

Aquisição de Conhecimento Implícito de Indicadores de Qualidade de Vida

Acquiring Implicit Knowledge of Indicators of Quality of Life

Celso Bilynkievycz dos Santos

Universidade Tecnológica Federal do Paraná (UTFPR) - Ponta Grossa - PR - Brasil

Universidade Estadual de Ponta Grossa (UEPG) - Ponta Grossa - PR - Brasil

Centro Universitário Claretiano (CEUCLAR) - Batatais - SP - Brasil

bilynkievycz@globocom.com

Luciano Scandelari

Universidade Tecnológica Federal do Paraná (UTFPR) - Ponta Grossa - Brasil

scandelari@cits.br

Deborah Ribeiro Carvalho

Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES) - Curitiba - PR - Brasil

Universidade Tuiuti do Paraná (UTP) - Curitiba - PR - Brasil

deborah@ipardes.pr.gov.br

Maria Salete Marcon Gomes Vaz

Universidade Estadual de Ponta Grossa (UEPG) - Ponta Grossa - PR - Brasil

Universidade Federal do Paraná (UFPR) - Curitiba - PR - Brasil

salete@uepg.br

Morgana das Graças Procz dos Santos

Universidade Estadual de Ponta Grossa (UEPG) - Ponta Grossa - PR - Brasil

Centro Universitário Claretiano (CEUCLAR) - Batatais - SP - Brasil

morgana@uepg.br

Resumo

A Qualidade de Vida é um estado de bem-estar, que intercorrelaciona diversos aspectos do cotidiano e que pode sofrer interferências de fatores externos, muitas vezes difíceis de serem identificados, principalmente quando são estudados grandes grupos de trabalhadores com diferentes características. A identificação destes fatores, potencialmente implícitos, poderá auxiliar gestores de empresas em suas tomadas de decisões para melhoria da qualidade de vida de seus colaboradores e, conseqüentemente, da produtividade. Realizou-se um estudo de caso em uma instituição pública e, como instrumento de coleta dos dados, o WHOQOL-100. Para aquisição de conhecimento implícito, foram percorridas as etapas do Processo de Descoberta de Conhecimento em Base de Dados. Conclui-se que as técnicas de *Data Mining*, mais especificamente, as de Regras de Associação, podem ser utilizadas para complementar a análise dos resultados de um instrumento de pesquisa estruturada para *surveys*, principalmente quando se dispõe de dados adicionais e se deseja descobrir conhecimento implícito.

Palavras-chave: mineração de dados, WHOQOL-100, qualidade de vida, processo de descoberta de conhecimento em base de dados.

Abstract

Quality of Life is a well-being state which links several different aspects of everyday life and which also may be suffer interferences from outside factors, being those at times difficult to be realized, mainly when are studied carries large groups of workers with different characteristics. The identification of such factors, potentially implicit, might serve managers as for their decisions when it comes to improve the quality of life of their collaborators and consequently the productivity. A Study of case was carried in a public institution, using the WHOQOL-100 as a data collector. In order to obtain implicit knowledge, the Procedure of Knowledge Discovery in Databases phases was taken, It is concluded that the Data Mining techniques, more specifically, the Rules of Association ones, can be used as an addition when it comes to analysis results of an instrument of study based upon surveys, mainly when additional data is held and the implicit knowledge is the main focus to be figured.

Keywords: data mining, WHOQOL-100, quality of life, knowledge discovery in databases.

1. Introdução

A Qualidade de Vida (QV), definida pela Organização Mundial de Saúde(OMS), como “um estado de completo bem-estar físico, mental e social, que não consiste apenas na ausência de doença ou de enfermidade”, foi um tema de estudo ao longo dos anos, em diferentes momentos e perspectivas, principalmente pós-revolução industrial.

Atualmente essa temática, também é foco de estudo da Engenharia de Produção, principalmente quando se aborda a Qualidade de Vida no Trabalho (QVT), pois as empresas estão em constante busca por qualidade, modificando suas atitudes, relacionadas à vida de seus consumidores internos (colaboradores) e externos, e cada vez mais, preocupadas com suas responsabilidades sociais.

Este estado de bem-estar, denominado Qualidade de Vida é o resultado de uma intercorrelação entre diversas facetas do nosso cotidiano, que por sua vez podem sofrer influência de fatores externos.

A identificação desses fatores, principalmente dos que prejudicam a Qualidade de Vida pode auxiliar na elaboração de medidas que contribuam para a sua melhoria. Porém, na maioria das vezes, eles são difíceis de serem identificados, devido a sua complexidade e interdependência, principalmente quando se deseja monitorar grupos de trabalhadores com diferentes características.

Diante desta realidade é de fundamental importância aos gestores de qualidade da empresa, buscar por alternativas para otimizar a identificação dos fatores que poderiam estar prejudicando a QV e que poderiam passar despercebidos.

Então, definiu-se a seguinte pergunta a nortear a pesquisa: Como extrair conhecimento implícito de Qualidade de Vida?

Partindo das premissas de que a QV deve ser observada de forma global (FOX-RUSHBY; PARKER, 1995) e, de que o processo de tomada de decisão pode ser auxiliado por ferramentas de apoio à decisão (FAYYAD et al., 1996), busca-se neste trabalho, a utilização de uma ferramenta de avaliação de qualidade de vida global e de uma ferramenta de apoio à tomada de decisões.

Nesta perspectiva, procurou-se um instrumento que melhor avaliasse a qualidade de vida de forma global, adotando-se como objeto de estudo, o Instrumento de Avaliação de Qualidade de Vida - WHOQOL-100, considerado por Fleck (1999), um instrumento complexo, por contemplar a qualidade de vida nos seus diferentes aspectos.

Considerando-se o grande número de questões do instrumento e de resultados gerados pela sintaxe, mais a possibilidade de acréscimo de dados cadastrais e, considerando-se também, as limitações da holística humana, recorreu-se à Inteligência Artificial (IA) para complementar a análise.

Segundo Fayyad (1998), o ser humano não está preparado para interpretar grandes volumes de dados e/ou espaços multidimensionais. Nesta conjuntura, a Tecnologia da Informação surge como a ferramenta necessária para exercer esta função, mais especificamente a IA, com suas técnicas de *Data Mining*.

A função do *Data Mining* é garimpar informações ocultas, geralmente em grandes bases de dados, ampliando ao máximo as comparações entre os dados e tornando o conhecimento implícito em explícito (FAYYAD,1998).

A vantagem na adoção do *Data Mining* ao invés de técnicas tradicionais da Tecnologia da Informação, reside no fato das tradicionais exigirem que, inicialmente, hipóteses sejam estabelecidas, para que então, sejam construídas consultas a fim de comprovar ou não estas suposições, enquanto as técnicas de *Data Mining* têm maior independência do estabelecimento de hipóteses iniciais.

Outro fator motivador é a inexistência de trabalhos utilizando técnicas de *Data Mining* em dados coletados a partir do WHOQOL-100, pois os trabalhos existentes, geralmente se limitam à análise proposta pelo Grupo WHOQOL (1993), para o instrumento, o que caracteriza o presente trabalho como inovador.

Neste trabalho foi realizado um estudo de caso utilizando técnicas de *Data Mining* para complementar a análise dos resultados de uma ferramenta de coleta e análise de dados.

Escolheu-se para estudo de caso a Universidade Estadual de Ponta Grossa (UEPG), principalmente pela acessibilidade aos dados e por ser considerada uma das instituições que mais emprega na região dos Campos Gerais, com grande diversificação de mão-de-obra.

Este trabalho tem como objetivo extrair conhecimento a partir dos resultados do WHOQOL-100 e como objetivos específicos inovar a análise dos resultados WHOQOL-100, com a utilização de técnicas de *Data Mining* e avaliar o quão interessante é o conhecimento extraído.

2. Instrumento de Avaliação de Qualidade de Vida - WHOQOL-100

A procura de um instrumento de avaliação de qualidade de vida dentro de uma expectativa legitimamente universal fez com que a Organização Mundial da Saúde desenvolvesse um projeto colaborativo internacional. Este projeto resultou na elaboração do WHOQOL-100, um instrumento de QV composto por 100 itens (FOX-RUSHBY; PARKER, 1995).

O WHOQOL-100 é composto por 100 questões que analisa seis domínios: Físico, Psicológico, Nível de Independência, Relações Sociais, Meio Ambiente e Espiritualidade / Crenças Pessoais.

O instrumento proporciona uma avaliação minuciosa de 25 facetas, das quais uma corresponde às questões de QV geral, e as demais correspondem a 24 aspectos que estão distribuídas entre os 6 domínios (Quadro 1).

Quadro 1 - Aspectos e contemplados em cada domínio do WHOQOL-100

Domínios		Aspectos Contemplados	
I	Físico	01	Dor e desconforto
		02	Energia e fadiga
		03	Sono e repouso
II	Psicológico	04	Sentimentos Positivos
		05	Pensar, aprender, memória e concentração
		06	Auto-estima
		07	Imagem corporal e aparência
III	Nível de Independência	08	Sentimentos negativos
		09	Mobilidade
		10	Atividades da vida cotidiana

		11	Dependência de medicação ou de tratamentos
		12	Capacidade de trabalho
IV	Relações Sociais	13	Relações pessoais
		14	Suporte (apoio) social
		15	Atividade sexual
V	Meio Ambiente	16	Segurança física e proteção
		17	Ambiente no lar
		18	Recursos financeiros
		19	Cuidados de saúde e sociais: disponibilidade e qualidade
		20	Oportunidades de adquirir novas informações e habilidades
		21	Participação em, e oportunidades de recreação/ lazer
		22	Ambiente físico: poluição/ ruído/ trânsito/ clima
23	Transporte		
VI	Aspectos Espirituais/ Crenças pessoais	Religião/ 24	Espiritualidade / religião / crenças pessoais

Fonte: Adaptado de WHOQOL Group (1993)

Segundo definições de Vasconcelos (2002), o WHOQOL é classificado quanto ao tipo de entrevista e possibilidade de tratamento estatístico das respostas, como um instrumento estruturado para *surveys*.

3. Data Mining X KDD

O *Data Mining* compreende apenas uma das fases do processo KDD, referente à aplicação de algoritmos para extração de padrões em um processo de prospecção de conhecimento/padrões em bases de dados, envolvendo a automação da identificação e do reconhecimento de padrões (CARVALHO, 1999).

3.1 O Processo de KDD

O Processo de KDD surgiu em 1989, com o propósito de encontrar o conhecimento existente em uma base de dados e enfatizar o alto nível das aplicações dos métodos de prospecção de dados (AMARAL, 2001).

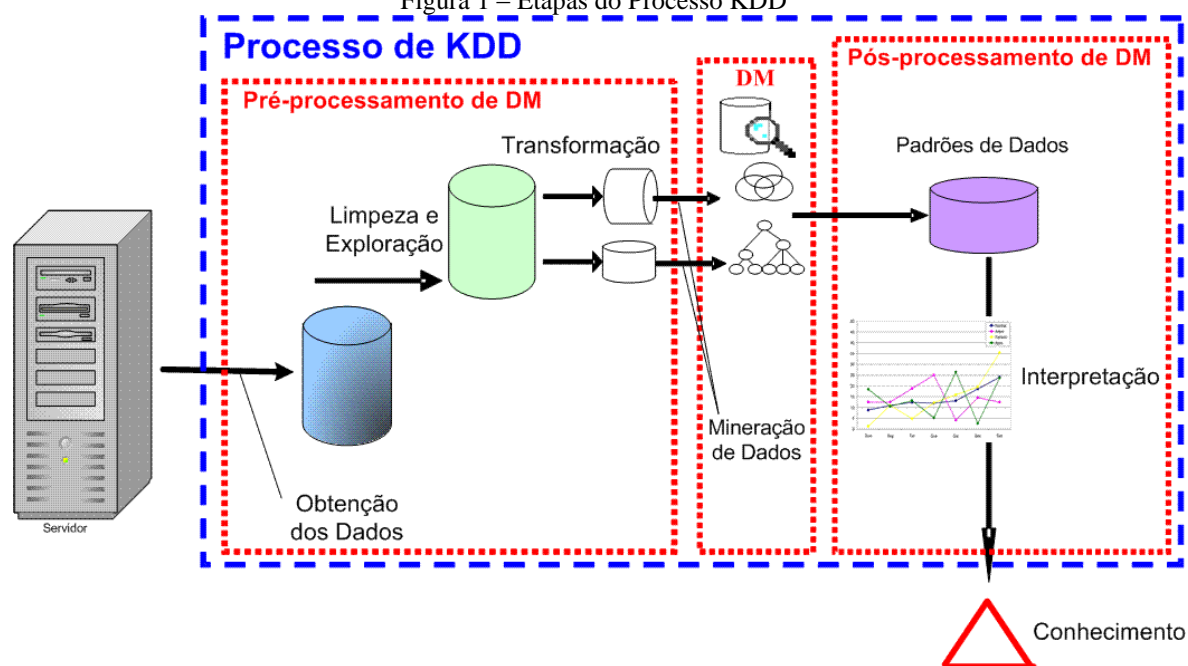
Em 1995, depois da realização de uma série de *workshops* (FAYYAD; UTHURUSAMY, 1994) sobre o assunto, foi realizada na cidade de Montreal no Canadá, a primeira Conferência Internacional de Prospecção de Dados (FAYYAD; UTHURUSAMY, 1995), quando se padronizou o termo “Descoberta de Conhecimento em Base de Dados” como KDD - *Knowledge Discovery in Databases*.

Esse processo tem por objetivo a extração do conhecimento implícito previamente desconhecido e a busca de informações potencialmente útil sobre os dados (FAYYAD, 1998).

3.1.1 Etapas do Processo KDD

Segundo Fayyad (1998) são cinco os procedimentos básicos para a implantação deste processo: seleção dos dados, pré-processamento (limpeza e exploração), transformação, *Data Mining*, interpretação dos resultados (Figura 1).

Figura 1 – Etapas do Processo KDD



Fonte: Santos (2007)

Em alto nível de abstração, as etapas de KDD fazem parte de três grandes grupos: pré-processamento, aplicação de um algoritmo de *Data Mining* e pós-processamento (MICHALSKI; KAUFMAN, 1998).

3.2 Data Mining

Data Mining é a etapa que efetivamente busca extrair informação válida, previamente desconhecida e de máxima abrangência geralmente de grandes bases de dados, usando-as para tomada de decisões, que permitem aos usuários explorar e inferir informação útil a partir dos dados, descobrindo relacionamentos escondidos no banco de dados (ADRIAANS; ZABTINGE, 1996).

O *Data Mining* é um área de estudo da Inteligência Artificial, que descende de três conceitos: da estatística, da inteligência artificial (IA) e da *Machine Learning (LM)* (FAYYAD et al., 1996).

3.2.1 Tarefas e Técnicas de *Data Mining*

As tarefas de KDD são dependentes do domínio da aplicação e do interesse do usuário. Cada tarefa de KDD extrai conhecimento do banco de dados em distintas formas de apresentação, podendo requerer um algoritmo diferente para a extração de conhecimento (LOPES, 1999). As tarefas em *Data Mining* têm como objetivo elementar, a predição e a descrição (FAYYAD et al., 1996).

A predição utiliza atributos para prognosticar o desconhecido ou os valores futuros de outras variáveis de interesse, enquanto a descrição contempla o que foi descoberto nos dados sob o ponto de vista da interpretação humana (CARVALHO, 1999).

Dentro desta perspectiva, segundo Viana (2004) as principais classes de tarefas de KDD, são: associação, classificação, *clustering*.

Alguns autores (RODRIGUES, 2000), apresentam um grande número de técnicas de *Data Mining*, considerando as tarefas. No entanto, cinco delas englobam as demais (CARVALHO, 1999). São elas: técnicas de redes neurais, regras de indução, árvores de decisão, análise de séries temporais e visualização (CARVALHO, 1999; CARVALHO, 2001).

O Quadro 2 apresenta as técnicas mais utilizadas na mineração de dados de acordo com as principais tarefas de KDD:

Quadro 2 - Problemas e respectivas Tarefas e principais Técnicas de *Data Mining* utilizadas para sua solução

Problema	Tarefa	Técnica	Exemplos de Algoritmos mais Utilizados
Predição	Regressão	Regressão	Cart, M5
	Classificação	Árvore de Decisão	J48
	Associação	Regras de Associação	Apriori, Dhp, Abs, Sampling
Descrição	Agrupamento	Particionamento	K-Means
		Hierarquia	Hac

Fonte: Adaptado de Viana (2004)

A escolha da técnica mais apropriada depende dos objetivos da pesquisa, do tipo de extração desejada e das características dos dados.

3.2.2 Tarefa de Associação

A tarefa de Associação (AGRAWAL; SRIKANT, 1994), tem como objetivo descobrir regras de associação (RA), que são expressões $X \rightarrow Y$, onde X e Y são conjuntos de itens, $X \cap Y = \emptyset$. Cada regra é um relacionamento SE (X) ENTÃO (Y), onde os conjuntos de itens X e Y frequentemente ocorrem juntos em uma mesma transação (registro), que indicam que na base de dados contém instâncias nas quais existe o conjunto X e também contem a presença de Y .

Um exemplo de uma regra do tipo $X \rightarrow Y$ poderia ser: 90% dos consumidores que compram carvão e carne (elemento X), também compram cerveja (elemento Y). O valor 90% é dito a confiança da regra, ou seja, representa o número de consumidores que compram cerveja, dividido pelo número de consumidores que compraram carvão e carne.

Além do valor de confiança, uma RA pode ser avaliada através de outras medidas de interesse que serão apresentadas na seção 4 deste trabalho.

4. Medidas de Interesse

As Medidas de Interesse são indicadores utilizados para avaliar a qualidade e/ou o grau de interesse de uma regra (padrão descoberto) e podem ser classificados em Subjetivos e Objetivos (MCGARRY, 2005).

Nas Medidas de Interesse Subjetivas, o grau de interesse da regra é avaliado pelo usuário, levando em consideração as suas expectativas e conhecimento prévio dos dados (SILBERSCHATZ; TUZHILIN, 1995). Enquanto, nas medidas de interesse objetivas, as regras são mensuradas a partir de formulações estatísticas/matemáticas.

4.1 Medidas de Interesse Subjetivas

As técnicas subjetivas de avaliação de uma regra, geralmente operam comparando as crenças de um usuário contra os padrões descobertos através de *Data Mining* (MCGARRY, 2005).

Através dessas medidas normalmente se determina, se um padrão é “útil” e/ou “inesperado”. No entanto, o padrão encontrado, poderá receber diferentes julgamentos, dependendo do usuário e dos seus objetivos (SILBERSCHATZ; TUZHILIN, 1995).

4.2 Medidas de Interesse Objetivas

As medidas de interesse objetivas mais conhecidas para avaliar uma RA são Suporte e a Confiança. A primeira corresponde ao índice de ocorrências de registros com os mesmos

antecedentes da RA na base de dados, e o segundo corresponde ao índice de registros com os mesmos antecedentes que atendem a RA.

A utilização apenas destes parâmetros de medidas, proporciona a geração de muitas regras, entre elas, regras óbvias, de baixo interesse, redundante ou contraditório, tornando a análise muitas vezes impraticável.

Para amenizar estes problemas, pesquisadores desenvolveram outras medidas estatísticas para avaliar as regras, entre eles, Tan (2002) apresenta em seu estudo aproximadamente vinte medidas de interesse.

Neste trabalho foram utilizadas cinco medidas, disponíveis no software WEKA, que serão descritas a seguir, são elas: Suporte, Confiança, *Lift*, *Rule Interest* (RI), Convicção.

4.2.1 Suporte

Existem várias, equações para calcular o suporte. Neste trabalho utilizou-se a fórmula proposta na implementada do algoritmo Apriori por Borgelt e Kruse (2002), que corresponde ao índice de frequência de um determinado item ou grupo de itens na base de dados, calculado através da Equação 1:

Equação 1 - Suporte

$$S_{(x)} = \frac{R_{(x)}}{Tr}$$

onde:

$S_{(x)}$ é o índice de suporte;

$R_{(A)}$ é o número de registros com o objeto x; e

Tr é o número total de registros na base de dados.

4.2.2 Confiança

Uma regra de associação é composta de dois elementos, um antecedente (condição) e um conseqüente (resposta). O antecedente deve implicar (\rightarrow) na ocorrência do conseqüente.

Dada a seguinte RA: $A \rightarrow B$, o seu valor de confiança é calculado a partir da Equação 2.

Equação 2 - Confiança

$$C_{(A \rightarrow B)} = \frac{S_{(A \cup B)}}{S_{(A)}}$$

onde:

$C_{(A \rightarrow B)}$ é o índice de confiabilidades da RA;

$S_{(A \cup B)}$ é suporte do antecedente e conseqüente, juntos; e

$S_{(A)}$ é o suporte do antecedente.

Esta medida representa, dentre as transações que contém o item “A”, o índice de vezes que as transações que também contém o item “B”

Os valores de confiança variam de 0 a 1.

4.2.3 Lift

O valor de *Lift*, utilizado para verificação do índice de dependência do conseqüente em relação ao seu antecedente, é calculado a partir da Equação 3:

$$\text{Equação 3 - Lift}$$
$$L_{(A \rightarrow B)} = \frac{C_{(A \rightarrow B)}}{S_{(B)}}$$

onde:

$L_{(A \rightarrow B)}$ é o índice de *Lift* da RA;

$C_{(A \rightarrow B)}$ é a confiabilidade da RA; e

$S_{(B)}$ é o suporte do conseqüente na base de dados.

O valores de *Lift* variam de 0 a ∞ .

O valor do índice igual a 1 indica independência (*Lift* = 1) entre os elementos da RA. Enquanto para os demais valores indicam, dependência positiva (*Lift* > 1) ou negativa (*Lift* < 1).

4.2.4 Rules Interest

O valor de RI, também conhecido por *Leverage*, nome do seu criador, também é utilizado para verificação do índice de dependência do conseqüente em relação ao seu antecedente, é calculado a partir da Equação 4:

Equação - Rules Interest ou Leverage

$$RI_{(A \rightarrow B)} = S_{(\text{real})} - S_{(\text{Esperado})}$$

onde:

$RI_{(A \rightarrow B)}$ é o índice de *lift* da RA;

$S_{(\text{real})}$ é o suporte rel, equivalente a $S_{(A \cup B)}$; e

$S_{(\text{esperado})}$ é o suporte esperado, equivalente a $S_{(A)} * S_{(B)}$.

Os valores dos índices de RI são bem menores que os de *Lift*, e variam de $-\infty$ a 0.

O valor do índice igual a zero indica independência (RI = 1) entre os elementos da RA. Enquanto para os demais valores indicam dependência positiva (RI > 1) ou negativa (RI < 1).

A diferença entre os dois índices: *Lift* e RI, está relacionada ao valor de suporte, o primeiro é mais preciso para identificar o nível de dependência entre elementos de RA de maior suporte, enquanto o segundo, para os de menor suporte.

Os dois índices apenas apresentam o grau de dependência entre os antecedentes e conseqüente, porém, não identificam o elemento implicado.

4.2.5 Convicção

A medida de *Convicção* foi introduzida por Brin et al. (1997) e é utilizada para indicar o grau de implicação do antecedente no conseqüente, que é calculado a partir da Equação 5.

Segundo Brin et al. (1997), as regras potencialmente implícitas (consideradas interessantes), possuem o índice de convicção entre 1.1 a 5.0.

Equação 5 - Convicção

$$Cv_{(A \rightarrow B)} = \frac{S_{(A)} * (1 - S_{(B)})}{S_{(A)} - S_{(A \cup B)}}$$

onde:

$Cv_{(A \rightarrow B)}$ é o índice de convicção da ra;

$S_{(A)}$ é o suporte do antecedente;

$S_{(\neg B)}$ é o suporte sem o conseqüente, equivale a $1 - S_{(B)}$; e

$S_{(A \cup B)}$ é suporte do antecedente sem o conseqüente, equivale a $S_{(A)} - S_{(A \cup B)}$

O valor do índice igual a zero indica independência ($Cv = 0$) entre os elementos da RA. Enquanto para os demais valores, indicam dependência positiva ($Cv > 0$) ou negativa ($Cv < 0$).

Para exemplificar, temos na Tabela 1, à esquerda, uma base de dados com 10 registros compostos de três itens cada e, à direita, a distribuição de cada item na base de dados.

Tabela 1 – Itens dos Registros em uma Base de Dados e sua Distribuição

N Registro	Itens			Distribuição dos itens na base de dados							
				♣	♠	♥	♦	●	■	○	
1	♣	●	♦	X			X	X			
2	♣	●	♠	X	X			X			
3	♣	●	♠	X	X			X			
4	♣	●	♠	X	X			X			
5	♣	♦	■	X			X			X	
6	♥	♦	♠		X	X	X				
7	♥	■	♠		X	X				X	
8	♥	■	♠		X	X				X	
9	♥	○	♦			X	X				X
10	♥	○	■			X				X	X
Total				5	6	5	4	4	4	4	2
Suporte				0.5	0.6	0.5	0.4	0.4	0.4	0.4	0.2

Fonte: Santos (2007)

O suporte de cada item do banco de dados foi calculado através da Equação 1.

Para exemplificar a aplicação dos índices, no processo de seleção objetiva das RA, temos as seguintes regras como suas respectivas unidades de interesse:

(1) ♣ → ● $S_{(♣)} = 0.4; S_{(●)} = 0.5; S_{(♣ \cup ●)} = 0.4; C = 1; L = 2; RI = 0.2; Cv = 0$

(2) ● → ♣ $S_{(●)} = 0.5; S_{(♣)} = 0.4; S_{(● \cup ♣)} = 0.4; C = 0.8; L = 2; RI = 0.2; Cv = 3$

(3) ♦ → ♠ $S_{(♦)} = 0.4; S_{(♠)} = 0.6; S_{(♦ \cup ♠)} = 0.1; C = 0.25; L = 0.42; RI = -0.14; Cv = 0.53$

$$(4) \spadesuit \rightarrow \diamond S_{(\spadesuit)}=0.4; S_{(\diamond)}=0.4; S_{(\spadesuit \cup \diamond)}=0.1; C=0.17; L=0.42; RI=-0.14; Cv=0.72)$$

$$(5) \diamond \rightarrow \heartsuit S_{(\diamond)}=0.4; S_{(\heartsuit)}=0.5; S_{(\diamond \cup \heartsuit)}=0.2; C=0.5; L=1; RI=0; Cv=1)$$

$$(6) \heartsuit \rightarrow \diamond S_{(\heartsuit)}=0.5; S_{(\diamond)}=0.4; S_{(\heartsuit \cup \diamond)}=0.2; C=0.4; L=1; RI=0; Cv=1)$$

Observando-se as RA 1 e 2, a partir das unidades de interesse de Lift e RI, conclui-se que existe dependência positiva entre o antecedente e conseqüente. E a partir do índice de Convicção, sabe-se que o item “●” depende do item “♣”, e o inverso não ocorre.

As RA 2 e 4 indicam que existe dependência negativa entre os termos “♦” e “♠”, que a medida que um deles ocorre na base de dados, diminui a possibilidade da ocorrência do outro e; que existe a possibilidade do elemento “♦” □ □ não ocorrer em função da ocorrência de “♠” de 0.53 vezes (RA 2), enquanto, a possibilidade do elemento “♠” □ □ não ocorrer em função da ocorrência de “♦” de 0.72 vezes (RA 4).

Como exemplos de RA que indicam independência entre os itens, temos as RA 5 e 6. Estas indicam que os elementos “♦” e “♥”, apesar de aparecerem juntas em 20% das transações, não dependem um do outro para ocorrerem.

5. Metodologia

Segundo as definições de Vasconcelos (2002) é classificada: quanto ao objeto como Temática ou Focal simples; quanto ao tipo do objeto e as fontes a serem investigadas classifica-se como Meta-análise; quanto à natureza dos dados e da análise é classificada como Mista; quanto aos objetivos, aplicações e tipo de conhecimento implicado é considerada Aplicada, pois articula dentro de diferentes ciências (Engenharia de Produção, Saúde, Ciências da Computação).

Aplicou-se o WHOQOL-100 através de amostragem probabilística por conglomerado, atingindo grupos de indivíduos de diversos níveis e segmentos, totalizando 267 agentes universitários (31,84% da população), equivalentes a 45,95% dos funcionários lotados no Campus da Universidade.

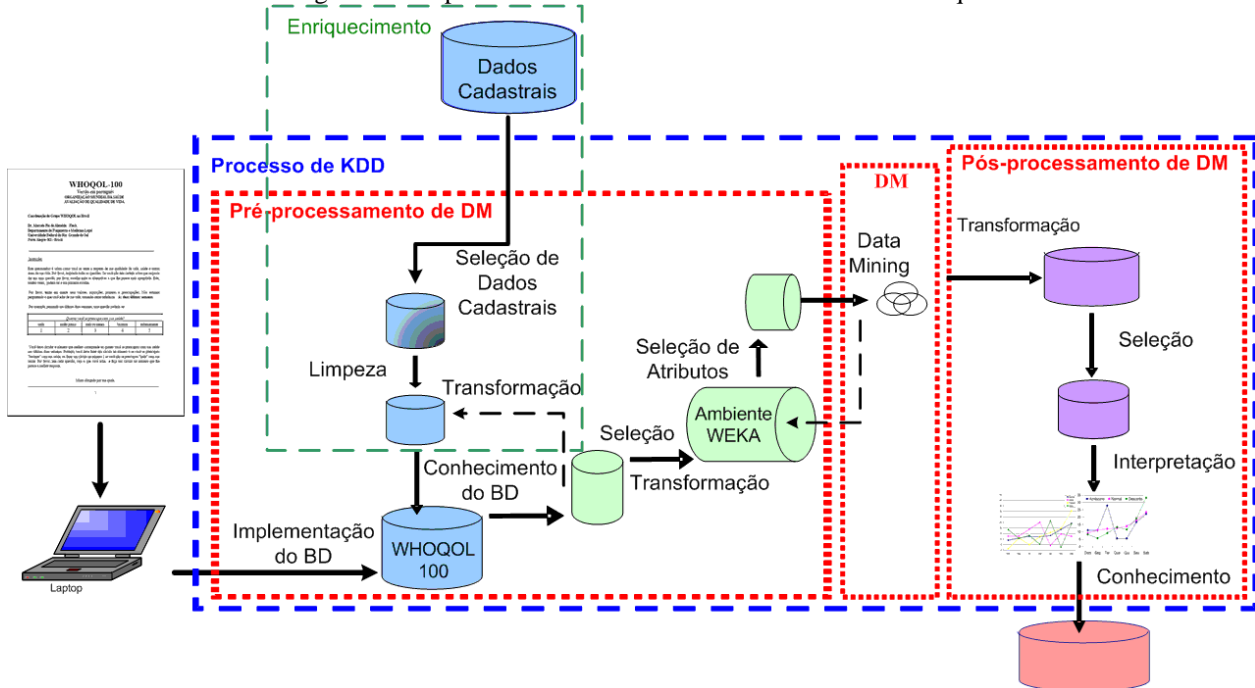
A análise de dados foi realizada em dois momentos. No primeiro foi utilizado o protocolo tradicional (WHOQOL Group, 1993), porém, sem a utilização do *software* SPSS. E no segundo momento, utilizou-se o Processo de KDD, como um protocolo complementar, destinado à aquisição de conhecimento implícito, através de técnicas de Data Mining de Regras de Associação.

6. Aquisição de Conhecimento Implícito de Indicadores de Qualidade de Vida

Nesta proposta de pesquisa, técnicas de *Data Mining* de Regras de Associação formam utilizadas para encontrar conhecimento implícito no relacionamento entre os resultados do WHOQOL-100 e informações cadastrais da amostra.

Para utilização das técnicas de *Data Mining* foram percorridas as etapas do processo de KDD (Figura 2).

Figura 2 – Etapas do Processo de KDD Desenvolvidas na Pesquisa



Fonte: Santos (2007)

O processo de KDD, segundo Michalski e Kaufman (1998), é composto basicamente de três macro-etapas, que serão descritas a seguir, são elas: Pré-processamento de *Data Mining*, *Data Mining* e Pós-processamento de *Data Mining*.

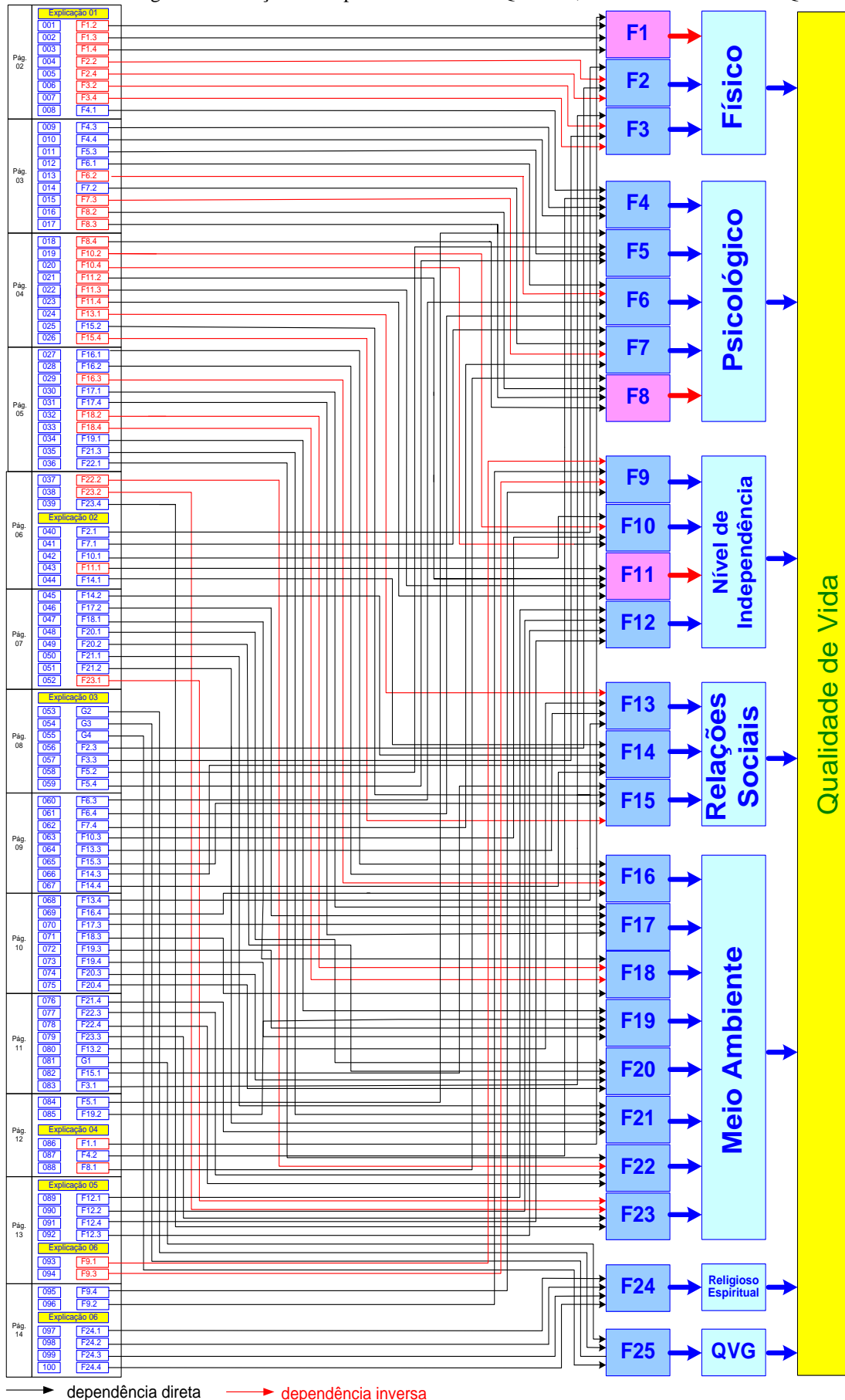
6.1 Pré-Processamento para *Data Mining*

Esta etapa foi realizada através de cinco sub-etapas. Iniciou-se com a implementação da base de dados e foi finalizada com a transformação dos dados.

6.1.1 Implementação da Base de Dados

A base de dados foi implementada a partir do entendimento da sintaxe de análise de dados (GRUPO WHOQOL, 1995) e da relação de dependência entre os resultados e os dados de entrada (respostas na escala Likert) (Figura 3).

Figura 3 – Relação de Dependência entre as Questões, Facetas e Domínios de QV

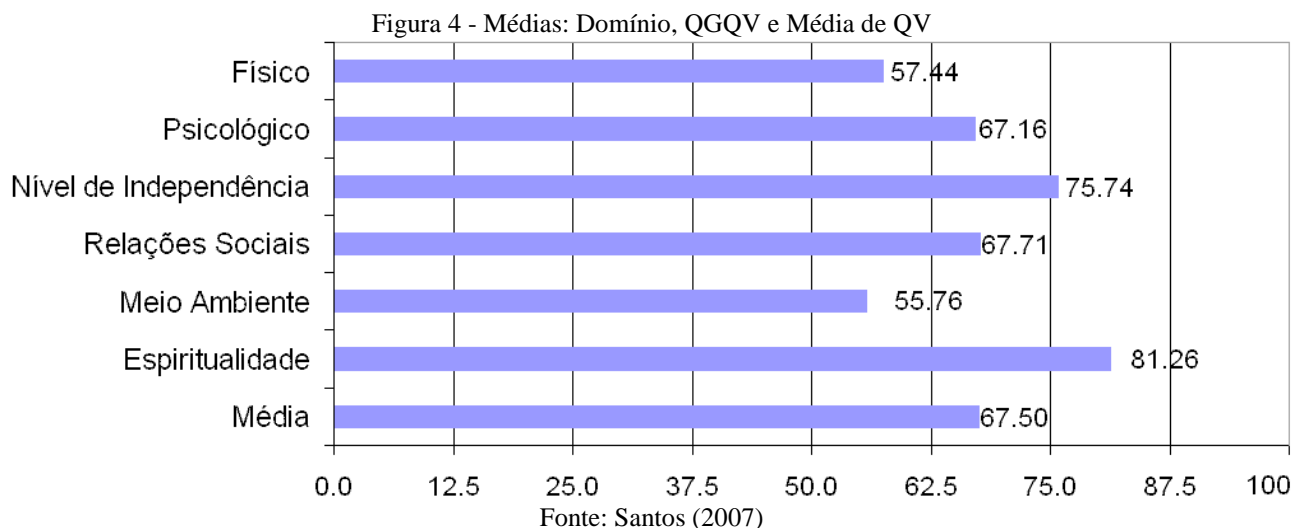


Fonte: Santos (2007)

A Figura 3 representa a relação de dependência entre domínios, facetas e as respostas do questionário.

6.1.2 Conhecimento da Base de Dados

Após o desenvolvimento da sintaxe do WHOQOL-100, através dos softwares MS Access e MS Excel, chegou-se aos seguintes resultados dos indicadores de QV.



O Domínio Meio Ambiente apresentou os piores indicadores de QV, enquanto o Domínio Espiritualidade apresentou os melhores escores de QV.

Foi convencionada, uma média de QV, a partir da média aritmética entre os domínios e QGQV, que resultou em 67,5% da possibilidade máxima da escala de QV.

A Figura 5 apresenta a média dos aspectos de QV.

Figura 5 - Aspectos de Qualidade de Vida



Fonte: Santos (2007)

O conhecimento da base de dados permitiu identificar os piores domínios e aspectos de QV, que foram definidos com atributos metas¹ para investigação através de RA, são eles: Domínio Meio Ambiente; Aspecto Dor e Desconforto e Cuidados de Saúde e Sociais (pior aspecto do pior domínio).

6.1.3 Enriquecimento da Base de Dados

O enriquecimento da base de dados foi realizado com a inclusão de 32 atributos cadastrais da amostra e de 106 atributos disponíveis em três atos públicos (UEPG, 2006).

6.1.3.1 Limpeza dos Dados

Esta etapa, apesar de apresentada nesta seqüência, iniciou-se durante o conhecimento da base de dados enriquecida, onde foram realizadas SQL específicas que resultaram na identificação de ruídos e registros duplicados, que em seguida foram eliminados ou substituídos por dados apropriados.

6.1.3.2 Conhecimento Base de Dados Enriquecida

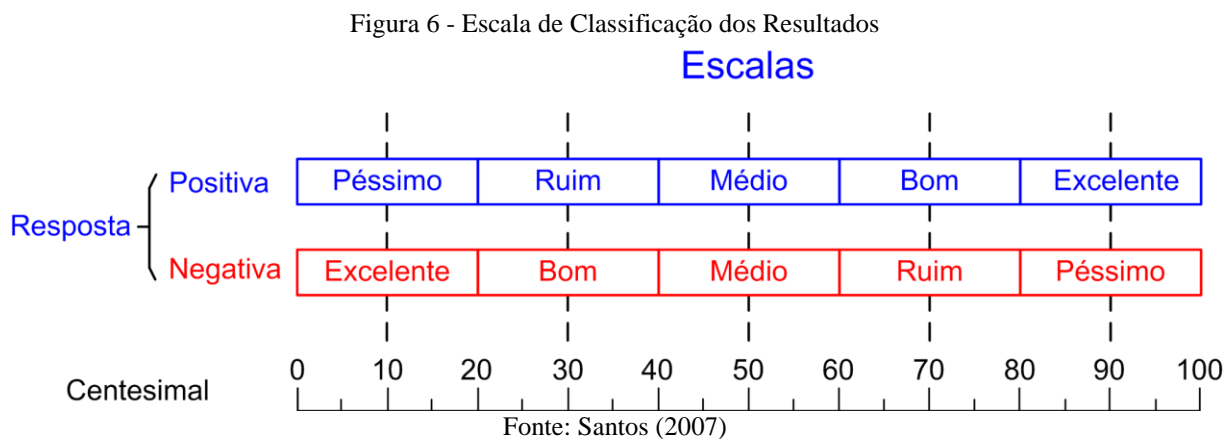
Durante a análise dos dados, através das ferramentas de pré-processamento do WEKA, observou-se que muitos atributos do tipo lógico, como por exemplo, os atributos correspondentes às competências dos servidores, evidenciavam baixos ou nenhum ganho de informação, devido à baixa ou inexistente freqüência de algumas classes. Então, os atributos lógicos, com as classes de menor freqüência, inferiores ao valor mínimo do suporte adotado foram eliminados (retorno à etapa anterior), pois naturalmente, estas classes, não apareceriam nas RA, enquanto as classes contrárias estariam presentes em grande parte das regras. Este procedimento resultou na diminuição de 17 atributos, o que diminuiria consideravelmente o número de RA com baixo grau de interesse (com conhecimento óbvio).

Os demais atributos com baixo ganho de informação foram mantidos, ficando em função dos algoritmos de seleção a sua eliminação. Pois, observou-se que a supressão de alguns desses atributos interferia na seleção de determinados grupos de atributos, devido ao fato de que, o algoritmo avaliador utilizado na seleção, busca preferencialmente, intercorrelações envolvendo classes de baixa frequência.

6.1.3.3 Transformação dos Dados Enriquecidos

Para utilização da técnica de Regras de Associação, o algoritmo minerador Apriori, exige que os dados de entrada estejam categorizados. Então, todos os atributos do tipo numérico foram substituídos por duas novas tipagens, lógica e categórica, totalizando 128 atributos cadastrais.

Segundo o Grupo WHOQOL (1995) não existe uma normalização padrão para os resultados do WHOQOL-100, e que havendo essa necessidade, sugerem que seja realizada através do seu respectivo grupo focal. Porém, tendo em vista a necessidade de categorização dos resultados e, conhecendo-se os valores máximo e mínimo da escala centesimal, adotou-se uma escala de Likert (Figura 6) para classificação dos resultados.



Através de SQL, os resultados foram transformados em variáveis ordinárias de avaliação de QV, considerando as características dos Aspectos ou Domínio e os intervalos da Figura 6.

6.1.4 Seleção e Transformação dos Dados

A seleção e transformação dos dados foram realizadas no ambiente WEKA.

Inicialmente configurou-se um *driver* ODBC (*Open Data Base Connectivity*) para conectar o software WEKA ao MS Access. Através desta integração foi possível executar SQL no WEKA, fazendo a transformação dos dados para o formato arff, exigido pelo ambiente.

Foram selecionados através de SQL 160 atributos: 32 indicadores de QV e 128 dados cadastrais.

6.1.4.1 Seleção de Atributos Relevantes

Nesta sub-etapa foram selecionados os atributos relevantes, para evitar a geração na etapa subsequente de regras óbvias ou casuais (resultante de correlações espúrias).

Para a seleção desses atributos foi utilizado o filtro *AttributeSelectio*, composto pelo avaliador de atributos *CfsSubsetEval* (*Correlation-based Feature Subset Seleccion*), através do método de busca *Best Fisrt*.

Esse algoritmo avalia o valor de um subconjunto de atributos considerando a capacidade preditiva individual de cada atributo, juntamente com o grau de redundância entre eles (HALL, 1998).

A seleção dos dados foi realizada de forma acíclica, buscando-se relacionamentos entre cada um dos indicadores de QV (Facetas, Domínios e MQV) e os dados cadastrais, gerando 32 grupos de dados.

Cada grupo de dados foi selecionado de forma supervisionada. Separou-se, inicialmente, apenas um indicador de QV e todos dados cadastrais (111 atributos) em um único arquivo. Cada indicador de QV foi configurado como atributo meta¹ (atributo de referência), durante a aplicação do filtro para seleção dos atributos relevantes em cada grupo de dados.

Depois de realizados os procedimentos acima mencionados, foram encontrados 14 atributos classificados como previsores dos aspectos e domínios de QV da amostra, apresentados na Tabela 2.

Tabela 2 – Percentual dos Atributos Metas atendidos pelos Atributos Previsores

Item	Atributo Previsor	% Atributos Metas	
		Domínios	Facetas
1	Lotação	100	100
2	Bairro	83.33	96
3	Últimas Férias	50	48
4	CSN	33.33	40
5	Alfabetizado	16.67	0
6	Competência: Agilidade	16.67	0
7	Competência: Atenção	16.67	0
8	Competência: Condicionamento Físico	16.67	0
9	Competência: Trabalho em Equipe	16.67	8
10	Função Anterior	16.67	48
11	Sanção Disciplinar	16.67	0
12	Competência: Asseio Pessoal	0	4
13	Disfunção Temporária	0	4
14	Partições em Atividades Administrativas	0	4

Fonte: Santos (2007)

O atributo lotação foi selecionado em todos lotes de dados como previsor dos diferentes aspectos e domínios de QV.

O atributo bairro foi classificado previsor de 83.33% dos Domínios e 96% dos aspectos de QV. Não é considerado atributo previsor apenas do Domínio Físico e do Aspecto: Participação em oportunidades de recreação/ lazer.

O salário representado pelo atributo “CSN” foi considerado atributo previsor de dois domínios: Psicológico e Meio Ambiente e, 40% dos aspectos de QV.

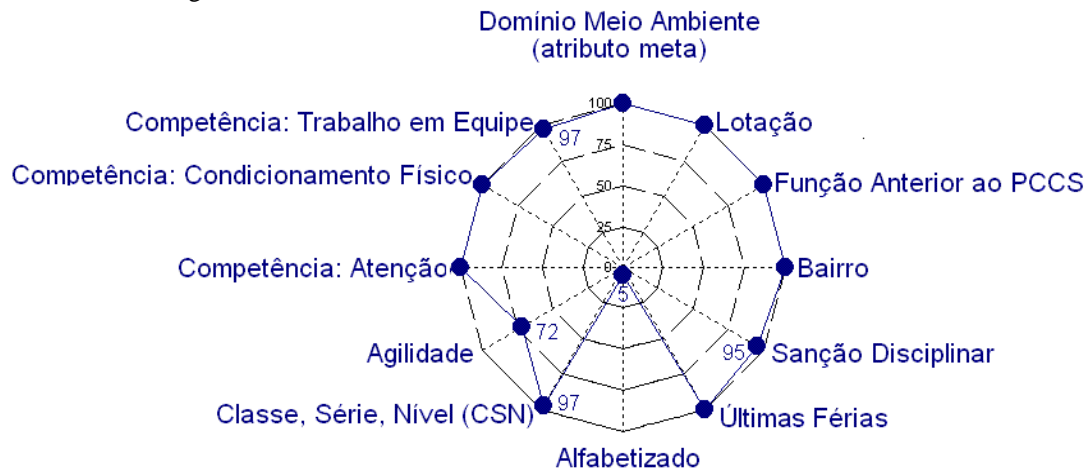
O período do gozo das últimas férias foi classificado como atributo previsor dos domínios: “Psicológico”, “Relações Sociais” e “Meio Ambiente” e, de 48% dos aspectos de QV.

¹ Atributo meta: termo adotado neste trabalho para identificar o atributo de referência utilizado na seleção dos atributos relevantes de um grupo de atributos. Termo geralmente utilizado em tarefas de Data Mining classificação.

A função ocupada antes do PCCS foi considerada um atributo predictor do domínio Meio Ambiente e 48% dos aspetos de QV.

O domínio “Meio Ambiente” apresentou o maior número de atributos previsores (Figura 7).

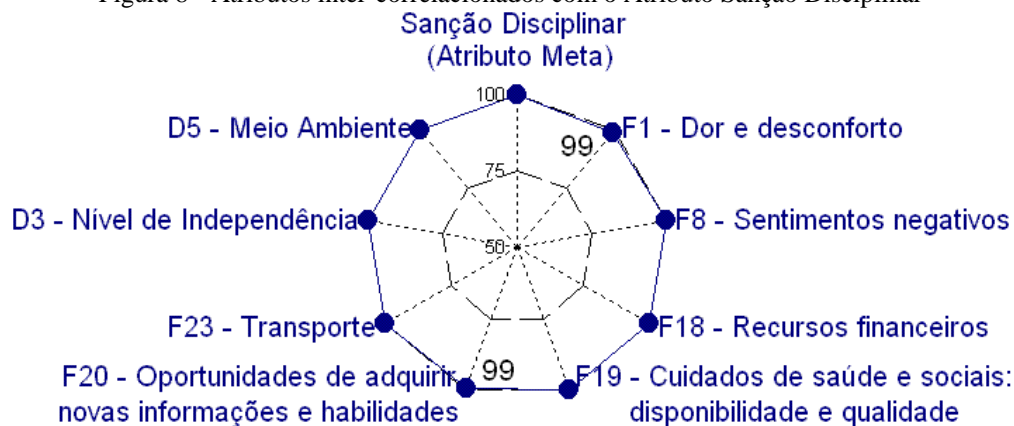
Figura 7 – Atributos inter-relacionados com o Domínio Meio Ambiente



Fonte: Santos (2007)

Neste trabalho, após a constatação de que o atributo “Sanção Disciplinar” e “Disfunção Temporária”, são atributos previsores de QV, mais especificamente do Domínio Meio Ambiente (Figura 8), decidiu-se verificar se existia dependência positiva destes atributos, com outros indicadores de QV. Esta investigação resultou na seleção dos indicadores de QV que poderiam estar contribuindo na ocorrência de sanções disciplinares ou disfunção temporárias. A Figura 8 apresenta oito atributos previsores do atributo “Sanção Disciplinar”.

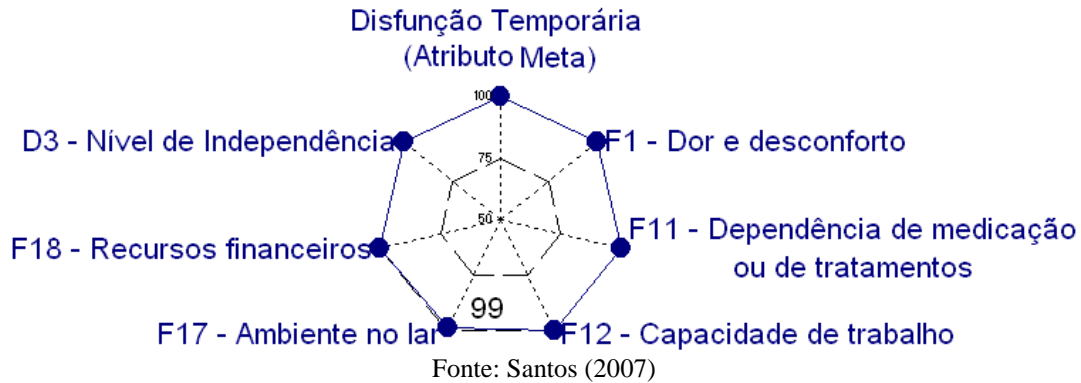
Figura 8 - Atributos inter-correlacionados com o Atributo Sanção Disciplinar



Fonte: Santos (2007)

A Figura 9 apresenta seis indicadores de QV que poderiam estar interferindo na ocorrência de disfunções temporárias.

Figura 9 - Atributos inter-correlacionados com o Atributo Sanção Disciplinar



Esta etapa de seleção de dados permitiu identificar os dados relevantes para a realização da mineração de dados. Uma vez identificada a correlação, buscaram-se modelos que possam prever as variáveis dependentes em função de variáveis previsoras (BATANERO, 1999).

6.2 Data Mining

Esta etapa também foi realizada de forma acíclica. Após cada seleção de atributos (etapa anterior), foi aplicado o algoritmo Apriori (AGRAWAL; SRIKANT, 1994).

O algoritmo permite trabalhos com grande número de atributos, gerando várias alternativas combinatórias entre eles. Realiza buscas contínuas em toda a base de dados, mantendo um ótimo desempenho em termos de tempo de processamento (AGRAWAL; SRIKANT, 1994).

A configuração do algoritmo no Ambiente WEKA, permitiu definir apenas duas medidas de interesse, o suporte e mais uma (Confiança ou *Lift* ou *Leverage* ou Convicção). Então, definiram-se as seguintes configurações mínimas: Suporte = 0.01 (equivalente a três registros) e Convicção = 1.1 (dependência positiva). Mantendo-se as demais configurações padrões.

Estes procedimentos resultaram na geração 327157 RA, distribuídas em 34 arquivos de saída.

6.3 Pós-Processamento de Data Mining

Esta etapa iniciou-se com a transformação dos resultados da mineração de dados e foi finalizada na sub-etapa de conhecimento adquirido.

6.3.1 Transformação

Após a geração das RA, os resultados foram transformados em uma base de dados com os resultados identificados por atributo meta.

6.3.2 Seleção das Regras de Associação

No processo de seleção das RA utilizaram-se medidas de interesse subjetivas e objetivas.

Definiu-se como medida de interesse a subjetividade, a preferência por regras que possuíssem no seu conseqüente, os piores indicadores de QV (identificados na etapa de conhecimento da base de dados); os dados cadastrais (identificados na etapa de seleção de

dados) que possivelmente sofreriam interferência dos indicadores de QV; a média de Qualidade de Vida e; as Questões Gerais de Qualidade de Vida.

Segundo Silberschatz e Tuzhilin (1995), o grau de interesse de uma RA é avaliado pelo usuário, levando em consideração as suas expectativas e conhecimento prévio dos dados.

Como expectativa do usuário, procurou-se identificar entre os indicadores de QV, os atributos mais preditivos entre eles, pois a inferência direta sobre o atributo previsor poderia contribuir na alteração dos atributos dependentes. Então, buscou-se identificar os atributos previsores do maior número de atributos dependentes, considerando-se preferencialmente, aqueles que não possuíam dependência funcional. Esta identificação foi possível através do uso do filtro *AttributeSelectio*, onde foi possível identificar como atributos mais promissores para sofrer algum tipo de inferência, as QGQV e o Aspecto Segurança física e proteção (F16), pois se apresentam com atributos previsores mais frequentes nos grupos de dados selecionados.

O atributo QGQV foi considerado atributo previsor de 11 indicadores de QV, e não possui dependência funcional de nenhum deles. Enquanto o Aspecto Segurança Física e Proteção foi considerado atributo previsor de 11 indicadores de QV, dois quais, apenas dois possuíam dependência funcional (D5 e MGQV).

As medidas de interesse objetivas adotadas foram:

- Dependência positiva do atributo meta;
- Dependência positiva potencialmente implícita (*Convicção* < 5.0);
- Confiança ≥ 0.60 (equivalente ao mínimo de 3/5 dos antecedentes);
- Separadas por classe do atributo meta; e
- Ordenadas pelos maiores valores de: Suporte, Confiança, Convicção, *Lift* e *Leverage*; e
- Maior número de itens no antecedente da RA.

Depois de delimitada a estratégia de seleção, as RA foram selecionadas através de SQL específicas ou de processos de buscas simples (em arquivos menores) e contabilizadas na, onde se pode verificar a significativa contribuição das unidades de medidas na diminuição dos custos com o processo de pós-processamento.

A partir dos dados da Tabela 3, observou-se uma diminuição no número de RA avaliadas de 4,20% de forma subjetiva, e de 99% de forma objetiva.

Tabela 3 – Total de RA com os Resultantes

Atributos metas da Intercorrelação: Indicadores de Qualidade de Vida e Dados Cadastrais		Número de Atributos Intercorrelacionados	Total de RA							Interessantes / Geradas
			Definidas no Algoritmo S ≥ 0.01 Cv ≥ 1.1	Processo de Busca C ≥ 0.60 L ≥ 1.01 ou Lv ≥ 0.01 Cv ≤ 5.0						
				Classes do Atributo Meta						
				Péssimo	Ruim	Médio	Bom	Excelente	Total	
MQV*		2	36	0	1	0	0	8	9	25%
Domínios de QV	D1	1	10							
	D2	4	338							
	D3	2	46							
	D4	3	163							
	D5*	11	198000	0	0	1336	222	0	1558	0.79%
	D6	2	26							
Aspectos de QV	F1*	3	136	0	14	11	2	0	27	19,85%
	F2	3	253							
	F3	5	1096							
	F4	3	269							
	F5	4	371							
	F6	3	161							
	F7	4	708							
	F8	3	214							
	F9	2	42							
	F10	4	896							
	F11	2	29							
	F12	6	4366							
	F13	4	787							
	F14	4	252							
	F15	2	36							
	F16*	4	755	0	18	49	6	0	73	9,67%
	F17	3	110							
	F18	3	104							
	F19	5	2358							
	F20	3	281							
	F21	1	12							
	F22	6	481							
	F23	4	300							
	F24	2	26							
QGQV*		4	730	0	0	25	43	0	68	9.32%
Sanção Disciplinar*		8	101166	Sim =35					35	0.01%
Disfunção Temporária*		6	12599	Sim =11					11	0.09%
Total		14	327157						1781	0.5%

*Atributos selecionados através de medidas de Interesse Subjetivas.

Fonte: Santos (2007)

As RA dos indicadores de QV e atributos cadastrais escolhidos para investigação, formam organizados em quadros e, ordenados por classes e por maior suporte (Exemplo: Quadro 3), para facilitar as suas interpretações na etapa a seguir.

Quadro 3 – Regras de Associação com Conseqüente: Média de Qualidade de Vida(MQV)

MQV	Antecedente→MQV(Suporte(S); Confiança(C); Lift(L); Leverage(Lv); Convicção(Cv))
Excelente S=0.62	*
Bom S=0.10	Bairro=Oficinas→MQV=Bom(S=0.056; C=0.93; L=1.5; Lv=0.02; Cv=2.85) Bairro=Cará-Cará→MQV=Bom(S=0.023; C=1; L=1.61; Lv=0.01; Cv=2.28) Lotação=Colégio Agrícola Estadual Augusto Ribas E Bairro=Uvaranas→MQV=Bom(S=0.023; C=1; L=1.61; Lv=0.01; Cv=2.28) Lotação=Seção de Paisagismo E Bairro=Uvaranas → MQV=Bom(S=0.019;C=1; L=1.61; Lv=0.01; Cv=1.9) Lotação=Colégio Agrícola Estadual Augusto Ribas→ MQV=Bom(S=0.053; C=0.85; L=1.36; Lv=0.01; Cv=1.65) Lotação=Laboratório Universitário de Análises Clínicas→MQV=Bom(S=0.015;C=1; L=1.61; Lv=0.01; Cv=1.52) Bairro=Núcleo Pimentel→MQV=Bom(S=0.015; C=1; L=1.61; Lv=0.01; Cv=1.52) Lotação=Departamento de Odontologia→MQV=Bom(S=0.053; C=0.79; L=1.27; Lv=0.01; Cv=1.33)
Médio S=0.21	Lotação=Seção de Obras→MQV=Médio(S=0.019; C=0.6; L=2.75; Lv=0.01; Cv=1.3)
Ruim S=0.005	*
Péssimo S=0.004	*
SR S=0.053	*

* Inexistência de padrões com os valores mínimos das Medidas de Interesse Objetivas adotadas.

Fonte: Santos (2007)

6.3.3 Interpretação dos Resultados

Para auxiliar na interpretação das RA, o ambiente WEKA permite configurar o algoritmo para apresentar a frequência, na base de dados, de todos os elementos envolvidos nas RA (conseqüentes ou antecedentes). Algumas dessas informações, como o suporte de cada classe do elemento conseqüente da RA, podem ser verificada na Quadro 3.

Nesta etapa as principais RA foram transformadas em Regras de Produção, para facilitar a sua interpretação, conforme o modelo a seguir:

SE

Lotação = Setor de Alimentação

ENTÃO

D5 = Médio

Medidas de Interesse Objetivas (S =0.045; C =0.92; Cv =2.77)

Esta RA indica que 4.5% dos servidores trabalham no Setor de Alimentação e, que 92% destes servidores tem o domínio de QV: Meio Ambiente classificado como Médio. E que a possibilidade de um servidor ter a sua QV classificada como média é 2.77 vezes maior, caso esteja lotado no Setor de Alimentação.

6.3.4 Aquisição de Conhecimento

Na etapa de Aquisição de Conhecimento ocorre a transformação do conhecimento implícito em explícito. Esta transformação ocorreu através da análise dos resultados dos indicadores de QV e da interpretação das RA de associação.

A média da QV da amostra pode ser classificada, segundo a escala adotada, como boa, mas percebe-se que pode ser melhorada em até 32,5%. Essa melhoria pode ser alcançada inferindo inicialmente, nos piores indicadores de QV, entre eles, os mais preditivos. E, posteriormente, nos fatores passíveis de inferências, apresentados nas RA como elementos implicadores das piores classes de dos indicadores de QV.

O Aspecto Segurança Física e Proteção apresentou o pior indicador de QV do Domínio Meio Ambiente.

Entre as facetas de QV, observa-se que o Aspecto Dor e Desconforto apresentou o pior julgamento.

A partir da interpretação das RA selecionadas, sabe-se que existe um padrão de comportamento entre os indivíduos com os piores indicadores, que serão descritos a seguir.

Os agentes universitários que estão lotados na Seção de Obras merecem atenção especial, pois obtiveram a menor classificação da MGQV com padrão de comportamento.

Em relação as QGQV, que examinam as maneiras pelas quais uma pessoa avalia a sua qualidade de vida, a menor classificação foi dada aos servidores: ocupantes da função de Ajudante Geral, que residem no Bairro de Uvaranas; lotados na Seção Elétrica; lotados no Setor de Alimentação que ocupavam a Função de auxiliar de Cozinha e os residentes no Bairro da Palmeirinha.

Os servidores que sofreram sanção disciplinar e/ou ocupavam a função de Ajudante Geral e/ou que a função exige determinadas competências (Agilidade e/ou Atenção e/ou Condicionamento Físico e/ou Trabalho), possui a menor classificação do Domínio Meio Ambiente, com um padrão de comportamento.

Apresentaram o Aspecto Dor e Desconforto classificado como ruim, os servidores: lotados no CAAR; que residem no Bairro Cara-Cará; que recebem salário máximo de 3 salários mínimos e/ou estão lotados na Seção de Alimentação ou de Vigilância ou no CAIC e/ou residem no Bairro de Uvaranas.

Em relação ao Aspecto Segurança Física e Proteção, obtiveram a pior classificação (F16 = ruim), os servidores: lotados na Seção Elétrica ou Seção de Obras ou Seção de Alimentação e/ou que residem nos Bairros: Cará-Cará, Palmeirinha e Uvaranas.

Observou-se, que alguns indicadores de QV poderiam estar contribuindo para a ocorrência de sanção disciplinar ou de disfunção temporária de servidores.

O Aspecto Recursos Financeiros e Cuidados de Saúde e Sociais, classificados respectivamente com ruim e péssimos; ou o aspecto Sentimentos Negativos e o Domínio Nível de Independência e/ou Domínio Meio Ambiente, classificados, respectivamente como péssimo e médios, poderiam estar contribuindo na ocorrência de uma sanção disciplinar.

Observou-se que a combinação de pelo menos dois dos indicadores de QV a seguir, poderiam estar contribuindo com a ocorrência de um desvio de função temporária do servidor:

- F11 - Dependência de medicação ou de tratamentos – médio;
- D3 - Nível de independência – ruim;
- F1 - Dor e desconforto – ruim;

- F18 - Recursos financeiros – ruim e;
- F12 - Capacidade de trabalho – péssimo;

Finalizando, observa-se que merecem atenção especial dos responsáveis pela gestão de QV da instituição, principalmente, os servidores ocupantes das funções de Ajudante Geral ou Cozinheiro; lotados na Seção Elétrica ou Seção de Obras ou Seção de Alimentação; que residem nos Bairros Cará-Cará ou Palmeirinha.

7. Conclusões

As técnicas de *Data Mining*, mais especificamente, as técnicas de Regras de Associação, podem ser utilizadas para complementar a análise dos resultados de um instrumento de pesquisa estruturada para *surveys*, principalmente quando se dispõe de muitos dados adicionais e se deseja descobrir conhecimento implícito.

O planejamento e a utilização combinada de medidas de interesse subjetivas e objetivas são de extrema importância para a economia de custos operacionais das tarefas de KDD de descrição.

O uso de medidas de interesse, neste trabalho, reduziu o número de RA a serem investigadas em 4,20% e 99,5%, através da respectiva utilização de medidas de interesse subjetivas e objetivas. Isto evidencia a importância do processo de filtragem, principalmente, através do uso das medidas de interesse objetivas; e do processo de planejamento, através das medidas de interesse subjetivas.

Os seguintes padrões de comportamento, potencialmente implícitos, dos agentes universitários da instituição objeto da avaliação de qualidade de vida foram evidenciados:

- A lotação do trabalhador é determinante na sua qualidade de vida, em seus diferentes aspectos domínios;
- O bairro residencial interfere na maioria dos domínios e aspectos de QV, exceto no Domínio Físico e Aspecto Participação em oportunidades de recreação/ lazer.
- O salário e o período de férias interferem em alguns aspectos de qualidade de vida;
- A exigência ou não de algumas competências da função, também pode contribuir na classificação da sua QV;
- As funções mais específicas, ocupadas na instituição, também possuem padrões de comportamento em relação à QV, principalmente no Domínio e Aspectos de QV do Meio Ambiente.
- A ocorrência de sanção disciplinar contribui para a diminuição na qualidade de vida no Domínio Meio Ambiente; e tem maiores chances de ocorrer, quando alguns aspectos de Qualidade de Vida encontram-se classificados, como péssimo ou ruim; e
- Uma disfunção temporária tem maiores possibilidades de ocorrer, quando existe a combinação de baixos indicadores de qualidade de vida.

Através desses resultados comprova-se a importância do uso de técnicas de Mineração de Dados, no monitoramento de fatores que possam estar interferindo nos indicadores de Qualidade de Vida, pois com a utilização apenas dos métodos estatísticos propostos no instrumento WHOQOL, seria possível apenas a identificação dos índices de QV da amostra ou de determinados segmentos, enquanto através do uso das técnicas de *Data Mining*, foi permitido identificar entre os dados disponíveis, fatores que contribuem na predição dos indicadores de qualidade de vida.

O produto de um processo de KDD, quando aplicado em um instrumento de avaliação de qualidade de vida, poderá auxiliar os gestores de qualidade de vida da instituição

em suas tomadas de decisões. Assim como, realização de treinamentos, acompanhamento psicológico, assistência social e outras tarefas, com focos mais precisos podem ser aplicadas.

A utilização das técnicas de *Data Mining* poderá contribuir na elaboração de instrumentos mais específicos para determinados grupos ou culturas. E, também, podem ser utilizadas para definir questões de re-testes para avaliar a eficiência de inferências pontuais.

Referências

ADRIAANS, P; ZABTINGE, D. **Data Mining**. Harlow: Addison-Wesley, 1996. 176 p.

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. In: CONFERENCE ON VERY LARGE DATABASES, 20., 1994, Santiago de Chile, Chile. **Proceedings ...** San Francisco, CA: Morgan Kaufmann, 1994. p. 487-499.

AMARAL, F. C. N. **Data Mining**: Técnicas e Aplicações para o Marketing Direto. 1. ed. São Paulo: Berkeley, 2001, 110p.

BATANERO, C. Análise Exploratória de dados nos cursos de segundo grau. In: CONFERÊNCIA INTERNACIONAL "EXPERIÊNCIAS E EXPECTATIVAS DO ENSINO DE ESTATÍSTICA - DESAFIOS PARA O SÉCULO XXI", 1999, Florianópolis. **Atas...** Florianópolis: UFSC, 1999.

BORGELT, C; KRUSE, R. Induction of Association Rules: Apriori Implementation. In: CONFERENCE ON COMPUTATIONAL STATISTICS, 15., 2002. Berlin, Germany. **Proceedings...** Heidelberg, Germany: Physica Verlag, 2002, p. 395-400.

BRIN, S.; MOTWANI, R.; ULMAN, J. D.; TSUR, S. Dynamic Itemset counting and implication rules for market basket data. In: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 1997, Tucson, Arizona, USA. **Proceedings ...** New York: AMC, 1997, p. 255-264.

CARVALHO, D. R. **Data Mining Através de Indução de Regras e Algoritmos Genéticos**, 1999. 142p. Dissertação (Mestrado em Informática Aplicada) - Pontifícia Universidade Católica do Paraná, Curitiba, 1999.

CARVALHO, L. A. V. **Data Mining**: a Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração. São Paulo: Erica, 2001. 237 p.

FAYYAD, U. M. Diving into databases: SQL is helpless in the face of massive, accumulating data stores. **Database programming and design**. San Mateo, v. 11, n. 3, p. 24-31, mar. 1998.

FAYYAD, U. et al. **Advances in Knowledge Discovery and Data Mining**. Menlo Park, CA: AAAI Press / The MIT Press. 1996, 625 p.

FAYYAD, U. M.; UTHURUSAMY, R. International Conference on Knowledge Discovery And Data Mining, 1., 1995. Montreal, Canadá: AAAI Press, 1995, 348 p.

FAYYAD, U. M.; UTHURUSAMY, R. Workshop, Technical Report Knowledge Discovery in Databases, 1994. **Proceedings ...** Seattle, WA: AAAI, 1994, 473 p.

- FLECK, M. P. A. A Application of the Portuguese version of the instrument for the assessment of the quality of life of the World Health Organization (WHOQOL-100), **Revista de Saúde Pública**. São Paulo, v. 33, n. 2, p. 198-205, abr. 1999.
- FOX-RUSHBY, J.; PARKER, M. Culture and the measurement of health-related quality of life, **Journal of European Applied Psychology**, Paris, v. 45, n. 4, p. 257-263, 1995.
- HALL, M. A. **Correlation-based feature subset selection for machine learning**. Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1998.
- LOPES, C. H. P. **Classificação de registros em banco de dados por evolução de regras de associação**: utilizando algoritmos genéticos, 1999, 136p. Dissertação (Mestrado em Engenharia Elétrica e Sistemas de Computação) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1999.
- MCGARRY, K. A survey of interestingness measures for knowledge discovery. **The Knowledge Engineering Review**, Cambridge University, New York, USA, v. 20, n. 1, p. 39-61, mar. 2005.
- MICHALSKI, R.; KAUFMAN, K. **Data Mining and Knowledge Discovery**: a review of Issues and multistrategy approach. In: MICHALSKI, R. S.; BRATKO, I. KUBAT, M. **Machine Learning and Data Mining Methods and Applications**, England: John Wiley, 1998.
- RODRIGUES, A. M. **Técnicas de Data Mining Classificadas do Ponto de Vista do Usuário**. Rio de Janeiro, 2000. 116f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000.
- SANTOS, C. B. **Análise dos Dados do WHOQOL-100 utilizando Data Mining**. Ponta Grossa, 2007. 108f. Dissertação (Mestrado em Engenharia da Produção) - Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2007.
- SILBERSCHATZ, A.; TUZHILIN, A. On subjective measures of interestingness in knowledge discovery. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 1., 1995, Montreal, Canadá. **Proceedings ...** Chicago, USA: AAAI Press, 1995, p.275-281.
- UEPG - UNIVERSIDADE ESTADUAL DE PONTA GROSSA. Portaria R. nº 161 - UEPG de 1º de agosto de 2006. Procede a correlação de funções e o enquadramento em três etapas, na carreira técnica universitária para os servidores ocupantes do cargo de Agente Universitário da UEPG, nos termos do anexo que passa a integrar este ato legal. Relator: Paulo Roberto Godoy. **Atos Oficiais da Universidade Estadual de Ponta Grossa. Ponta Grossa**, 1º ago. 2006. p.1-30.
- VASCONCELOS, E. M. **Complexidade e pesquisa interdisciplinar**: epistemologia e metodologia operativa. Petrópolis: Vozes, 2002.
- VIANA, R. Mineração de dados: teoria e prática. **Revista SQL Magazine**, Rio de Janeiro: Neofício, n. 10, 2004.
- WHOQOL Group. Study protocol for the World Health Organization project to develop a quality of life assessment instrument (WHOQOL). **Quality of Life Research**. v. 2, n. 2, p. 153-159, apr. 1993.