max(nr_sequencia),0)+1)
69      into nr_sequencia_w
70      from w_relacao_conj;
71
72      insert into w_relacao_conj (ds_conjunto, nr_sequencia) values (ds_conjunto_w, nr_sequencia_w);
73      ds_conjunto_w := '';
74      qt_possibilidade_w := qt_possibilidade_w - 1;
75      end;
76  end loop;
77
78  end gerar_conjunto;

```

Figura 10 - Fragmento da rotina que realiza o relacionamento entre atributos

Após realizar o relacionamento entre os atributos é realizada a definição dos conjuntos P-elementares, a rotina de definição dos conjuntos P-elementares pode ser vista na figura 11 a seguir.

```

189         end if;
190     end if;
191
192     end;
193
194 end loop;
195 close c06;
196
197 delete from w_relacao_conj_item where ds_seq_conj = ds_seq_conj_w and nr_linha = nr_linha_w;
198 if (ds_linha_igual_w is null) then
199     delete from w_relacao_conj_item where ds_seq_conj = ds_seq_conj_w and nr_linha in (ds_linha_igual_w);
200 end if;
201 ds_linha_igual_w := '';
202
203 ds_conj_p_w := substr(ds_conj_p_w, 1, length(ds_conj_p_w)-1);
204
205 select nvl(instr(ds_lista_analisados_w, nr_linha_w),0)
206 into pos_linha_w
207 from dual;
208
209 if (ds_conj_p_w is not null) and
210 (pos_linha_w = 0) then
211     insert into w_conj_p (ds_conj_p) values (ds_conj_p_w);
212     ds_conj_PElementar_w := ds_conj_PElementar_w || '(' || ds_conj_p_w || ')';
213 end if;
214 ds_conj_p_w := '';
215 end;
216 end loop;
217 close c05;
218
219 if (ds_conj_PElementar_w is not null) then
220     ds_conj_PElementar_w := substr(ds_conj_PElementar_w, 1, length(ds_conj_PElementar_w)-1);
221
222     select (nvl(max(nr_sequencia),0)+1)
223     into nr_sequencia_w
224     from w_relacao_conj_pelementar;
225
226     insert into w_relacao_conj_pelementar (

```

Figura 11 – Fragmento da rotina que define os conjuntos P-elementares

Com isso, pode-se encontrar através da rotina implementada os cálculos das definições de aproximações P-Superior P^Y_1 , P-Inferior P_{Y_1} , P-Fronteira $Frp(Y_1)$ e a precisão de aproximação $\alpha_p(Y_1)$, bem como P-Superior P^Y_2 , P-Inferior P_{Y_2} , P-Fronteira $Frp(Y_2)$ e a precisão de aproximação $\alpha_p(Y_2)$, como mostra a figura 12.

```

352      /* Verifica as cardinalidades dos conjuntos SIM */
353      qt_card_pinf_sim_w := nvl(obter_cardinalidade_conjunto(ds_conj_pinferior_sim_w),0);
354      qt_card_psup_sim_w := nvl(obter_cardinalidade_conjunto(ds_conj_psuperior_sim_w),0);
355
356      /* Calcula a precisão de aproximação do SIM */
357      if (qt_card_psup_sim_w <> 0) then
358          vl_prec_aprox_sim_w := qt_card_pinf_sim_w / qt_card_psup_sim_w;
359      else
360          vl_prec_aprox_sim_w := 0;
361      end if;
362
363      /* Verifica as cardinalidades dos conjuntos NÃO */
364      qt_card_psup_ao_w := nvl(obter_cardinalidade_conjunto(ds_conj_psuperior_ao_w),0);
365      qt_card_pinf_ao_w := nvl(obter_cardinalidade_conjunto(ds_conj_pinferior_ao_w),0);
366
367      /* Calcula a precisão de aproximação do NÃO */
368      if (qt_card_psup_ao_w <> 0) then
369          vl_prec_aprox_ao_w := qt_card_pinf_ao_w / qt_card_psup_ao_w;
370      else
371          vl_prec_aprox_ao_w := 0;
372      end if;
373
374      /* Calcula a qualidade do atributo */
375      qt_card_universo_w := nvl(obter_cardinalidade_conjunto(ds_conj_decisao_sim_w) +
376                              obter_cardinalidade_conjunto(ds_conj_decisao_ao_w),0);
377      qt_card_PElementar_w := nvl(qt_card_pinf_ao_w, 0) + nvl(qt_card_pinf_sim_w, 0);
378
379      if (qt_card_universo_w <> 0) then
380          vl_qua_aprox_w := qt_card_PElementar_w / nvl(qt_card_universo_w,0);
381      else
382          vl_qua_aprox_w := 0;
383      end if;
384
385      insert into w_calculo_aprox (
386          ds_conj_pinferior_sim,
387          ds_conj_psuperior_sim,
388          ds_conj_pinferior_ao,
389          ds_conj_psuperior_ao,
390          vl_qua_aprox,
391          vl_prec_aprox_sim,
392          vl_prec_aprox_ao,
393          qt_card_pinf_sim,
394          qt_card_psup_sim,
395          qt_card_pinf_ao,
396          qt_card_psup_ao,
397          ds_seq_conj,
398          nr_seq_analise,
399          nr_seq_conj_pelementar)
400      values(
401          ds_conj_pinferior_sim_w,
402          ds_conj_psuperior_sim_w,
403          ds_conj_pinferior_ao_w,
404          ds_conj_psuperior_ao_w,
405          nvl(vl_qua_aprox_w, 0),
406          nvl(vl_prec_aprox_sim_w, 0),
407          nvl(vl_prec_aprox_ao_w, 0),
408          nvl(qt_card_pinf_sim_w, 0),
409          nvl(qt_card_psup_sim_w, 0),
410          nvl(qt_card_pinf_ao_w, 0),
411          nvl(qt_card_psup_ao_w, 0),
412          ds_conj_PElementar_w,
413          nr_seq_analise_p,
414          nr_seq_conj_pelementar_w);
415

```

Figura 12 – Fragmento da rotina que define os valores P-Superior, P-Inferior, Fronteira e qualidade de aproximação dos conjuntos P-elementares.

Por fim a rotina cria as regras reduzindo o universo de atributos, consistindo os atributos redundantes, porém sem perder a qualidade da informação dos atributos do universo completo, esse conjunto de regras será usado para buscar as informações na base de dados e demonstrá-las através de gráficos na tela, como mostra a figura 13.

```

43 ds_conj_item_w := substr(ds_conj_w, nr_pos_abre_par_w + 1, nr_pos_fecha_par_w - 2);
44 nr_pos_virgula_w := instr(ds_conj_item_w, ',');
45 end;
46 if (nr_pos_virgula_w > 0) then
47   begin
48
49     /* Se entrar aqui significa que ele tem mais de uma linha dentro do parenteses */
50     if (nr_pos_virgula_w > 0) then
51       begin
52         ds_conj_item_w := ds_conj_item_w;
53         while (ds_conj_item_w is not null) loop
54           begin
55             /* Verifica onde esta a virgula separando as linhas dentro deste grupo */
56             nr_pos_virgula_w := instr(ds_conj_item_w, ',');
57             if (nr_pos_virgula_w > 0) then
58               begin
59                 nr_linha_w := substr(ds_conj_item_w, 1, nr_pos_virgula_w-1);
60
61                 select nvl(instr(ds_conj_decisao_sim_w, nr_linha_w),0),
62                        nvl(instr(ds_conj_decisao_nao_w, nr_linha_w),0)
63                 into qt_scontem_w,
64                    qt_ncontem_w
65                 from dual;
66
67                 if (qt_scontem_w > 0) then
68                   begin
69                     insert into w_conjunto_regra
70                       (
71
72                     values (
73                         nr_sequencia,
74                         cd_tributo_tca,
75                         cd_item_tributo_tca,
76                         cd_decisao,
77                         nr_seq_analise,
78                         nr_item,
79                         nr_linha);
80                   end;

```

Figura 13 - Fragmento da rotina responsável por gerar as regras para a mineração na base de dados

A ferramenta utilizada para gerenciar a base de dados foi o Oracle, por ser um banco robusto e por obter inúmeras ferramentas de desenvolvimento que auxiliam o desenvolvimento do sistema.

1.10.2 Operacionalidade da implementação

O sistema desenvolvido serve para minerar as informações, acessando a base de dados de atendimento do ERP, utilizando a TCA listando os resultados obtidos em formas de relatórios e gráficos com o intuito de auxiliar no diagnóstico da doença.

O sistema é constituído por um único módulo, onde o usuário pode acessar o sistema através de uma das bases de dados configuradas em sua máquina, a base de dados do ERP, podendo ser configurada mais de uma base para o mesmo usuário.

Após acessar o sistema é possível realizar o cadastramento de outros usuários, cadastrar os atributos, que são os sintomas e os itens de atributos, que serão levados em conta para a realização do cálculo TCA, realizar o cálculo TCA, visualizar os históricos de resultados da TCA e visualizar o relatório de TCA.

A seguir é apresentado um exemplo de funcionamento da implementação, desde a entrada no sistema, os principais cadastros até o cálculo da TCA, os procedimentos são apresentados preservando a ordem de execução do sistema.

Ao iniciar o aplicativo será apresentada a tela de login, onde o usuário irá informar seu usuário, senha e base de dados, caso estiver tudo correto o sistema irá apresentar a tela principal onde o usuário poderá realizar as operações convenientes a ele, conforme figura 14 e figura 15, respectivamente.

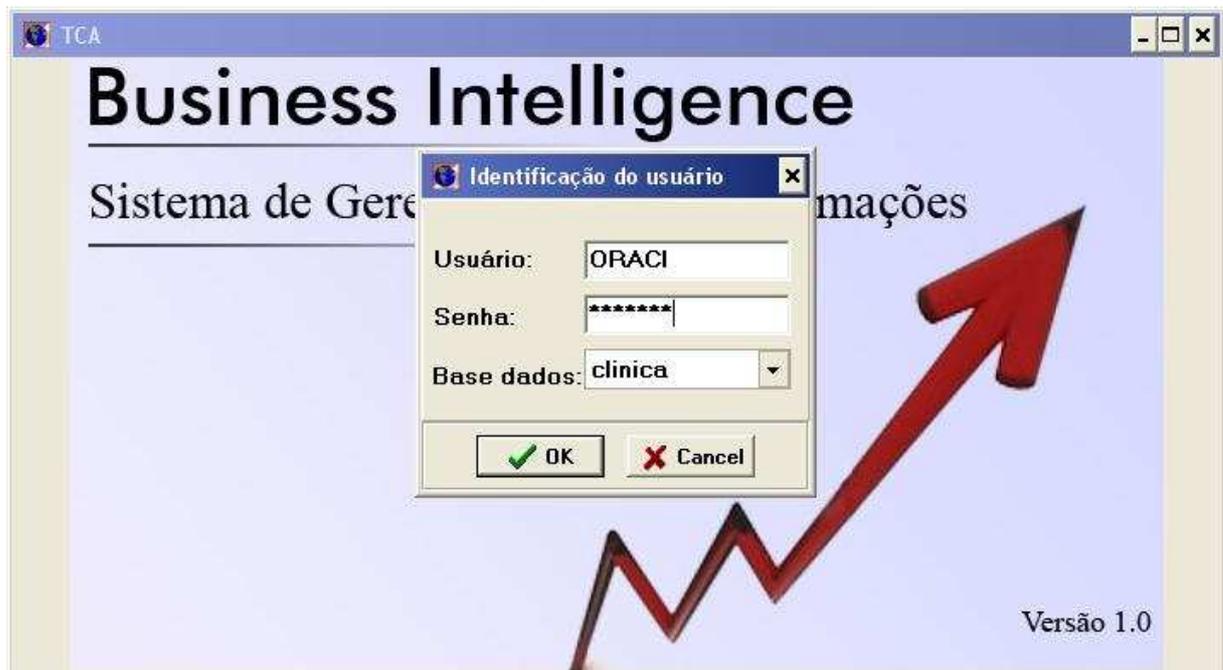
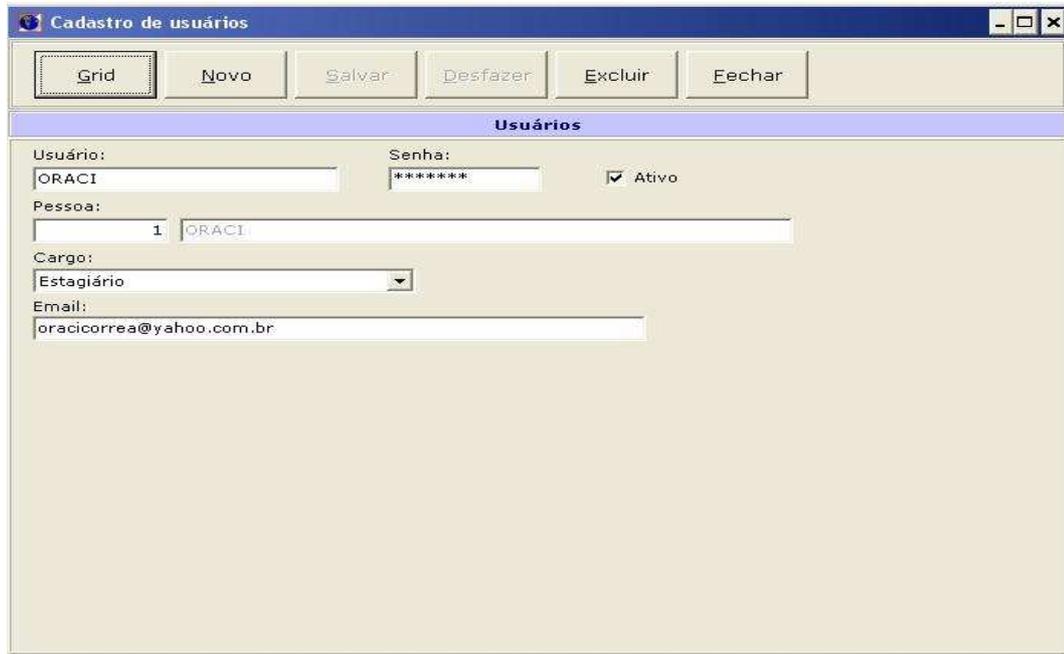
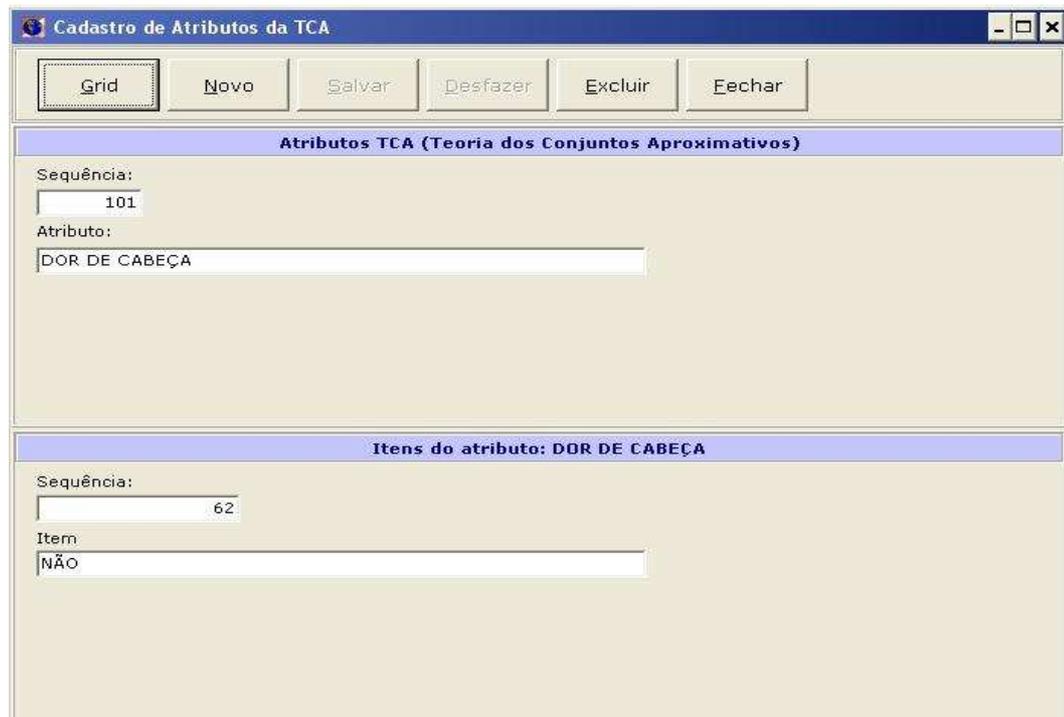


Figura 14 - Tela de login



Figura 15 - Tela principal do sistema

No menu cadastros pode-se acessar o cadastro de usuários e de atributos TCA, conforme demonstram a figura 16 e figura 17.

**Figura 16 - Cadastro de usuário****Figura 17 - Cadastro de atributo TCA**

Para a realização da demonstração da operacionalidade da aplicação do cálculo TCA será utilizado o mesmo exemplo citado na sessão anterior a qual demonstrava como o cálculo TCA é realizado, ou seja, deseja-se trazer os atendimentos dos pacientes que encontram-se com os sintomas da gripe, para isso o usuário informa uma série de parâmetros ao qual o sistema aplica o cálculo TCA, minerando a base do ERP para trazer os atendimentos que atendem as regras geradas pelo cálculo.

No menu Análise de dados existem as opções análise TCA, que é onde o sistema realiza o cálculo TCA, e o histórico de resultados TCA. Ao acessar a opção de análise TCA o sistema apresenta a tela inicial do processo TCA, conforme figura 18.

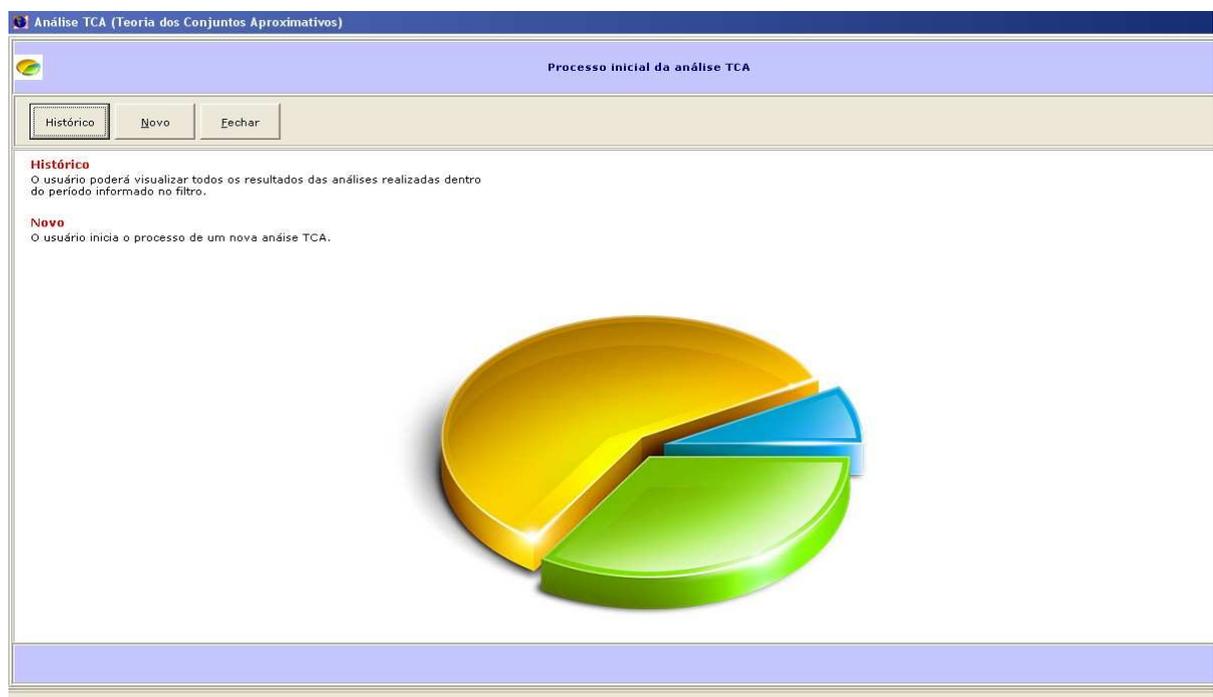


Figura 18 - Tela inicial análise TCA

Após pressionar o botão novo o sistema apresenta a tela de seleção de atributos, figura 19, onde é possível selecionar os atributos que irão compor os atributos de análise, basta que o usuário clique nos itens desejados para que os mesmos sejam selecionados, estes são os sintomas que serão levados em conta para a realização do cálculo TCA.

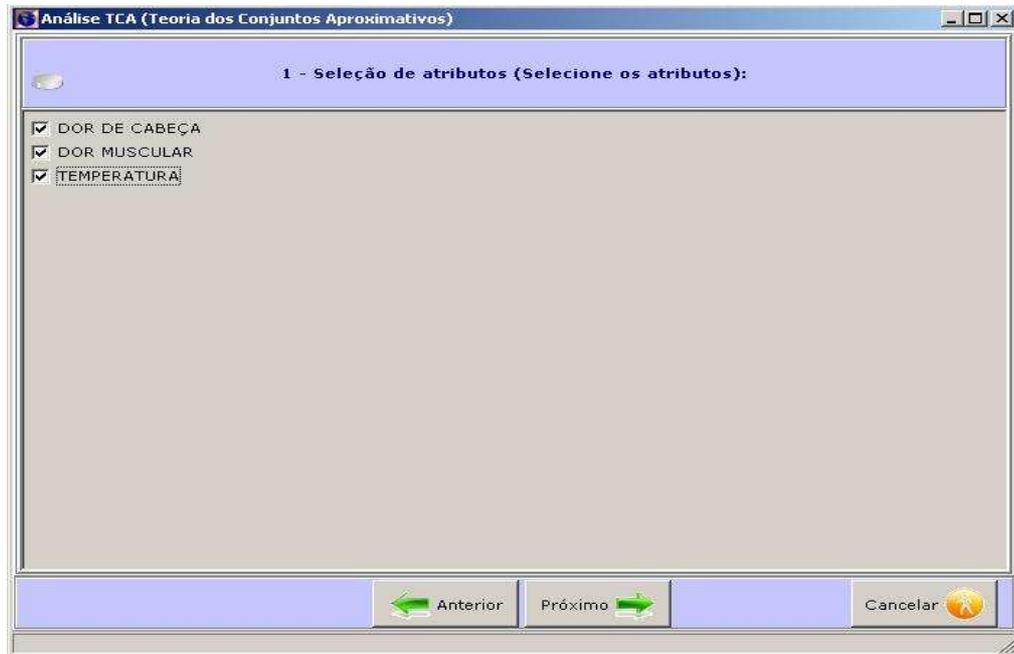


Figura 19 - Seleção de atributos de análise

Após o usuário selecionar os atributos desejados, no presente exemplo os atributos selecionados foram: dor de cabeça, dor muscular e temperatura. Em seguida deve-se pressionar o botão próximo para que o sistema mostre a tela de universo de objetos, ou seja, onde o usuário irá informar os valores desejados para os respectivos atributos, sintomas selecionados anteriormente, construindo o universo para que seja aplicado o cálculo TCA, conforme figura 20.

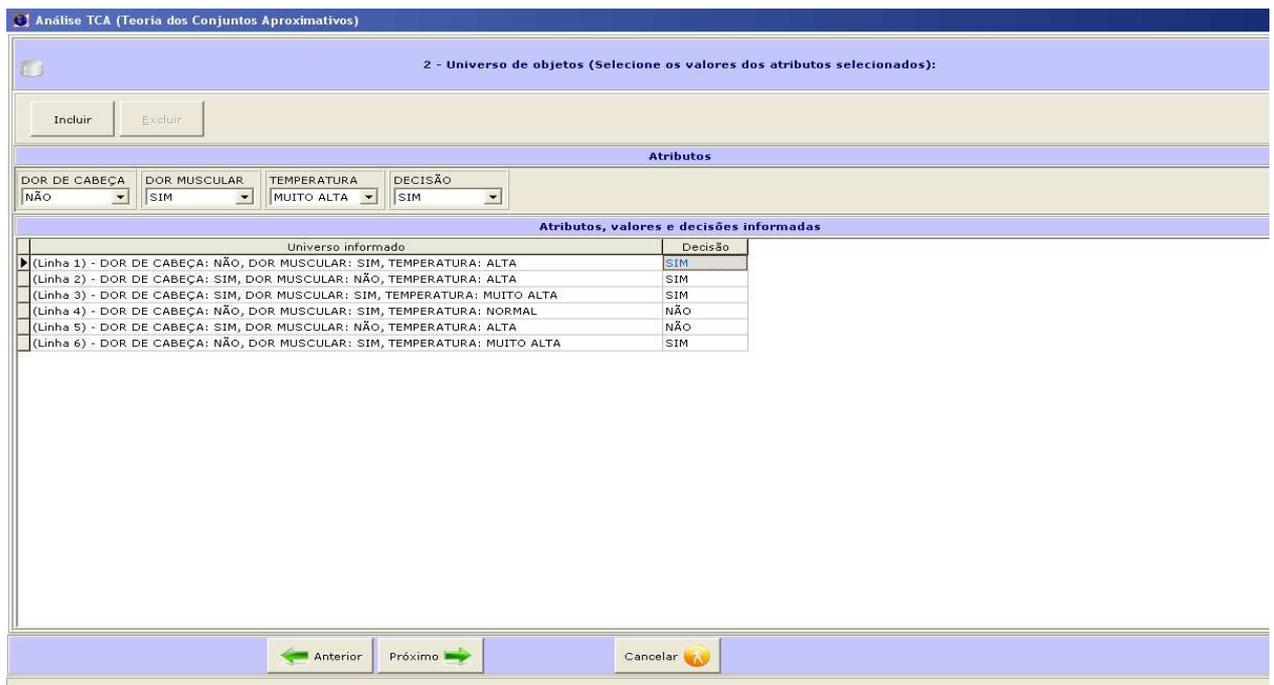


Figura 20 - Universo de objetos

Para que o sistema realize o cálculo TCA, após o usuário inserir o universo desejado, deve-se pressionar o botão próximo, a partir daí o sistema efetua todos os cálculos necessários para realizar a redução do universo informado, inicialmente são relacionados os atributos, após são formados os conjuntos P-elementares, baseados nos valores que o usuário informou.

Considerando o universo de atributos informado na figura 20, tomando como base os atributos Dor muscular (D) e Temperatura (T), serão obtidos os seguintes resultados, como demonstra o quadro 22.

- $U / IP = \{\{1\},\{2,5\},\{3,6\},\{4\}\}$
- $P^Y(\text{sim}) = \{1,3,6\}$
- $P_Y(\text{sim}) = \{1,2,3,5,6\}$
- Precisão de aproximação para sim = 0,67
- $P^Y(\text{não}) = \{4\}$
- $P_Y(\text{não}) = \{2,4,5\}$
- precisão de aproximação para não = 0,33
- qualidade de aproximação = 0,67

Quadro 22 - Resultados obtidos

Os resultados da qualidade de aproximação de todos os conjuntos P-Elementares obtidos através da operacionalidade para o exemplo citado são demonstrados no quadro 23.

Atributos P	Qualidade de Aproximação $\gamma_p(Y)$	Conjuntos P-Elementares U / I_p
{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}

Quadro 23 - Resultados obtidos no exemplo de operacionalidade

Após a realização do cálculo e da mineração de dados no ERP o sistema apresenta as informações do cálculo TCA. Apresentando as regras geradas pelo universo de parâmetros informado pelo usuário, neste exemplo, em específico as regras geradas foram: Temperatura = Muito Alta e Dor Muscular = Sim, Temperatura = Normal e Dor muscular = Sim e Temperatura Alta e Dor Muscular = Sim, apresenta as qualidades das informações calculadas, os números e a quantidade de atendimentos que foram encontrados na base ERP, considerando as regras citadas e o universo inicial que o usuário informou, como mostra a

figura 21.

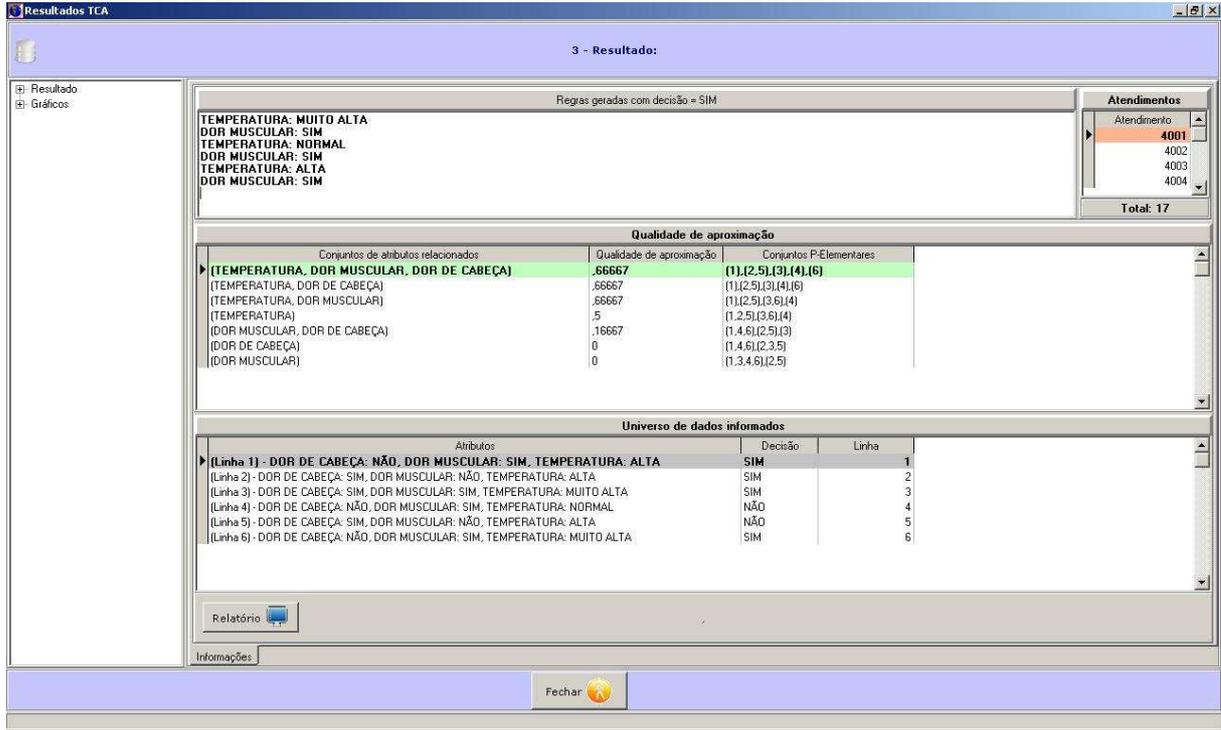


Figura 21 - Informações cálculo TCA

O usuário poderá acessar o cubo de decisão com os atendimentos trazidos do ERP, como mostra a figura 22, podendo cruzar as informações como desejar, pois o cubo permite arrastar as chamadas dimensões que aqui neste exemplo são: Tipo atendimento, Sintomas, Resultado, Atendimento e Pessoa.

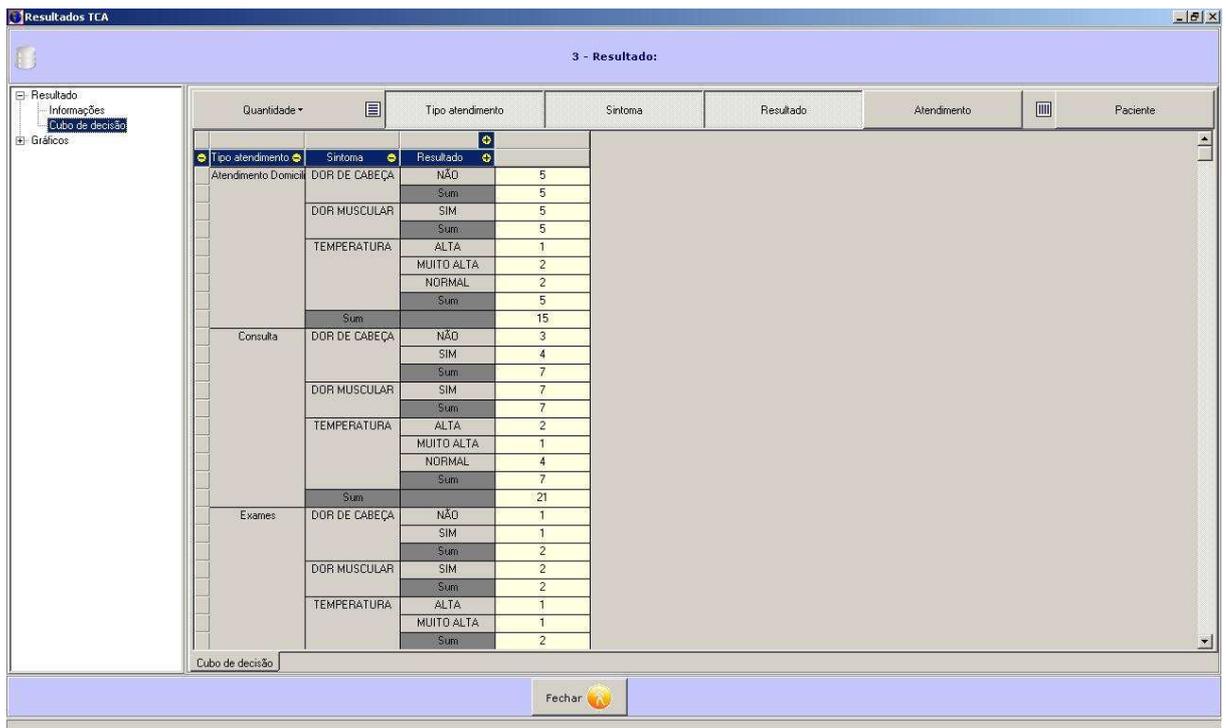


Figura 22 - Cubo de decisão

Na figura 23, pode ser visto fragmento da rotina que foi responsável em trazer os dados, demonstrados na figura 22, após realizar o cálculo TCA e gerar as regras a rotina realiza a busca das informações na base de dados do ERP.

```

procedure TBI_Analise_TCA.MineraDados;
var
    qDados : TQuery;
begin
    qDados.Active := False;
    qDados.SQL.Clear;
    qDados.SQL.Add(' select a.nr_atendimento, ');
    qDados.SQL.Add(' a.nr_prontuario, ');
    qDados.SQL.Add(' a.dt_nascimento, ');
    qDados.SQL.Add(' a.cd_pessoa_fisica, ');
    qDados.SQL.Add(' a.ie_tipo_atendimento, ');
    qDados.SQL.Add(' a.cd_convenio, ');
    qDados.SQL.Add(' a.nr_carteira, ');
    qDados.SQL.Add(' a.nr_carteira_tiss, ');
    qDados.SQL.Add(' a.dt_validade, ');
    qDados.SQL.Add(' a.cd_plano, ');
    qDados.SQL.Add(' a.nr_guia, ');
    qDados.SQL.Add(' a.nr_guia_principal_tiss, ');
    qDados.SQL.Add(' a.nr_autorizacao, ');
    qDados.SQL.Add(' a.dt_entrada ');
    qDados.SQL.Add(' from atendimento a ');
    qDados.SQL.Add(' sintomas b ');
    qDados.SQL.Add(' where a.nr_atendimento = b.nr_atendimento ');
    qDados.SQL.Add(' and a.cd_sintomas in (:listaregras) ');
    qDados.ParamByName(' listaregras').AsString := ListaRegras;
    qDados.Active := True;
end;

```

Figura 23 - Fragmento da rotina de mineração

Ainda pode ser demonstrada a tabela 5, a lista de regras que foram geradas e estão sendo passadas para a rotina de mineração.

Dor muscular (M)	Temperatura(T)	Gripe (G)
Sim	Alta	Sim
Sim	Muito Alta	Sim
Sim	Normal	Sim

Tabela 5 - Regras geradas

Após visualização do cubo de decisão o usuário poderá ver os atendimentos trazidos do ERP em forma de gráficos, os quais apresentam filtros de data. O gráfico a seguir mostra os atendimentos por tipos de atendimento, como mostra a figura 24.

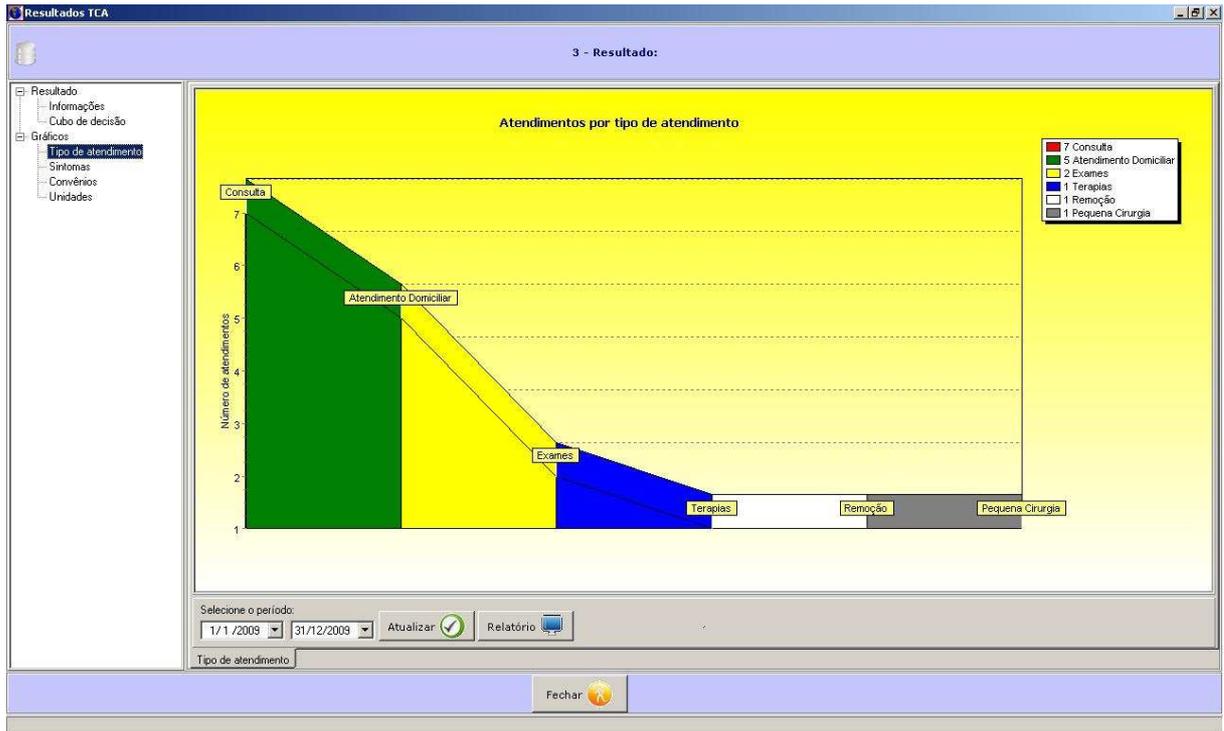


Figura 24 - Gráfico atendimentos por tipo de convênio

Neste outro gráfico, como demonstra a figura 25, o usuário pode ver a unidade de internação que o paciente se encontra.

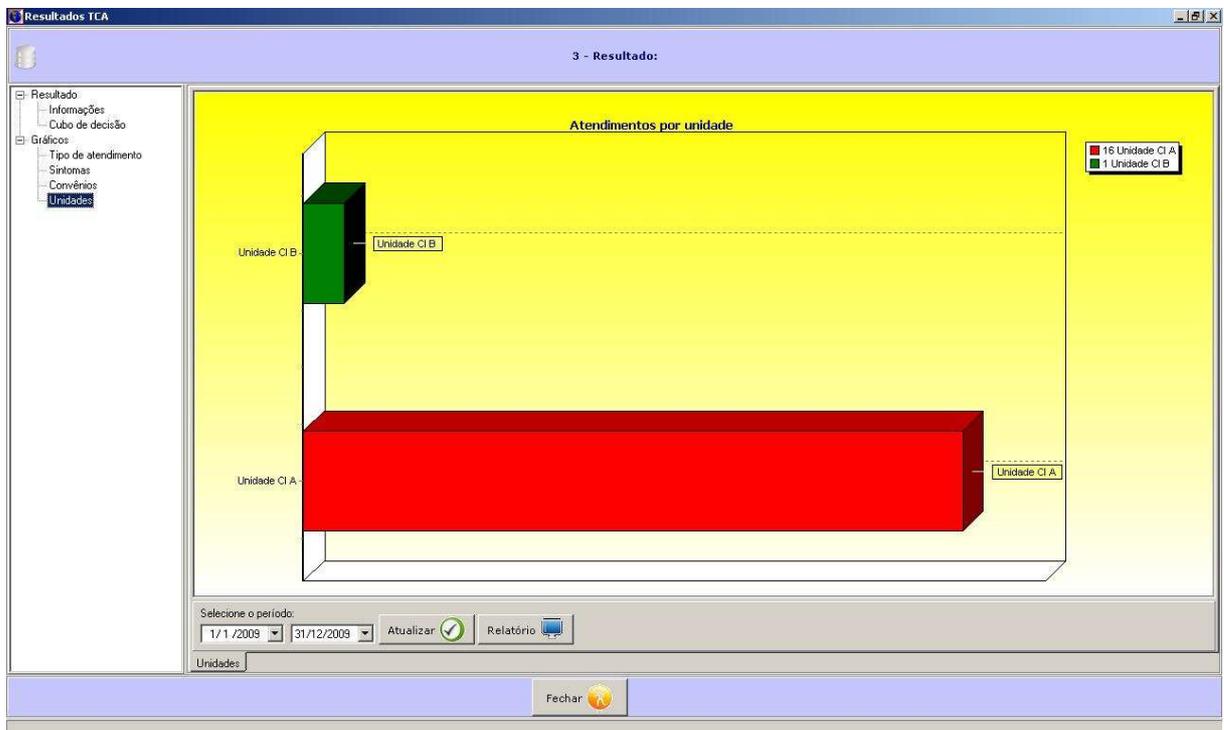


Figura 25 - Atendimentos por unidade internação

O usuário ainda pode verificar os dados através de relatórios, como atendimentos classificados por tipo de atendimento, como mostra a figura 26.

Print Preview

Relatório de atendimentos por tipo de atendimento

Data inicial: 1/1/2009 Data final: 31/12/2009

Tipo atendimento	Atendimento	Paciente	Prontuário	Data entrada	Data nascimento
Atendimento Domiciliar					
	4011	Pessoa_2011	2012	15/5/2009	11/12/1962
	4012	Pessoa_2012	2013	15/5/2009	1/5/1964
	4013	Pessoa_2013	2014	15/5/2009	3/11/1986
	4015	Pessoa_2015	2016	15/5/2009	1/1/2000
	4016	Pessoa_2016	2017	15/5/2009	15/3/2001
Consulta					
	4001	Pessoa_2001	2002	11/5/2009	5/6/1940
	4002	Pessoa_2002	2003	11/5/2009	29/6/2009
	4003	Pessoa_2003	2004	11/5/2009	5/7/1975
	4004	Pessoa_2004	2005	11/5/2009	16/8/1983
	4005	Pessoa_2005	2006	11/5/2009	13/11/1984
	4009	Pessoa_2009	2010	11/5/2009	15/7/1986
	4014	Pessoa_2014	2015	15/5/2009	15/7/1935
Exames					
	4010	Pessoa_2010	2011	11/5/2009	9/11/1960
	4032	Pessoa_2032	2033	15/5/2009	15/7/1986
Pequena Cirurgia					
	4007	Pessoa_2007	2008	11/5/2009	13/12/1965
Remoção					
	4006	Pessoa_2006	2007	11/5/2009	15/10/1962
Terapias					
	4008	Pessoa_2008	2009	11/5/2009	10/2/1968
Total:					17

Page 1 of 1

Figura 26 - Relatório de pacientes e sintomas

Neste outro relatório pode-se observar os dados referentes ao convênio de cada atendimento trazido na mineração da base de dados do ERP, como mostra a figura 27.

Print Preview

Relatório de atendimentos por convênio

Data inicial: 1/1/2009 Data final: 31/12/2009

Convênio	Atendimento	Paciente	Prontuário	Data entrada	Data nascimento	Plano	Nº carteira	Nº autorização
Unimed	4011	Pessoa_2011	2012	15/5/2009	11/12/1962	Master	118888	441132
	4012	Pessoa_2012	2013	15/5/2009	1/5/1964	Master	136748684	545132
	4013	Pessoa_2013	2014	15/5/2009	3/11/1986	Master	113164	45546
	4015	Pessoa_2015	2016	15/5/2009	1/1/2000	Master	131188	11132
	4016	Pessoa_2016	2017	15/5/2009	15/3/2001	Master	13113818	45454561
	4001	Pessoa_2001	2002	11/5/2009	5/6/1940	Standardt	89889898	1123132
	4002	Pessoa_2002	2003	11/5/2009	29/6/2009	Standardt		1112123
	4003	Pessoa_2003	2004	11/5/2009	5/7/1975	Standardt	11687898	545
	4004	Pessoa_2004	2005	11/5/2009	16/8/1983	Standardt	1564887	654645456
	4005	Pessoa_2005	2006	11/5/2009	13/11/1984	Standardt	113133989	888
	4009	Pessoa_2009	2010	11/5/2009	15/7/1986	Master	4898	46546546
	4014	Pessoa_2014	2015	15/5/2009	15/7/1935	Master	121336748	123123312
	4010	Pessoa_2010	2011	11/5/2009	9/11/1960	Master	446464687	11313
	4032	Pessoa_2032	2033	15/5/2009	15/7/1986	Enfermaria	221184	113135
	4007	Pessoa_2007	2008	11/5/2009	13/12/1965	Standardt	13168989	87878
	4006	Pessoa_2006	2007	11/5/2009	15/10/1962	Standardt	22134867	12132654
	4008	Pessoa_2008	2009	11/5/2009	10/2/1968	Standardt	13889889	112132
Total:						17		

Page 1 of 1

Figura 27 - Relatório de atendimentos por convênio

1.11 RESULTADOS E DISCUSSÃO

Nos trabalhos correlatos pode-se verificar que Zimmermann (2006) desenvolveu uma aplicação de Sistema de Apoio à Decisão utilizando BI, com a possibilidade de visualização das informações através do cubo de decisão, neste trabalhou também optou-se pelo uso do cubo de decisão, com isso o usuário pode verificar de uma forma mais organizada os atendimentos que foram trazidos na consulta das tabelas do sistema ERP.

A dissertação de Barreto (2003) teve como objetivo trazer a definição e a comparação da utilização do conjunto de ferramentas em Business Intelligence, neste trabalhou optou-se por uma técnica em específica DM, ou seja, realizar uma mineração na base ERP trazendo os atendimentos de acordo com as regras geradas pela técnica TCA.

No trabalho de Schmitt (2007), basicamente consistiu em aplicar a TCA em uma base de dados de recursos humanos, para a identificação de perfis dos profissionais cadastrados que mais se adequavam para comporem as equipes de projetos de software, neste trabalhou optou-se por aplicar a TCA no auxílio a diagnósticos de doenças. A qual foi possível observar que os resultados obtidos auxiliaram a tomada de decisão, pois a informação foi tratada, afim retirar os atributos repetidos e incertos, gerando ao final um conjunto das melhores regras para assim o sistema realizar a busca dos atendimentos dentro da base ERP.

Pode-se observar no exemplo da sessão anterior que após realizar o cálculo TCA, através das regras geradas: Temperatura = Muito Alta e Dor Muscular = Sim, Temperatura = Normal e Dor muscular = Sim e Temperatura Alta e Dor Muscular = Sim, o sistemas encontrou 17 atendimentos dentro da base ERP, com isso auxiliando o usuário na hora de diagnosticar o paciente, pois ele poderá verificar varias informações utilizando como base os atendimentos encontrados após a aplicação da técnica.

Partindo do principio que a informação é a principal arma nas tomadas de decisões, o que a torna um bem de muito valor dentro das instituições e com o intuito de proporcionar maior agilidade, confiabilidade e segurança, no que se refere à manipulação dessas informações, com o objetivo de prover e aumentar a qualidade da gestão do conhecimento decidiu-se por aplicar a TCA.

A TCA é uma técnica extremamente eficaz, pois trata os aspectos de redundância de dados reduzindo as variáveis envolvidas, facilitando o processo de mineração, através do relacionamento dos conjuntos de atributos informados pelo gestor. Dimensionando a cada conjunto o valor da qualidade de aproximação, o qual revela o quanto a variável é responsável

em gerar algum resultado.

No DM o sistema buscou implementar todas as etapas da geração do conhecimento, desde a seleção e processamento dos dados, transformação e mineração das informações, demonstrando-as por fim, através das regras geradas após a aplicação da TCA.

Através deste trabalho o gestor poderá verificar todos os atendimentos dos pacientes de acordo com os atributos de sintomas informados, o sistema irá ler as informações, irá aplicar o cálculo TCA, no qual irá gerar as regras de acordo com os relacionamentos de cada atributo e irá trazer os resultados através do cubo de decisão.

O uso de ferramentas como o Delphi, PL/Sql e banco de dados Oracle facilitou e motivou a realização da implementação, pois são tecnologias utilizadas em grande escala em nossa região.

Por fim, pode-se afirmar que todos os requisitos funcionais do sistema foram contemplados, atingindo o resultado final. Em relação aos requisitos não-funcionais todos foram atingidos também, devido a utilização das tecnologias mencionadas com o objetivo de aplicar a TCA na mineração de informações na área da saúde.

CONCLUSÕES

O principal objetivo deste trabalho foi de através do diagnóstico de doenças auxiliar a tomada de decisão dos gestores envolvidos, a partir de um universo de atributos conhecido pelo usuário, aplicando a TCA para que ao final sejam trazidos apenas os atributos que obtêm as melhores qualidades das informações. Desconsiderando os atributos com valores incertos e repetidos visando gerar as regras para trazer apenas os atendimentos dos pacientes, minerados através da base do ERP, apenas aqueles que podem ajudar no atendimento ao diagnóstico do paciente que esta sendo atendido. O sistema disponibiliza as informações através de um cubo de decisão, capaz de gerar gráficos para facilitar visualmente a análise das informações dos atendimentos que foram trazidos.

Todos os objetivos específicos deste trabalho foram realizados com sucesso, aplicação da técnica de mineração de dados utilizando a TCA, onde o gestor envolvido no processo informa os atributos de universo para que o cálculo seja aplicado, realizando todas as etapas dentro da TCA e dentro da mineração de dados.

O acesso dinamicamente as informações nas bases de dados hospitalares pode ser feita através da pesquisa do banco de dados, pois os resultados são guardados em tabelas especificamente desenvolvidas para este trabalho.

Quanto à apresentação das informações, após a realização da mineração o sistema demonstra as informações obtidas através de gráficos e relatórios. Com isso o usuário tem uma visão da informação mais eficaz e organizada auxiliando-o na tomada de decisão.

Este trabalho agrega a percepção de gerar informações coerentes em tempo hábil auxiliando nas tomadas de decisão das pessoas envolvidos. O uso da técnica TCA mostrou-se muito interessante, pois seus métodos de tratamentos de incertezas e de redundâncias acabam facilitando as gerações das informações em um menor tempo e com maior qualidade.

Uma das maiores dificuldades encontradas neste trabalho foi o entendimento do cálculo TCA, pois a teoria requer uma série de comparações, cada atributo deve ser comparado separadamente e em conjunto, pois com isso é possível tratar todas as incertezas e do universo de dados, requerendo certo cuidado para que a TCA seja calculada da maneira correta.

1.12 EXTENSÕES

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, como raciocínio baseado em casos, redes neurais ou árvores de decisão para que seja possível realizar um comparativo dos resultados obtidos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, Leandro M. et al. **Uma ferramenta para extração de padrões**. 2004. 13 f. Centro Universitário Luterano de Palmas, Palmas. disponível em: <<http://www.sbc.org.br/reic/edicoes/2003e4/cientificos/UmaFerramentaParaExtracaoDePadroes.pdf>>. Acesso em: 22 maio. 2009.
- BARBIERI, Carlos. **BI – Business Intelligence: modelagem & tecnologia**. Rio de Janeiro: Axel Books do Brasil, 2001.
- BARRETO, David Guaspari; **Business Intelligence: comparação de ferramentas**. 2003. 73 f. Dissertação (Mestrado em Computação). Curso de Pós-Graduação em Ciência da Computação - Universidade Federal do Rio Grande do Sul - Porto Alegre.
- BEZERRA, Eduardo. **Princípios de análise e projetos de sistemas com UML**. Rio de Janeiro: Campus, 2002.
- DALFOVO, Oscar; AMORIM, Sammy Newton. **Quem tem informação: é mais competitivo**. Blumenau: Santa Catarina Acadêmica Ltda, 2000.
- DOMINGOS, Marcos A.; UCHÔA, Joaquin Q. **Representação de Conhecimento Usando Teoria dos Conjuntos Aproximativos**. Lavras, 2006. Disponível em:< <http://www.dcc.ufla.br/infocomp/artigos/v2.1/art10.pdf>>, Acesso em: 10 março 2009.
- GHODDOSI, Nader M.A; DALFOVO, Oscar; MAIA, Luiz Fernando Jacinto. **Sistema de apoio à tomada de decisões, utilizando técnicas de data warehouse e tecnologia web**. Florianópolis. 2004. Disponível em: <<http://inf.unisul.br/~ines/workcomp/cd/pdfs/2806.pdf>>. Acesso em: 19 maio 2008.
- LEME FILHO, Trajano. **Business intelligence no microsoft Excel**. Rio de Janeiro: Axel Books do Brasil, 2004.
- MANZANO, José A. N. G; MENDES, Sandro S. V. **Estudo dirigido – Delphi 7**. Érica, 2003.
- PAWLAK, Zdzislaw. **Rough Sets. International Journal of Information & Computer Sciences**. [S.l.], 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.

RAMOS, Isabel; SANTOS, Maribel Yasmina. **Data mining no suporte à construção de conhecimento organizacional**. Porto. 2003. Disponível em: <http://repositorium.sdum.uminho.pt/bitstream/1822/2302/1/CAPSI2003_IMR_MYS.pdf>. Acesso em: 23 mar. 2009.

REZENDE, Denis Alcides; ABREU, Aline França de. **Tecnologia da informação: aplicada a sistemas de informação empresariais**. 3. ed. São Paulo: São Paulo Atlas S/A, 2003.

ROSINI, Alessandro M.; PALMISANO, Ângelo. **Administração de Sistemas de Informação e a Gestão do Conhecimento**. São Paulo: Thompson, 2003.

SERRA, Laércio. **A essência do Business Intelligence**. 1ª Edição, São Paulo: Editora Berkely Brasil, 2002.

SCHMITT, Sidnei; **Gestão do conhecimento: aplicação em Data Mining utilizando a Teoria dos Conjuntos Aproximados para Geração do Capital Intelectual**. 2007. 60f. Trabalho de Conclusão de Curso - Fundação Universidade Regional de Blumenau - Blumenau - SC.

STAIR, Ralph M.; REYNOLDS, George W. **Princípios de sistemas de informação: uma abordagem gerencial**. Tradução: Flávio Soares Corrêa da Silva. São Paulo: Pioneira Thomson Learning, 2006.

ZIMMERMANN, Thiago Rafael; **Desenvolvimento de um sistema de apoio à decisão baseado em business intelligence**. 2006. 76f. Trabalho de Conclusão de Curso – Fundação Universidade Regional de Blumenau – Blumenau – SC.

DICIONÁRIO DE DADOS

O dicionário de dados desenvolvido para especificar o sistema é apresentado nos quadros a seguir.

ATRIBUTO_TCA: permite armazenar os dados referentes aos atributos, sintomas serão a base para o cálculo TCA, dentro do ERP a tabela de atendimentos paciente contém ligação com esta tabela.

Campo	Descrição	Tipo	Tamanho
CD_ATRIBUTO_TCA	Código do atributo	number	10
DS_ATRIBUTO_TCA	Descrição do atributo	Varchar2	100
Chave Primária: CD_ATRIBUTO_TCA			

Quadro 24- Tabela ATRIBUTO_TCA

ITEM_ATRIBUTO_TCA: permite armazenar os dados referentes aos itens dos atributos, resultados dos sintomas que serão levados em conta para o cálculo TCA.

Campo	Descrição	Tipo	Tamanho
CD_ITEM_ATRIBUTO_TCA	Código do item de atributo	number	10
DS_ITEM_ATRIBUTO_TCA	Descrição do item de atributo	Varchar2	255
Chave Primária: CD_ITEM_ATRIBUTO_TCA			

Quadro 25 - Tabela ITEM_ATRIBUTO_TCA

CONTROLA_ATRIB_TCA: tabela originária da relação entre as tabelas APLICA_DADOS_TCA com a ATRIBUTO_TCA.

Campo	Descrição	Tipo	Tamanho
CD_ITEM_ATRIBUTO_TCA	Código do item de atributo	Number	10
DS_ITEM_ATRIBUTO_TCA	Descrição do item de atributo	Varchar2	255
Chave Primária: CD_ITEM_ATRIBUTO_TCA			

Quadro 26 - Tabela CONTROLA_ATRIB_TCA

CONTROLA_ITEM_ATRIB_TCA: tabela originária da relação entre as tabelas APLICA_DADOS_TCA com a ITEM_ATRIBUTO_TCA.

Campo	Descrição	Tipo	Tamanho
CD_ITEM_ATRIBUTO_TCA	Código do item de atributo	number	10
DS_ITEM_ATRIBUTO_TCA	Descrição do item de atributo	Varchar2	255
Chave Primária: CD_ITEM_ATRIBUTO_TCA			

Quadro 27 - Tabela CONTROLA_ITEM_ATRIB_TCA

ANALISE_TCA: permite armazenar os dados referentes às análises que foram realizadas.

Campo	Descrição	Tipo	Tamanho
NR_SEQUENCIA	Seqüência da análise TCA	Number	10
DT_ANALISE	Data da análise	Date	
NR_SEQ_USUARIO	Seqüência do usuário	Number	10
Chave Primária: NR_SEQUENCIA			
Chave Estrangeira: NR_SEQ_USUARIO (Referência tabela USUARIO)			

Quadro 28 - Tabela ANALISE_TCA

APLICA_DADOS_TCA: permite armazenar os dados referentes aos itens informados pelo usuário que serão levados em conta na análise TCA.			
Campo	Descrição	Tipo	Tamanho
NR_SEQUENCIA	Seqüência da aplica dados	Number	10
CD_ITEM_ATRIBUTO_TCA	Código do item de atributo	Number	10
CD_DECISAO	Código decisão	Number	10
NR_ITEM	Número do item	Number	10
NR_LINHA	Número da linha	Number	10
CD_ATRIBUTO_TCA	Código do atributo	Varchar2	255
NR_SEQ_ANALISE	Seqüência da análise	Number	10
Chave Primária: NR_SEQUENCIA			
Chave Estrangeira: NR_SEQ_ANALISE (Referência tabela ANALISE_TCA),			
Chave Estrangeira: CD_ATRIBUTO_TCA (Referência tabela ATRIBUTO_TCA)			
Chave Estrangeira: CD_ITEM_ATRIBUTO_TCA (Referência tabela ITEM_ATRIBUTO_TCA)			

Quadro 29 - Tabela APLICACAO_DADOS_TCA

RELACAO_CONJ_PELEMENTAR: permite armazenar os dados referentes aos conjuntos denominados P-Elementares que são levados em conta pela TCA.			
Campo	Descrição	Tipo	Tamanho
NR_SEQUENCIA	Seqüência do conjunto P-Elementar	Number	10
DS_CONJUNTO	Conjunto P-Elementar	Varchar2	255
DS_CONJUNTO_REL	Conjunto relacionado	Varchar2	255
NR_SEQ_APLICA_DADOS	Seqüência de aplica dados TCA	Number	10
Chave Primária: NR_SEQUENCIA			
Chave Estrangeira: NR_SEQ_APLICA_DADOS (Referência tabela APLICACAO_DADOS_TCA)			

Quadro 30 - Tabela RELACAO_CONJ_PELEMENTAR

TOTAL_APROX: permite armazenar os dados referentes aos resultados do cálculo TCA.			
Campo	Descrição	Tipo	Tamanho
NR_SEQUENCIA	Seqüência do total de aproximação	Number	10
DS_CONJ_PINFERIOR_SIM	Conjunto P-Inferior sim	Varchar2	255
DS_CONJ_PSUPERIOR_SIM	Conjunto P-Superior sim	Varchar2	255
DS_CONJ_PINFERIOR_NAO	Conjunto P-Inferior não	Varchar2	255
DS_CONJ_PSUPERIOR_NAO	Conjunto P-Superior não	Varchar2	255
VL_QUA_APROX	Valor qualidade aproximação	Number	12,5
VL_PREC_APROX_SIM	Valor da precisão de aproximação sim	Number	12,5
VL_PREC_APROX_NAO	Valor da precisão de aproximação não	Number	12,5
QT_CARD_PINF_SIM	Cardinalidade do conjunto P-Inferior sim	Number	10
QT_CARD_PSUP_SIM	Cardinalidade do conjunto P-Superior sim	Number	10
QT_CARD_PINF_NAO	Cardinalidade do conjunto P-Inferior não	Number	10
QT_CARD_PSUP_NAO	Cardinalidade do conjunto P-Superior não	Number	10
NR_SEQ_CONJ_PELEMENTAR	Seqüência do conjunto P-Elementar	Number	10
Chave Primária: NR_SEQUENCIA			
Chave Estrangeira: NR_SEQ_CONJ_PELEMENTAR (Referência tabela RELACAO_CONJ_PELEMENTAR)			

Quadro 31 - Tabela TOTAL_APROX