

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
MESTRADO EM ENGENHARIA DE PRODUÇÃO

DANIEL HENRIQUE PERUCELLI ROSAS

APLICAÇÃO DE TÉCNICAS DE VISUALIZAÇÃO DE INFORMAÇÕES
EM UMA FERRAMENTA DE DESCOBERTA DE CONHECIMENTO

DISSERTAÇÃO

PONTA GROSSA

2013

DANIEL HENRIQUE PERUCELLI ROSAS

**APLICAÇÃO DE TÉCNICAS DE VISUALIZAÇÃO DE INFORMAÇÕES
EM UMA FERRAMENTA DE DESCOBERTA DE CONHECIMENTO**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Engenharia de Produção, no Programa de Pós-Graduação em Engenharia de Produção, Universidade Tecnológica Federal do Paraná. Área de concentração: Gestão Industrial.

Orientadora: Professora Dra. Simone Nasser Matos.

PONTA GROSSA

2013

Ficha catalográfica elaborada pelo Departamento de Biblioteca
da Universidade Tecnológica Federal do Paraná, Campus Ponta Grossa
n.026/13

R788 Rosas, Daniel Henrique Perucelli

Aplicação de técnicas de visualização de informações em uma ferramenta de descoberta de conhecimento. / Daniel Henrique Perucelli Rosas. -- Ponta Grossa, 2013.

98 f. : il. ; 30 cm.

Orientadora: Profa. Dra. Simone Nasser Matos

Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção. Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2013.

1. Sistemas de informação. 2. Mineração de dados - (Computação). 3. Gestão do conhecimento. I. Matos, Simone Nasser. II. Universidade Tecnológica Federal do Paraná, Campus Ponta Grossa. III. Título.

CDD 670.42



Universidade Tecnológica Federal do Paraná
Campus Ponta Grossa
Diretoria de Pesquisa e Pós-Graduação
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO



FOLHA DE APROVAÇÃO

Título da Dissertação Nº **228/2013**

APLICAÇÃO DE TÉCNICAS DE VISUALIZAÇÃO DE INFORMAÇÕES EM UMA
FERRAMENTA DE DESCOBERTA DE CONHECIMENTO

por

Daniel Henrique Perucelli Rosas

Esta dissertação foi apresentada às **16 horas** de **09 de julho de 2013** como requisito parcial para a obtenção do título de MESTRE EM ENGENHARIA DE PRODUÇÃO, com área de concentração em Gestão Industrial, Programa de Pós-Graduação em Engenharia de Produção. O candidato foi argüido pela Banca Examinadora composta pelos professores abaixo citados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

Prof. Dr. Luciano José Senger (UEPG)

Prof. Dr. Pedro Paulo de Andrade Junior
(UTFPR)

Prof. Dr. Antonio Carlos de Francisco (UTFPR)

Prof. Dra. Simone Nasser Matos (UTFPR)
- *Orientador*

Prof. Dr. Aldo Braghini Junior (UTFPR)
Coordenador do PPGEP

A FOLHA DE APROVAÇÃO ASSINADA ENCONTRA-SE NO DEPARTAMENTO DE REGISTROS ACADÊMICOS DA UTFPR-CÂMPUS PONTA GROSSA

AGRADECIMENTOS

Este trabalho representou um passo muito importante em minha vida, foi um tempo de aprendizado, de conhecimento dos meus limites e perceber que poderia atingir meus objetivos.

Neste espaço venho agradecer a todos que tiveram participação no desenvolvimento e conclusão desta longa jornada.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo incentivo financeiro concedido por meio de bolsa de estudo.

Agradeço aos meus professores do mestrado que transmitiram suas experiências, contribuindo para meu crescimento intelectual.

Sou grato, em especial, a minha orientadora Professora Doutora Simone Nasser Matos por sua enorme paciência em me acompanhar e pela sabedoria com que me auxiliou neste trabalho.

Aos meus colegas de curso que me ajudaram em muitas situações de dúvida durante o período do curso.

Aos autores das referências que me ajudaram no desenvolvimento e fundamentação de minha pesquisa.

Agradeço antecipadamente, aqueles que vão tomar meu trabalho como referência para construírem os seus.

E, principalmente, a Deus por ter escutado minhas orações, pedidos, me confortado e me fortalecido para continuar e chegar até este momento.

“Não seremos limitados pela
informação que temos. Seremos limitados
por nossa habilidade de processar esta
informação.”

Peter Drucker

RESUMO

ROSAS, Daniel Henrique P. **Aplicação de Técnicas de Visualização de Informações em uma Ferramenta de Descoberta de Conhecimento**. 2013. 98f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2013.

Ferramentas de descoberta de conhecimento que usam a mineração de dados para extração de informações geram relatórios que são analisados e podem servir de base para a tomada de decisão. Mas, a percepção do conhecimento obtido pode ser comprometida caso os resultados fornecidos pelas ferramentas utilizem termos técnicos, linguagem de máquina ou o formato de exibição das informações seja inteligível. Por isso, técnicas que utilizam recursos gráficos para expor informações passam a ser importantes na descoberta de conhecimento podendo ser aplicadas aos resultados de modo a facilitar a compreensão da informação exposta. Este trabalho correlaciona o conhecimento fornecido pelas técnicas de mineração de dados com as características gráficas das técnicas de visualização de informações de modo que auxiliem na compreensão dos resultados fornecidos pelas ferramentas de descoberta de conhecimento. Apresenta também as técnicas de visualização de informações mais apropriadas para transmitir o conhecimento obtido além de exibir os protótipos das representações gráficas dos resultados gerados por técnicas e algoritmos de mineração da ferramenta Weka. A aplicação de recursos visuais visa simplificar a interpretação do conhecimento extraído e fortalecer a base de informações para tomada de decisão das organizações.

Palavras-chave: Visualização de Informações. Mineração de Dados. Sistemas de Informação.

ABSTRACT

ROSAS, Daniel Henrique P. **Application of Information Visualization Techniques in a Knowledge Discovery Tool**. 2013. 98p. Dissertation (Master in Production Engineering) – Federal Technological University of Paraná. Ponta Grossa, 2013.

Tools for knowledge discovery using data mining to extract information generate reports that are reviewed and may serve as a basis for decision making. But the perception of knowledge gained can be compromised if the results provided by the tools using technical terms, machine language or the display format of the information to be intelligible. Therefore, techniques that use information to display graphics resources become important in knowledge discovery can be applied to the results in order to facilitate the understanding of the information displayed. This work correlates the knowledge provided by the techniques of data mining with graphic features of information visualization techniques in order to assist in the understanding of the results provided by the tools of knowledge discovery. It also presents information visualization techniques most appropriate to convey the knowledge gained in addition to displaying prototypes of graphical representations of the results generated by techniques and algorithms mining tool Weka. The use of visuals to simplify the interpretation of the extracted knowledge and strengthen the information base for decision-making organizations.

Keywords: Information Visualization. Data Mining. Information Systems.

LISTA DE FIGURAS

Figura 1 – Processo Geral de Desenvolvimento do Trabalho.....	38
Figura 2 – Categorias e técnicas de visualização de informações.....	43
Figura 3 – Relação tarefas e métodos das técnicas de mineração	52
Figura 4 – Cabeçalho do resultado de mineração da base Consumidor	53
Figura 5 – Conhecimento Novo resultante da mineração pela técnica de Classificação .	54
Figura 6 – Informações Adicionais fornecidas pela técnica Classificação	55
Figura 7 – Cabeçalho fornecido pela técnica Clusterização	56
Figura 8 – Conhecimento Novo fornecido pela técnica Clusterização	57
Figura 9 – Informações Adicionais fornecida pela técnica Clusterização	57
Figura 10 – Cabeçalho e Informações Adicionais fornecidos pela técnica Associação	58
Figura 11 – Conhecimento Novo fornecido pela técnica Clusterização	59
Figura 12 – Técnicas de visualização que correspondem às características do algoritmo J48.....	65
Figura 13 – Técnicas de visualização que correspondem as características do algoritmo K-means.....	66
Figura 14 – Técnicas de visualização que correspondem as características do algoritmo Apriori.....	66
Figura 15 – A raiz dá início à árvore Bifocal Tree.....	69
Figura 16 – Nós-filhos da raiz exibidos na área de detalhe	69
Figura 17 – Raiz “kitchen ≤ 3 ” na área de contexto e o nó-filho “frequent $\leq 2,38$ ” na área de detalhe	70
Figura 18 – Nó-pai “frequent $\leq 2,38$ ” na área de contexto e nó-filho “statecod” na área de detalhe.....	70
Figura 19 – Nó-pai “statecod” na área de contexto e o nó-filho na área de detalhe.....	71
Figura 20 – Raiz “kitchen ≤ 3 ” tem a maior divisão.....	71
Figura 21 – Nó-pai “statecod” e seus nós-filhos	72
Figura 22 – Nó-pai “statecod =ma” e seus nós-filhos.....	72
Figura 23 – Estrutura da árvore de decisão por Cheops.....	73
Figura 24 – Cheops, as informações compactadas aparecem com a seleção do icone	74
Figura 25 – Árvore de decisão pela técnica Grafos.....	74
Figura 26 – Árvore de decisão pela técnica Information Slices.....	75
Figura 27 – Disco à direita mostra os subníveis referentes ao nó-filho selecionado	75
Figura 28 – Esquema de apresentação da técnica Information Cube.....	76

Figura 29 – Classes distribuídas na janela de exibição.....	77
Figura 30 – Conteúdo detalhado do cluster movido para o centro	77
Figura 31 – Classes representadas pela técnica Coordenadas Paralelas	78
Figura 32 – Conhecimento Novo representado pela técnica Grafos	78
Figura 33 – Representação das clusteres do algoritmo K-means por Treemap	79
Figura 34 – Representação das clusteres do algoritmo K-means por Treemap	80
Figura 35 – Pixeis de cores diferentes para classes diferentes.....	80
Figura 36 – Cada atributo de um cluster é representado por um pixel colorido.....	81
Figura 37 – Relações de dependência por cores na Figura de Arestas	82
Figura 38 – Relação de dependência representada pela técnica Grafo.....	82

LISTA DE QUADROS

Quadro 1 – Exemplos de ferramentas de mineração de dados	25
Quadro 2 – Sistemas de mineração e os critérios de usabilidade	40
Quadro 3 – Tarefas e métodos dos algoritmos das técnicas de mineração de dados	42
Quadro 4 – Tarefas, métodos e suporte de informações das técnicas de visualização	44
Quadro 5 – Correlação entre técnicas de mineração de dados e técnicas de visualização de informações	45
Quadro 6 – Elementos visuais usados pelas técnicas de visualização de informações	47
Quadro 7 – Relação Base x Técnica de Mineração x Algoritmo	50
Quadro 8 – Tarefas e métodos dos algoritmos das técnicas de mineração de dados	51
Quadro 9 – Tarefas, métodos e suporte das técnicas de visualização	60
Quadro 10 – Correlação entre técnicas de mineração de dados e técnicas de visualização de informações	64
Quadro 11 – Elementos visuais usados pelas técnicas de visualização de informações ...	67

SUMÁRIO

1 INTRODUÇÃO	13
1.1 JUSTIFICATIVA.....	14
1.2 OBJETIVOS	15
1.2.1 Objetivo Geral.....	16
1.2.2 Objetivos Específicos	16
1.3 ORGANIZAÇÃO DO TRABALHO	16
2 REFERENCIAL TEÓRICO	17
2.1 SISTEMAS DE INFORMAÇÃO	17
2.1.1 Sistemas de Mineração de Dados.....	19
2.2 REPRESENTAÇÃO VISUAL DE INFORMAÇÕES	25
2.2.1 Técnicas de Visualização de Informações	28
2.2.2 Níveis de Interação do Usuário	33
2.3 TRABALHOS RELACIONADOS.....	34
3 METODOLOGIA	37
3.1 CLASSIFICAÇÃO DA PESQUISA.....	37
3.2 PROCESSO DE DESENVOLVIMENTO DA PESQUISA	38
3.2.1 Escolher a Ferramenta e Técnicas de Mineração de Dados	39
3.2.2 Analisar as Técnicas de Mineração de Dados	41
3.2.3 Analisar as Técnicas de Visualização de Informações.....	43
3.2.4 Correlacionar as Técnicas de Mineração e Visualização	44
4 ADEQUAÇÃO DAS TÉCNICAS DE VISUALIZAÇÃO ÀS TÉCNICAS DE MINERAÇÃO DE DADOS	49
4.1 BASES DE DADOS PARA O EXPERIMENTO	49
4.2 ANÁLISE DOS RESULTADOS GERADOS PELAS TÉCNICAS DE MINERAÇÃO. 50	
4.2.1 Experimento com a Base Consumidor	53
4.2.2 Experimento com a Base Seguros.....	56
4.2.3 Experimento da Base Supermercado.....	58
4.3 ANÁLISE DAS TÉCNICAS DE VISUALIZAÇÃO	60
4.3.1 Técnicas de Visualização Geométricas	61
4.3.2 Técnicas de Visualização Orientadas a Pixel	61
4.3.3 Técnicas de Visualização Iconográficas.....	61
4.3.4 Técnicas de Visualização Hierárquicas	62

4.4 CORRELAÇÃO DAS TÉCNICAS DE VISUALIZAÇÃO COM TÉCNICAS DE MINERAÇÃO	63
4.5 PROTÓTIPO DE INTERFACE GRÁFICA PARA VISUALIZAÇÃO DO CONHECIMENTO NOVO	68
4.5.1 Protótipo para Representação Visual do Algoritmo J48	68
4.5.2 Protótipo para Representação Visual do Algoritmo K-means.....	76
4.5.3 Protótipo para Representação Visual do Algoritmo Apriori	81
4.6 CARACTERÍSTICAS RELEVANTES DO USO DAS TÉCNICAS DE VISUALIZAÇÃO EM UMA FERRAMENTA DE DESCOBERTA DE CONHECIMENTO	83
5 CONCLUSÃO.....	85
5.1 TRABALHOS FUTUROS	86
REFERENCIAS.....	87

1 INTRODUÇÃO

O grande volume de dados gerados pelas organizações em suas relações internas e externas impossibilita a análise manual. Para resolver esta dificuldade as empresas usam sistemas de gestão de informações que capturam os registros e extraem informações que descrevem a situação da empresa ou de seus concorrentes.

Segundo Hoffmann; Oliveira; Zeferino (2012), os sistemas de informação são introduzidos para melhorar o processo de informação dentro da organização, dando suporte aos gerentes na tomada de decisão e orientar as atividades da empresa. Quando são usados na gestão operacional, por exemplo, estão relacionados com as informações de produção, como volume e custo, além de investimentos para atender ao crescimento do mercado e acompanhar as tendências em tecnologia e informação. Entretanto, sistemas tradicionais de gerenciamento apenas reúnem os registros num só ambiente e não possibilitam a construção de estratégias para a organização.

A fim de explorar estes registros na identificação de informação útil, outros tipos de ferramentas de gestão estão sendo criadas. Ferramentas como *Weka* que exploram grandes concentrações de dados em busca de padrões, tendências, problemas ou características de um determinado setor da empresa ou do ambiente mercadológico utilizando algoritmos de mineração de dados.

Mineração de dados é uma das etapas de um processo maior conhecido como Descoberta de Conhecimentos em Banco de Dados ou em inglês KDD (*Knowledge Discovery in Databases*) que inclui as tarefas de seleção, preparação e exploração das informações e por fim a análise e interpretação dos resultados obtidos em busca de assimilar o conhecimento extraído no processo (GALVÃO; MARIN, 2009).

A etapa de análise e interpretação dos resultados da mineração de dados pode se apresentar como um obstáculo à completa utilização do processo de descoberta de informação por exigir do usuário o domínio das informações e a compreensão do conhecimento extraído. Desta forma, um usuário que não esteja diretamente ligado aos registros coletados para serem minerados poderá ter problemas para entender as informações encontradas e os resultados gerados pelo processo.

Para facilitar o entendimento dos resultados obtidos pelos algoritmos de mineração de dados, pode-se incorporar um conjunto de recursos gráficos os quais representem

visualmente os resultados com o objetivo de explorar com alto desempenho o conhecimento, utilizando da capacidade perceptiva do usuário, facilitando a compreensão e a interpretação das informações apresentadas, na busca de novos conhecimentos (ALEXANDRE; TAVARES, 2007).

O objetivo principal da visualização de informações é facilitar a compreensão de conjuntos de informações que, à primeira vista, são volumosos ou não são facilmente compreendidos, tornando as informações que são relevantes objetos mais naturais, mais semelhantes a outros que já são conhecidos e criando metáforas que possam realizar essas aproximações da melhor maneira possível. Assim, a área de Visualização de Informações é capaz de unir os dois maiores sistemas de processamento de informações conhecidos: a mente humana e o computador (RAUTER; BENATO, 2006).

Na prática, o gerenciamento e a tomada de decisão tornam-se mais seguros e confiáveis quando as informações necessárias são entendidas de forma correta e facilitada.

Este trabalho tem a finalidade de estabelecer uma correlação entre os formatos de apresentação do conhecimento extraído por meio de técnicas de mineração de dados com as estruturas de apresentação das técnicas de visualização de informações de tal maneira que aperfeiçoe o entendimento de padrões, tendências ou regras nos resultados gerados pelos algoritmos da ferramenta de descoberta de conhecimento *Weka*.

O resultado deste trabalho é um subproduto de auxílio para o projeto do Grupo de Pesquisa em Sistemas de Informação da Universidade Tecnológica Federal do Paraná - GPSI (2013), que está desenvolvendo um *framework* na área de formação de preço de venda, tendo em um dos módulos que o compõe a descoberta de padrões usando a ferramenta *Weka*. Como o *framework* será usado por usuários leigos em informática, necessita-se de proporcionar a ele uma forma fácil de interpretar o conhecimento gerado.

1.1 JUSTIFICATIVA

A descoberta de informação por meio de ferramentas de mineração de dados tem se tornado um investimento necessário nas organizações, isto ocorre pelo número de registros que as relações organizacionais produzem a todo momento. Estes registros podem conter informações relevantes que podem apontar oportunidades de negócio como tendências no

comportamento dos clientes, o modo operante dos fornecedores e concorrentes, abrindo espaço para novos investimentos, renegociações, metas de produção e geração de novos produtos ou serviços.

Por este motivo, cada vez mais ferramentas que tratam grandes conjuntos de dados visando a extração de informações são colocadas no mercado. Quando se trata de apresentar informações armazenadas em bases de dados, deve-se pensar no uso dos recursos computacionais para enriquecer a interpretação de informações por parte dos usuários.

Uma das principais considerações a ser feita no processo de tomada de decisão é a disseminação das informações que fundamentam tal decisão, de tal forma que todos conheçam e tomem para si os mesmos objetivos. A transmissão destas informações precisa ser de fácil absorção e compreensão e uma forma de fazê-lo é utilizando recursos visuais que sejam familiares aos observadores.

A determinação de qual técnica será empregada é certamente dependente do tipo de informação que está sendo tratada, das características da técnica e das tarefas que podem ser realizadas pelo usuário. Além disto, é cada vez mais necessária a integração com sistemas de mineração de dados, já que a busca por facilitar o entendimento das informações passa pelo reconhecimento de padrões, estruturas e outros relacionamentos ocultos (FREITAS et al, 2001).

Procurando colaborar na descoberta de conhecimento no *framework* criado pelo Grupo de Pesquisa da UTFPR (GPSI, 2013) e observando o formato atual dos resultados da ferramenta de gestão da informação por meio da mineração de dados *Weka*, reunindo conceitos de representação visual de informações surgiu a seguinte questão: quais técnicas de visualização de informações podem melhorar a apresentação do conhecimento obtido pelos algoritmos de mineração de dados da ferramenta de gestão *Weka*?

1.2 OBJETIVOS

Visando estabelecer uma combinação de representação visual de informações entre as técnicas de visualização de informações com as técnicas de mineração de dados, os objetivos estão assim traçados.

1.2.1 Objetivo Geral

Correlacionar e aplicar as técnicas de visualização com as de mineração usando como base os resultados obtidos pela ferramenta de descoberta de conhecimento *Weka*.

1.2.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Identificar os modelos de conhecimento gerados pelas técnicas de mineração de dados.
- Levantar os modelos de conhecimento e a carga de informações apresentados por cada técnica de visualização de informação.
- Caracterizar os elementos visuais das técnicas de visualização adequadas.
- Propor protótipos usando técnicas de visualização para a representação do conhecimento obtido pela ferramenta *Weka*.

1.3 ORGANIZAÇÃO DO TRABALHO

Este trabalho está dividido em cinco capítulos. O capítulo 2 aborda a revisão da literatura sobre os assuntos relacionados ao tema, sendo conceitos de sistemas de informação, mineração de dados e visualização de informações. O capítulo 3 relata a classificação da pesquisa e a descrição da metodologia seguida para o desenvolvimento do trabalho. O capítulo 4 mostra os resultados da análise das características das técnicas de mineração e de visualização e da correlação entre ambas. Por fim, o capítulo 5, apresenta a conclusão desta pesquisa.

2 REFERENCIAL TEÓRICO

Este capítulo apresentará a revisão de literatura realizada para o desenvolvimento deste trabalho. A seção 2.1 relata sobre a importância do uso dos Sistemas de Informação e Mineração de Dados e das técnicas e ferramentas de mineração de dados. A seção 2.2 aborda sobre a Representação Visual de informações, apresentando as técnicas de visualização mais conhecidas. Na seção 2.3 estão os trabalhos que serviram de base para o desenvolvimento desta dissertação.

2.1 SISTEMAS DE INFORMAÇÃO

Nos últimos anos, têm-se verificado um crescimento exponencial na quantidade de informações produzidas e disponibilizadas aos usuários nos mais diversificados domínios de aplicação, gerando um grande volume de informações de diversas origens e formatos (VALIATI, 2008).

A informação influencia o cotidiano das pessoas, nas organizações essa influência é mais determinante, pois atinge o gerenciamento de processos, o comportamento de setores e pessoal, que por sua vez traz efeitos para a tomada de decisão (GOUVEIA; RANITO, 2004). Ter informação passou a ser um dos recursos mais importantes para as organizações, pois ela garante não somente o cumprimento de metas, mas também um posicionamento estratégico favorável perante seus concorrentes.

Independente da sua natureza, toda organização precisa se abastecer de informações para poder fundamentar e seguir com suas atividades, com isso há a necessidade de dedicar maiores esforços visando construir uma infraestrutura que permita o recolhimento, armazenamento, tratamento, representação e distribuição adequada destas informações. Os prejuízos advindos de decisões baseadas em informações não absolutamente entendidas podem resultar em perda irreparáveis, e, características como precisão, coerência e facilidade de compreensão da informação passam a ser indispensáveis para a sobrevivência das empresas (GOUVEIA; RANITO, 2004).

Uma informação de qualidade trás consigo aspectos importantes como objetividade, confiabilidade, relevância e interpretabilidade. Essas características somadas classificam a informação como válida, ativa e de fato pode ser utilizada pelo usuário ou organização. No entanto, ter posse de informação que contenha todos estes predicados não está ao alcance de muitas organizações, ou por consequência do volume de dados que são armazenadas sem o adequado cuidado ou pela falta de preparo das organizações em explorar esses dados (NEVES, 2006).

A busca de padrões, tendências, problemas ou características de um determinado setor da empresa ou do ambiente mercadológico motivou a criação de ferramentas que explorassem concentrações de informações a fim de auxiliar na identificação de conhecimento útil que possa estar contida nelas. A obtenção das informações, de maneira ágil e adequada pode ser conseguida com a implementação de um Sistema de Informações (SI), um instrumento administrativo muito importante para a otimização da comunicação e o processo decisório em uma empresa (FARIA; LAUDELINO; DOMINGUES, 2006).

Dentro das organizações um Sistema de Informação tem por objetivo orientar a tomada de decisão, além de garantir informações seguras e protegidas e permitir a sua disponibilidade para demais componentes da organização. A eficiência de um SI é aferida pela forma como cumpre os objetivos definidos e a capacidade de fornecer dados e informação à organização em tempo e custo adequado (GOUVEIA; RANITO, 2004). O aumento na participação de mercado, redução de custos de produção, desenvolvimento de novos produtos são resultados que somente são alcançados se o SI adotado for projetado de acordo com as estratégias da organização (REZENDE; ABREU, 2000).

Os Sistemas de Informação evoluíram de forma extraordinária, sendo visivelmente percebidos nas indústrias e nos ambiente empresariais em geral através dos diversos *softwares* utilizados para o controle e gerenciamento das atividades internas e externas (FONSECA; GARCIA, 2007).

Como principais características de um Sistema de Informação eficiente podem ser destacados as seguintes (REZENDE; ABREU, 2000):

- a) Segurança ao trabalhar com grandes volumes de dados.
- b) Melhoria da qualidade.
- c) Aumento da produtividade e rentabilidade.
- d) Vantagem competitiva.
- e) Aperfeiçoamento de tecnologias e procedimentos técnicos.

- f) Ampla abrangência e dinamicidade.
- g) Controle de processamento complexos.

O objetivo da elaboração de um Sistema de Informação é de criar um cenário de informações confiáveis. Desta forma, um bom sistema garante informações disponíveis, de qualidade e úteis para tomada de decisão. Mas para que essas características ocorram é preciso que o conhecimento de Sistemas de Informação envolva (BATISTA, 2006):

- a) Domínio das tecnologias da informação.
- b) Compreensão do relacionamento da organização com seus funcionários e parceiros.
- c) Capacidade de solucionar problemas.

Um Sistema de Informação possui três atividades básicas: a entrada (*input*), que envolve a captação ou coleta de fontes de dados brutos dentro da organização ou em seu ambiente externo; o processamento, que envolve a conversão dessa entrada bruta em uma forma mais útil e apropriada; e a saída (*output*), que permite a transferência da informação processada às pessoas ou atividades que a usarão (MARTINS JUNIOR et al., 2012).

As tecnologias desenvolvidas para coleta, armazenamento e apresentação de informações vêm evoluindo e disponibilizando técnicas, métodos e ferramentas computacionais automatizadas, capazes de auxiliar na extração e exibição de conhecimento útil contido em grandes volumes de informações, dentre essas tecnologias vêm ganhando destaque os Sistemas de Mineração de Dados, cuja utilização representa oportunidades de ações estratégicas em *marketing*, produção e inovação, dentro do mercado onde as organizações estão inseridas ou avançar em novos mercados (BURNHAM et al., 2005; PENTEADO; CARVALHO, 2009).

Estes sistemas são explicados na próxima subseção.

2.1.1 Sistemas de Mineração de Dados

O crescimento do volume de registros tem gerado uma urgente necessidade de novas técnicas e ferramentas capazes de transformar, de forma inteligente e automática, *terabytes* de registros em informações significativas e em conhecimento. Essas informações de grande valia para o planejamento, gestão e tomadas de decisão, estão, na verdade, implícitas e/ou escondidas sob uma montanha de registros e não podem ser descobertas ou, no mínimo,

facilmente identificadas utilizando-se sistemas convencionais de gerenciamento de banco de dados (SFERRA; CORREA, 2003).

Os registros gerados pelas organizações de médio e grande porte superam a capacidade humana de interpretar, analisar e compreender tanta informação. Por isso, são necessários novos sistemas capazes de analisar automaticamente o volume de registros produzidos, fornecendo o conhecimento para auxiliar nos processos decisórios. Conseqüentemente, as grandes empresas estão começando a explorar as possibilidades oferecidas pelas diversas técnicas e ferramentas atualmente disponíveis para aprimorar o processo de tomada de decisão (BARBOSA; MACHADO, 2007).

Surgem, então, as ferramentas e técnicas de Mineração de Dados, que estão sendo cada vez mais empregadas em organizações, pois oferecem de forma rápida e automatizada uma alternativa para a geração de informações e produção do conhecimento, identificando aspectos relevantes que possam ser utilizados em nível estratégico como apoio ao processo de tomada de decisão (SCHUCH et al., 2010).

Para Rodrigues Filho (2001), Mineração de Dados é um sistema usado para revelar informação estratégica escondida em grandes massas de registros. Ela é utilizada em diversas áreas, como análise de riscos, *marketing* direcionado, controle de qualidade, análise de registros científicos, entre outros.

De modo geral, primeiramente é feita a escolha das fontes de registros a serem utilizadas e a definição dos objetivos. Uma parcela desses registros é selecionada, pré-processada e submetida a métodos de exploração contendo algoritmo específicos de mineração e ferramentas adequadas com o objetivo de encontrar padrões ou modelos que representem o conhecimento obtido. Depois de extraídos, os padrões são pós-processados e o conhecimento adquirido é avaliado quanto a sua qualidade para determinar a viabilidade de sua utilização no apoio a algum processo de tomada de decisão (YAMAGUCHI, 2010).

As funções da Mineração de Dados podem ser divididas em dois focos: preditivos e descritivos. A predição, consiste em utilizar-se de um grupo de registros que reportam características conhecidas e, a partir dela, construir-se um modelo que poderá ser utilizado para prever as mesmas características em novos casos ou situações. Já a descrição, pode ser vista como uma aproximação sobre registros nos quais não há uma característica de interesse e deseja-se ter somente uma ideia de como esses registros se agrupam segundo variáveis especificadas (RODRIGUES FILHO, 2001).

Para atingir seu objetivo, o Sistema de Mineração de Dados usa diversas técnicas disponíveis para análise de registros em seu sentido mais amplo (RODRIGUES FILHO, 2001). A escolha das técnicas de mineração de dados dependerá da tarefa específica a ser executada e dos dados disponíveis para análise, devendo-se levar em conta a natureza dos registros disponíveis em termos de conteúdo e a estrutura das relações entre os registros (SCHUCH et al., 2010).

Conforme Castanheira (2008), o conhecimento descoberto durante a fase de mineração de dados pode ser conduzido de acordo com seis técnicas:

- a) Classificação: características são pré-definidas pelo usuário e o conjunto de informações é analisado pelos algoritmos de mineração em busca de semelhanças às descrições de cada característica (CARDOSO; MACHADO, 2008).
- b) Clusterização: registros são processados em busca de elementos de informação que apresentem semelhanças. A diferença para a classificação é que aqui não há características pré-definidas, são os próprios algoritmos que classificam os elementos em classes distintas (KEIM; WARD, 2002).
- c) Associação: procura encontrar relações entre os atributos de um conjunto de registros de tal forma que a presença de alguns atributos implique na presença de outros. O objetivo com a geração de “regras de associação” é encontrar padrões e tendências em conjuntos de informações (CARDOSO; MACHADO, 2008; PERNOMIAN, 2008).
- d) Predição: baseados em dados históricos, faz-se a projeção do comportamento mais provável (RODRIGUES FILHO, 2001).
- e) Padrões sequenciais: um padrão sequencial é uma expressão da forma $\langle I_1, I_2 \dots I_n \rangle$, em que cada I_i é um conjunto de itens. A ordem em que estão alinhados os conjuntos reflete a cronologia com que aconteceram os fatos representados por eles. Encontrar padrões previsíveis em um período de tempo significa que um comportamento particular em um dado momento pode ter como consequência outro comportamento ou sequência de comportamentos dentro de um mesmo período de tempo (CARDOSO; MACHADO, 2008).
- f) *Outliers*: um banco de dados pode conter informações que não apresentam o comportamento geral da maioria. Eles são denominados *outliers* (exceções) (CARDOSO; MACHADO, 2008).

Nas técnicas de mineração de dados são aplicados algoritmos específicos para extração de padrões e comportamentos dentro de um banco de dados. Algoritmos de

mineração de dados são os mecanismos que exploram e criam os modelos de mineração para definir padrões e tendências. A partir destes modelos e padrões são estabelecidos os parâmetros de mineração para a extração de informações dos registros (SCHUCH et al., 2010).

De acordo com Rodrigues Filho (2001), Schuch et al. (2010), Domingues (2004), Romão et al. (1999) e Mafra et al. (2008), os principais algoritmos de mineração de dados são:

- a) De classificação: os algoritmos mais comuns são os que geram árvores de decisão. As árvores de decisão constroem uma estrutura de relacionamentos na forma de árvore onde folhas representam classificações e ramos conjunções de características que levam às classificações. Um exemplo de algoritmo de classificação por árvores de decisão é o algoritmo *J48*; com este algoritmo cada ‘nó-pai’ da árvore representa uma condição, envolvendo um atributo e um conjunto de valores. Enquanto os ‘nós-filho’, correspondem às conclusões que indicam a atribuição de um valor ou conjunto de valores a um atributo do problema (RODRIGUES FILHO, 2001).
- b) De clusterização: esses algoritmos dividem os registros em grupos, baseados em semelhanças encontradas pelos próprios algoritmos nos registros, explicitando relações que não podem ser notadas antes. Um algoritmo de agrupamento é o *K-Means*, o qual utiliza o conceito de centróides como protótipos representativos dos grupos, onde o centróide representa o centro de um grupo, sendo calculado pela média de todos os objetos do grupo. O objetivo deste algoritmo é encontrar a melhor divisão de P dados em K grupos distintos entre si, de maneira que a semelhança entre os dados de um grupo e o seu respectivo centro, seja maior e mínima se comparado com dados dos outros grupos (MAFRA et al., 2008).
- c) De associação: os algoritmos de regras de associação efetuam varreduras na base de dados definindo indicações quanto à presença de um conjunto de variáveis dependente de outro conjunto distinto nos registros (SCHUCH et al, 2010). Um exemplo de algoritmo de associação é o algoritmo *Apriori* que encontra todos os conjuntos de itens frequentes contidos numa base de dados. Esse algoritmo gera um conjunto de k -*itemsets* candidatos e então percorre a base de dados para determinar se os mesmos são frequentes, identificando desse modo todos os k -*itemsets* frequentes (DOMINGUES, 2004; ROMÃO et al, 1999).

- d) De Predição, Padrões sequenciais e *Outliers*: os algoritmos explicitados anteriormente podem ser utilizados também para prever comportamentos, sequências e excessões no conjunto de dados.

Dentre as dificuldades enfrentadas no processo de mineração de dados está a manipulação de grandes volumes de dados. Então, para auxiliar nesta tarefa foram criadas ferramentas operacionais que executam a exploração de dados de forma automática e fornecem o resultado (GOLDSCHMIDT; PASSO, 2005).

Essas ferramentas são utilizadas para prever tendências e comportamentos futuros, permitindo aos tomadores de decisões basearem-se em fatos históricos e não em suposições.

A seguir será feita uma breve apresentação das ferramentas para a mineração de dados mais populares:

a) *Clementine*

É uma ferramenta de mineração de dados que permite que se desenvolvam rapidamente modelos preditivos utilizando conhecimento de negócio e implante-os em operações para melhorar a tomada de decisão. Organizações utilizam o conhecimento adquirido a partir de *Clementine* para manter clientes rentáveis, atrair novos clientes, detectar fraudes, reduzir o risco e melhorar a prestação de serviços (SPSS, 2007).

b) *PolyAnalyst*

É uma ferramenta multi-estratégia para descoberta de conhecimento em grandes volumes de texto e dados estruturados. O sistema permite responder perguntas de negócios diversificados, aprendendo a partir de dados históricos prever os resultados de situações futuras. Usando uma interface simples de arrastar e soltar, pode-se facilmente desenvolver diversos cenários de análise de dados (MEGAPUTER, 2007).

c) *Waikato Environment for Knowledge Analysis (Weka)*

É um programa desenvolvido na Universidade de Waikato, na Nova Zelândia e seu nome significa Ambiente para Análise do Conhecimento Waikato. O sistema é escrito em Java e distribuído sob os termos da GNU General Public License que lhe classifica como de distribuição gratuita. Rodando em qualquer plataforma, fornece uma interface uniforme em todas as etapas de sua utilização (WITTEN; FRANK, 2005).

Weka tem a finalidade de identificar a informação a partir de dados brutos. O software trabalha com arquivos específicos e seu enfoque principal está em ser um classificador com os algoritmos de filtro. Porém, também inclui implementação de algoritmos

para aprender com associação de regras, como também agrupar dados para os quais nenhum valor de classe é especificado (MORAES; BASTOS; BITTENCOURT, 2002).

O *Weka* permite sua utilização em quatro ambientes de exploração: *Explorer*, *Experimenter*, *Knowledge flow* e *Simple CLI*. De acordo com Witten, Frank (2005) a maneira mais fácil de usar *Weka* é através da interface gráfica chamada de *Explorer*, por que dá acesso a todas as suas instalações, utilizando-se de menu e preenchimento de formulários e se mostra melhor para usuários iniciantes.

d) *IBM Intelligent Miner*

O *Intelligent Miner* dá suporte às tradicionais técnicas de mineração de dados, tais como: análise de agrupamentos, análise de afinidades, classificação, estimativa e previsão. Adicionalmente, componentes de apresentação estão disponíveis para possibilitar uma análise visual dos resultados (GONÇALVES et al., 2011).

e) *WizRule*

Criado pela empresa *WizSoft*, o *WizRule* foi desenvolvido pra examinar e decrever conjuntos de dados, detectando erros dentre os dados analisados. Se mostra útil na detecção de desvios por meio da extração de regras tipo se X então Y (GOLDSCHMIDT; PASSOS, 2005).

f) *SAS Enterprise Miner*

Este *software* de mineração permite a descoberta de relações, tendências e padrões nos dados; para as organizações se apresenta como um meio de ganhar vantagem competitiva, possibilitando compreender o comportamento dos clientes e reconhecimento de oportunidades de negócio. Sua interface gráfica interativa mostra todo o processo de mineração de dados dando acesso ao usuário na eliminação de variáveis, adição de conjuntos de dados e filtragem de erros (SAS, 2012).

g) *Oracle Data Mining*

É um *software* de mineração de dados que opera integrado ao gerenciador de banco de dados da *Oracle* (GOLDSCHMIDT; PASSOS, 2005).

O Quadro 1 apresenta um resumo das ferramentas de mineração de dados, com as tarefas que os algoritmos executam, a forma como apresentam seus resultados e o fabricante.

Nome da ferramenta	Tarefas realizadas	Apresentação dos resultados	Fabricante	Licença de uso
SPSS/ClementinePoly	Classificação, regras de Associação, Clusterização, Sequências, Detecção de Desvios	Tabela e Texto	SPSS Inc.	Comercial
PolyAnalyst	Classificação, Regressão, Regras de Associação, Clusterização, Sumarização, Detecção de Desvios	Textos e Gráficos	Magaputer Intelligence	Comercial
Weka	Classificação, Clusterização, Regras de Associação	Conjunto de Regras e Árvore de Decisão	University of Waikato	Gratuito
Intelligent Miner	Classificação, Regras de Associação, Sequências, Clusterização, Sumarização	Gráficos, Tabelas, Árvores e Clusters	IBM Corp.	Comercial
WizRule	Sumarização, Classificação, Detecção de Desvios	Planilha e Gráficos	WizSoft Inc.	Gratuito
SAS Enterprise Miner	Classificação, Regras de Associação, Regressão, Sumarização	Gráficos, Textos e Árvores	SAS Inc.	Comercial
Oracle Data mining	Classificação, Regressão, Associação, Clusterização	Gráficos, Tabelas e Histogramas	Oracle	Comercial

Quadro 1 – Exemplos de ferramentas de mineração de dados

Fonte: Adaptado de Goldschmidt e Passos (2005)

Quando se trata de apresentar informações armazenadas em bases de dados, deve-se pensar no uso dos recursos computacionais para enriquecer a interpretação de informações por parte dos usuários.

Técnicas de mineração podem apoiar-se na exploração visual de grandes conjuntos de registros, e também serem apoiadas por recursos visuais. A integração de técnicas de mineração de dados com técnicas de visualização de informações facilita a análise das informações tornando-se um valioso recurso na tomada de decisões, de forma que a combinação pode ser usada para agilizar a compreensão do conhecimento (NETO et al., 2010).

As formas de representação gráfica de informações serão detalhadas na próxima seção.

2.2 REPRESENTAÇÃO VISUAL DE INFORMAÇÕES

Existem duas formas para representar os resultados de um problema: a gráfica e a não gráfica. Informações de pequeno porte são mais fáceis de serem absorvidas, porém quando envolvem grandes volumes tornam-se difíceis de ser compreendidas, situação que

pode-se agravar quando sua disponibilidade é apenas em formato textual (DIAS; CARVALHO, 2007).

A tecnologia da informação permitiu a construção de estruturas visuais que demonstrassem melhor o fluxo de informações e fornecesse um significado mais amplo a informação. O papel das estruturas de visualização de informações é o de acelerar a apropriação de algum conhecimento, ou seja, de aumentar a velocidade da transferência do conhecimento por meio de estruturas gráficas, e quanto maior for a complexidade do conhecimento a ser transmitido, maior será o grau de interação da estrutura de visualização com seu observador (DIAS, 2007).

Informações representadas graficamente tendem a ser processadas de maneira mais automática pela visão, em um processo mais superficial, rápido e de capacidade elevada. Dessa forma, representar graficamente os dados a serem analisados é interessante do ponto de vista da obtenção de informação, pois faz com que não apenas mecanismos computacionais sejam usados para a análise de dados, mas também recursos da visão e da cognição humana (SILVA, 2007).

Conceitualmente, entende-se por visualização a transformação de informação em representações gráficas com o objetivo de tornar essas informações mais inteligíveis para a mente humana. Ela se apoia na faculdade da visão humana de captar de forma mais rápida a informação mostrada quando apresentada graficamente, usando cores, formas diferenciadas do que quando apenas em formato texto (NETO et al., 2008).

A Visualização de Informações é um campo que estuda o uso de representações visuais de dados abstratos, aproveitando-se da capacidade da percepção visual humana, para detectar de maneira mais rápida, padrões e tendências que de outra forma seriam dificilmente percebidos pelo usuário (DARIO, 2010). Dias (2007) define Visualização de Informações como uma área da ciência dedicada ao estudo da apresentação de informações por meio de recursos gráficos com a finalidade de facilitar seu entendimento e auxiliar na dedução de novos conhecimentos.

Gomes e Tavares (2011) apontam como vantagens na utilização de visualização de informação:

- a) A imediata percepção da informação importante.
- b) As características das informações são realçadas visualmente.
- c) A indicação de padrões e relações entre informações.
- d) Informações incorretas são facilmente apontadas.

Para que uma representação visual seja considerada adequada, ela necessita ser intuitiva, de modo que o observador possa entender as informações e seu fluxo sem nenhuma dificuldade. As estruturas de visualização precisam ser apropriadas para o tipo de informação a ser apresentada, pois elas têm papel importante na formulação de hipóteses e podem contribuir para a tomada de decisão (ESTIVALET, 2000).

São várias as maneiras de se representar uma informação graficamente, a área de Visualização de Informações possui diversas técnicas que podem ser utilizadas individualmente ou combinadas. As chamadas técnicas de visualização procuram otimizar a capacidade visual do ser humano facilitando o processo de entendimento do conhecimento a partir das estruturas apresentadas.

Alguns tipos de representação visual de informações são (YAMAGUCHI, 2010):

- a) Diagramas: representa esquemas que categorizam as informações, reduzindo a complexidade sobre elas, de modo que seja possível visualizar de forma mais clara a relação entre os seus valores, aumentando a compreensão sobre o significado.
- b) Imagens: podem ser utilizadas para descrever metáforas visuais, que representam uma nova ideia a partir de elementos familiares ao receptor, como elementos da natureza, objetos físicos, atividades ou conceitos abstratos, utilizados com um significado diferente para expressar um conceito.
- c) Mapas: seguem as convenções da cartografia para a construção da visualização que indique um conhecimento, como a delimitação territorial de uma cidade, por exemplo.
- d) Objetos: podem ser utilizados para explorar o conhecimento de uma composição espacial sob diferentes pontos de vista.

Além desses citados por Yamaguchi (2010), outro recurso de visualização de informações é o Fluxograma, representação gráfica de etapas sequenciais que envolvem um processo. Tem a função de mostrar as direções que devem ser seguidas na realização de atividades ou no gerenciamento de processos. Com o uso de fluxograma, é possível representar conhecimentos sob a forma lógica para a resolução de problemas. É um recurso útil dentro no processo de tomada de decisão, pois, de maneira rápida, são realizados testes e é detectado qual problema está ocorrendo. Desse modo, qualquer pessoa, que não seja técnica no assunto, conseguirá iniciar uma ação corretiva até a resolução da situação (DIAS, 2007; KHAN; KHAN, 2011).

Diferentes técnicas de Visualização de Informações utilizam recursos computacionais para representar informações de maneira gráfica e interativa, procurando

otimizar o uso das capacidades visuais humanas para compreender fenômenos que não possuem em si mesmos uma representação espacial própria. Perceber informação relevante exige que essa seja apresentada de maneira clara e não ambígua às pessoas envolvidas na análise, de modo que elas possam interpretá-las corretamente e abstrair conhecimento relevante a ser posteriormente compartilhado e disseminado (SILVA, 2007).

Nascimento e Ferreira (2005) afirmam que para uma técnica de visualização de informações ser considerada útil é preciso que ela atenda a dois parâmetros: a expressividade e a efetividade. A expressividade se caracteriza pela capacidade da visualização de expor as informações relevantes em sua totalidade. Já na efetividade o destaque está na facilidade de compreensão das informações expostas, inibindo erros de interpretação. Esses são dois aspectos que necessariamente devem ser atendidos por qualquer sistema de visualização de informação, do contrário a eficácia estará comprometida.

Na próxima subseção serão descritas algumas técnicas que se utilizam da união de imagens e objetos para a exibição de relacionamentos e tendências entre informações.

2.2.1 Técnicas de Visualização de Informações

O uso de técnicas de visualização da informação pode minimizar a dificuldade enfrentada em notar resultados relevantes, pois proporciona uma interface que permite a busca, localização e visualização da informação de seu interesse, viabilizando a possibilidade de compreensão de como está organizada a informação, oferecendo uma visão geral dos assuntos tratados, e contextualizando o resultado de forma a satisfazer a necessidade informacional do usuário (VIEIRA; CORRÊA, 2010).

Existem muitas técnicas de visualização de informações, as mais utilizadas estão divididas em quatro categorias (NETO, 2008).

As técnicas de visualização geométricas: entre as técnicas desta categoria estão as Coordenadas Paralelas e Gráfico de Dispersão (KEIM; WARD, 2002; VALIATI, 2008).

- Técnica Coordenadas Paralelas

Nessa técnica representa-se informações utilizando linhas verticais e horizontais. Cada linha vertical indica uma dimensão ou atributo e os valores das dimensões são mapeados para pontos nessas linhas verticais, de forma que cada item de informação seja representado

como uma linha poligonal que intercepta cada linha vertical em seu ponto correspondente ao valor de cada atributo, formando um aglomerado de linhas horizontais (VALIATI, 2008).

- Técnica Gráfico de Dispersão

É uma técnica onde pontos são plotados em coordenadas (x, y) semelhante a outros gráficos bidimensionais. Na extensão desta técnica, conhecida como Matriz de Dispersão, vários Gráficos de Dispersão são utilizados para representar todo o conjunto de informações com seus atributos, cada gráfico individual mostra cada parte isolada da informação sendo identificadas nas linhas e colunas do gráfico (RABELO, 2007).

As técnicas de visualização orientadas a pixel: esta categoria baseia-se no mapeamento de um atributo de informação para um pixel colorido do dispositivo de exibição, proporcionando a análise simultânea do maior número de informações possível. Para cada atributo dos itens de informação, uma imagem pode ser construída e exibida em uma janela separada. A técnica se preocupa em como os pixels serão organizados na tela, otimizando a percepção do usuário em relação ao conjunto de informações (BOSCARIOLI; TABUSADANI; BIDARRA, 2008).

As técnicas de visualização iconográficas: que têm como componentes as Faces de Chernoff, Glifo em Estrela e Figura de Arestas.

- Técnica Faces de Chernoff

Uma das mais conhecidas visualizações desta categoria, onde as dimensões das informações são mapeadas para características da face humana, como formato do nariz, boca, olhos e da face como um todo, resultando em diferentes formatos segundo os valores mapeados. Contudo, a validade desta técnica é amplamente discutida devido a dificuldade de percepção de diferenças muito pequenas nas imagens (RABELO, 2007).

- Técnica Glifo em Estrela

Outra visualização bastante utilizada, onde as dimensões são representadas como raios de ângulos iguais partindo do centro de um círculo. O comprimento dos raios indica o valor de cada atributo, raios mais extensos representam valores máximos e o centro do círculo indica o valor mínimo da dimensão. É eficiente na representação de conjuntos de subdivisões ou partes de informação, permite a exibição de mais subdivisões de informação do que a técnica de *Faces de Chernoff*, mas possui como a outra técnica, um limite para a exibição dessas partes (RABELO, 2007).

- Técnica Figura de Arestas

Caracteriza-se por ser um ícone básico, onde várias dimensões das informações são mapeadas para parâmetros do ícone, tais como o comprimento e o ângulo de rotação de uma aresta ou galho. Variações de comprimento, espessura e cor das arestas fornecem outras possibilidades de representação, explorando a capacidade humana na percepção de texturas e interpretação de padrões (VALIATI, 2008; KEIM; WARD, 2002).

As técnicas de visualização hierárquicas: nesta categoria os atributos e dimensões da informação são apresentados na área de exibição dividida em áreas de tamanhos dependentes da importância e localização na hierarquia da informação. *Perspective Wall*, *Cone Tree*, *TreeMap*, *Cheops* e *Hyperbolic Tree* são exemplos desta categoria (KEIM; WARD, 2002).

A seguir são descritas técnicas pertencentes a esta categoria.

- Técnica *Bifocal Display*

Usa a distorção bidimensional para exibir informações. Esta técnica possui três regiões diferentes para distribuir as informações. Na região central, de tamanho maior, encontra-se as informações de interesse, enquanto que nas laterais estão contidas as demais informações que dão contexto ao que está em foco (NETO, 2008; GOMES; TAVARES, 2011).

- Técnica *Flip Zooming*

Esta técnica divide a tela em pequenas regiões retangulares de tamanho igual para representar as informações de contexto, já a informação de interesse é representada no centro da tela por um retângulo maior. As informações de contexto são distribuídas em sequência ao redor da informação principal. Caso as informações não necessitem de uma sequência lógica, uma ordem poderá ser estabelecida pelo observador. Tanto as regiões menores como a maior são exibidas frontalmente, permitindo uma visualização geral do todo (LUZZARDI, 2003).

- Técnica *Perspective Wall*

Nessa técnica, a área exibida é representada como um retângulo horizontal longo. A ideia é desenhar a informação no retângulo e, então, dobrar o retângulo para trás nos limites à direita e à esquerda. Uma perspectiva tridimensional do resultado é gerada. Desta forma, uma seção detalhada pode ser visualizada em primeiro plano enquanto o resto da estrutura aparece em segundo plano. Quando um usuário seleciona um objeto da parede, o item é movido para o centro da parede usando animação. Na *Perspective Wall* ocorre a criação de uma visualização que provê uma visão detalhada de uma região de interesse, enquanto o contexto global das estruturas adjacentes é mantido. Tem como vantagem manter sempre o contexto, fazendo com

que o usuário não perca a localização das informações (LUZZARDI, 2003; ESTIVALET, 2000).

- Técnica *Information Cube*

As informações são exibidas, num espaço tridimensional, como cubos translúcidos encaixados um dentro do outro, onde o mais externo representa o nó raiz da hierarquia. Os cubos que denotam o segundo nível estão localizados no interior e assim por diante. Títulos são exibidos na superfície dos cubos representando as informações contidas nos nós. O aspecto transparente dos cubos permite que o usuário possa visualizar seu interior sem ser atrapalhado por muitos detalhes. Pela variação do nível de transparência, o sistema pode controlar a complexidade da visualização, além de que distinguir facilmente a estrutura hierárquica da informação nos cubos, e as superfícies semi-transparentes separam claramente os diferentes níveis de informação (LUZZARDI, 2003; ESTIVALET, 2000).

- Técnica *Cone Tree*

A técnica representa informações hierárquicas onde o elemento raiz de uma árvore, representado por um retângulo, é localizado no local mais alto de um cone transparente e todos os seus filhos ficam dispostos na base circular do cone. Os cones apresentam a mesma altura para cada nível da árvore mas os seus diâmetros são reduzidos de um nível para outro, de modo a que toda a estrutura seja visível na área disponível da tela de exibição (GOMES; TAVARES, 2011).

- Técnica *TreeMap*

Divide o espaço disponível para exibição em regiões e cada região é dividida novamente para cada nível na hierarquia. Esta técnica utiliza o conceito de preenchimento de espaço, isto significa que o espaço inteiro de exibição para visualização é dividido em nós. Cada nó da hierarquia é representada por um retângulo. Os filhos de um nó interior são desenhados dentro do retângulo do nó-pai. O tamanho do retângulo desenhado depende do peso do nó representado. Assim, nós que representam informações de maior peso possuem mais espaço na área de exibição (NETO, 2008; PUTZ, 2005).

- Técnica *Cheops*

Esta técnica mantém uma visão geral da estrutura informacional e proporciona um acesso fácil aos detalhes por meio de uma visualização comprimida dos componentes da informação, onde alguns de seus nós são representados com um mesmo componente visual, um triângulo. Desta forma, numa mesma área de tela pode-se exibir uma quantidade de

informação muito maior se comparada aos métodos tradicionais de visualização (LUZZARDI, 2003).

- Técnica *Information Slices*

Utiliza um ou mais discos semicirculares para visualizar hierarquias com vários níveis em duas dimensões de forma compacta. Assim, cada disco representa uma hierarquia com múltiplos níveis. Em cada nível da hierarquia, os filhos são dispostos no espaço de acordo com o seu tamanho. Hierarquias mais profundas são representadas usando uma série de discos em cascata (LUZZARDI, 2003).

Cores e tamanho são usados para indicar o tipo de nível hierárquico e a natureza da informação representada. Além disso, uma fatia do disco semicircular é expandida para uma área à direita da primeira, também como um semicírculo. Para exibir mais do que dois níveis de uma hierarquia, os de ordem superior são representados por ícones (GOMES; TAVARES, 2011; ANDREWS; HEIDEGGER, 1998).

- Técnica *Bifocal Tree*

Permite uma melhor percepção da informação, pois divide o conteúdo em duas regiões integradas: uma região de contexto e outra de detalhe. À direita, é exibida de forma detalhada a subárvore cuja *raiz* foi selecionada pelo usuário. À esquerda, fornece uma visão global da estrutura. Cada área tendo o seu próprio nó foco. Quando um nó é selecionado, este passa a ser o foco na área de detalhe e o seu nó-pai torna-se o foco na área de contexto. Desta forma, a técnica permite ao usuário uma visão detalhada das relações da subárvore que contém o nó de interesse sem que a percepção da hierarquia completa seja perdida (CAVA; LUZZARDI; FREITAS, 2002).

- Técnica Grafos

Representam relacionamentos entre instâncias de uma classe, sendo cada nó a representação de uma instância, identificada pelo nome, e cada aresta o relacionamento entre as instâncias. Da maneira como a estrutura é apresentada, são exibidos os nós mais importantes, que são os que possuem maior número de relacionamentos ou as palavras-chave. Nas organizações, esse tipo de estrutura é utilizado para mapear os fatores críticos de sucesso, classificando os nós em pais e filhos (DIAS, 2007).

- Técnica Olho de Peixe

Essa técnica apresenta uma característica interessante a respeito do sentido de visão dos usuários; pois, quando o mouse se movimenta no sentido vertical, de acordo com o deslocamento do seu ponteiro, as informações modificam de tamanho (DIAS, 2007).

Todas estas técnicas não precisam ser utilizadas isoladas uma das outras, poderá haver situações onde a união de características de cada categoria de visualização seja necessária, para aplicação em uma visualização, de forma a aumentar a captação de conhecimento da informação que está sendo alvo de análise.

Gerada a representação visual das informações, uma apresentação estática, pode não contemplar todo o significado da mensagem a ser transmitida, impedindo a compreensão da informação pelo usuário. Para que haja melhor exploração visual, são disponibilizadas funções aos usuários as quais possibilitam que sejam alteradas as representações de acordo com suas necessidades. Essas funções podem ser referenciadas como operações de interação ou distorção na visualização de informação e podem ser utilizadas em diferentes níveis de ação, como mostrar apenas a região de interesse, selecionada pelo usuário, como também expandir ou reduzir a quantidade de informações a ser exibida (ESTIVALET, 2000; PERNOMIAN, 2008).

2.2.2 Níveis de Interação do Usuário

Freitas et al. (2001) indica diferentes níveis de interação do usuário com a representação visual. Para ele, num nível mais básico estão as funções de navegação e reposicionamento do observador, quer pelo deslocamento horizontal e vertical no caso de uma representação plana, como pelo deslocamento de uma câmera virtual no espaço, no caso de uma representação tridimensional. Num segundo nível, estão funções de seleção de partes da informação que podem provocar um reposicionamento para uma melhor visualização ou um detalhamento de um componente da informação ou a supressão de outro.

Nestas operações podem ocorrer três situações (FREITAS et al., 2001):

- a) A representação é alterada para mostrar apenas a região de interesse selecionada.
- b) A região de interesse ocupa o campo de visão principal e o restante do conjunto de dados é mantido em área à parte.
- c) A região de interesse e a visão geral são exibidas concomitantemente.

Existem várias operações de interação e a escolha destas ações irá depender da técnica de visualização adotada. Nenhuma delas deverá prejudicar o sentido de orientação do

usuário no conjunto de informações, nem dificultar a exploração do conteúdo da informação (DARIO, 2010). As operações são:

- a) Visão geral: quando todo o conteúdo informacional é exibido em uma mesma estrutura. Aqui exige-se que alguns detalhes sejam suprimidos, relacionamentos mais importantes são mantidos, de forma que o usuário possa se localizar e evitar pesquisas desnecessárias (DIAS, 2007).
- b) Visão geral+detalhe: permite que a região de interesse do usuário ocupe o campo de visão principal e o restante do conjunto de informações fique em uma área separada (ESTIVALET, 2000).
- c) Visão foco+contexto: quando a região de interesse é exibida em conjunto com o restante da informação. Essa técnica é útil quando a pessoa quer uma visão mais detalhada de uma determinada região mas não deseja perder o contexto da visualização dos arredores dessa região de interesse (ESTIVALET, 2000).
- d) Filtragem: permite aos usuários ajustar dinamicamente a quantidade de informações a serem exibidas, para concentrar-se na informação de interesse (KHAN; KHAN, 2011).
- e) *Zoom*: o usuário pode aproximar aumentando o nível de detalhes da região de interesse. Tarefa utilizada quando há a necessidade de se obter um detalhamento maior ou focalizar uma determinada parte da informação (ROMANI, 2000).

As operações citadas, devem também permitir a visualização das relações entre informações, oferecer condições para desfazer e repetir ações, possibilitar o agrupamento de informações relacionadas pelo usuário e possuir as opções de armazenar e extrair os resultados de consultas realizadas (DARIO, 2010).

2.3 TRABALHOS RELACIONADOS

Diferentes autores reúnem diferentes motivos para a escolha de uma técnica em detrimento de outra para representar informações.

Luzzardi (2003), utilizou o conceito de usabilidade para avaliar as características das técnicas de visualização de informações do tipo hierárquica. Para isso estabeleceu critérios próprios e dividiu sua análise em dois tipos: a primeira avaliava a estrutura de apresentação das técnicas, na segunda avaliou os mecanismos de interação presentes em cada uma das técnicas.

O resultado desta análise foi comparado a critérios de usabilidade tradicionais. Para o autor, apesar da evolução na pesquisa sobre ferramentas de avaliação da visualização gráfica de informações ainda não havia um conjunto de parâmetros suficientemente específico para técnicas de visualização de informações. Ao final, o autor concluiu que seus critérios são mais eficientes na identificação de problemas de usabilidade em técnicas de visualização de informações hierárquicas do que outros conjuntos utilizados.

Na análise procedida por Yamaguchi (2010), avaliou-se as técnicas de visualização de informações existentes, com base nos dados a serem representados. Esses dados foram classificados em uni, bi ou tridimensionais, além disso o tipo de tarefa realizada pelo usuário foi analisado, o volume de informações apresentadas e a posição dos atributos das informações no gráfico também. Desta forma, as representações visuais foram divididas em gráficos de projeção unidimensional, bidimensional e tridimensional; técnicas iconográficas; técnicas geométricas, técnicas orientadas a pixel e técnicas hierárquicas ou baseada em grafos.

Rabelo (2007), observou que o potencial de uma técnica de visualização está ligada a sete características, sendo:

- a) A quantidade de informações que a técnica pode exibir.
- b) Se a técnica consegue suportar todas as partes que dão significado a informação (atributos).
- c) A natureza da informação, diferenciando-a em qualitativa e quantitativa.
- d) nível de interação permitido pela representação visual.
- e) A facilidade de perceber o conhecimento exposto.
- f) Os níveis de relacionamento e correlação entre as partes da informação.

Adicionalmente, Dias et al. (2012) identificaram como influências na escolha da técnica, os elementos de visualização fornecida pela técnica e a localização da informação no espaço de apresentação. Essas influências são descritas a seguir:

- a) Elementos da visualização: a visualização fornecida pela técnica para a representação da informação é outro aspecto impactante na escolha de uma técnica de representação visual. Elementos como pixels, ícones, retângulos, divisão de tela, cores e ponto num plano devem funcionar como agentes na transmissão da informação.
- b) Localização da informação: a ordem dos elementos apresentados pode facilitar ou prejudicar a localização das informações, algumas técnicas de visualização permitem ao usuário modificar a ordem de apresentação, sem alterar o significado geral, em

busca de novos padrões nas informações; esta função serve para verificar novas correlações.

Foi necessário conhecer as características intrínsecas das representações visuais fornecidas por essas técnicas e quais parâmetros são avaliados pelos autores citados anteriormente, na busca de levantar as características que poderiam ser compatíveis com a finalidade deste trabalho que é de correlacionar as estruturas das técnicas de visualização com os resultados obtidos pela ferramenta de descoberta de conhecimento *Weka*.

Este capítulo reuniu os conceitos relacionados ao tema abordado nesta pesquisa de modo que o objetivo fosse melhor entendido. Os próximos capítulos trataram do desenvolvimento do trabalho.

3 METODOLOGIA

Este capítulo apresenta a metodologia usada para o desenvolvimento deste trabalho. A seção 3.1 aborda a classificação científica da presente pesquisa. A seção 3.2 descreve o processo de desenvolvimento da pesquisa, com suas respectivas etapas.

3.1 CLASSIFICAÇÃO DA PESQUISA

Essa seção classifica a presente pesquisa pela metodologia científica adotada para seu desenvolvimento. Para a classificação foi levado em consideração a natureza, o objetivo, o problema abordado e os métodos aplicados nesta pesquisa. Desta forma, este trabalho é classificado como (TURRONI; MELLO, 2012):

- Quanto a sua natureza, a pesquisa é aplicada porque seu resultado trata da demonstração das diferentes maneiras de se representar o conhecimento obtido nos resultados da ferramenta de gestão de informações *Weka*.
- Quanto aos seus objetivos, classifica-se esta pesquisa como exploratória, já que sua finalidade é correlacionar aspectos importantes presentes nas técnicas de mineração com as técnicas de visualização e utilizar esses aspectos para demonstrar maneiras diferentes de exibir os resultados minerados pela ferramenta de gestão de informações *Weka*.
- Quanto a abordagem ao problema, o trabalho é qualitativo porque faz uma avaliação das características de apresentação dos resultados de mineração da ferramenta *Weka* com o foco em estabelecer uma relação entre o conhecimento gerado pelos algoritmos de mineração com recursos gráficos das técnicas de visualização de informações.
- Do ponto de vista dos métodos adotados, a pesquisa se apresenta como um experimento pois culminará na construção de protótipos utilizando os elementos visuais das técnicas de visualização de informações para a representação do conhecimento extraídos pelos algoritmos e técnicas de mineração na ferramenta *Weka*.

A próxima seção reporta o processo de desenvolvimento utilizado nesta pesquisa.

3.2 PROCESSO DE DESENVOLVIMENTO DA PESQUISA

A presente pesquisa está focada em como apresentar os conhecimentos gerados por técnicas de mineração de dados em uma ferramenta de gestão de informações utilizando-se de técnicas de visualização de informações.

Este trabalho foi desenvolvido ao longo de quatro fases: Escolher a Ferramenta e as Técnicas de Mineração, Analisar as Técnicas de Mineração, Analisar as Técnicas de Visualização, Correlacionar as Técnicas de Mineração e Visualização, ilustradas na Figura 1.

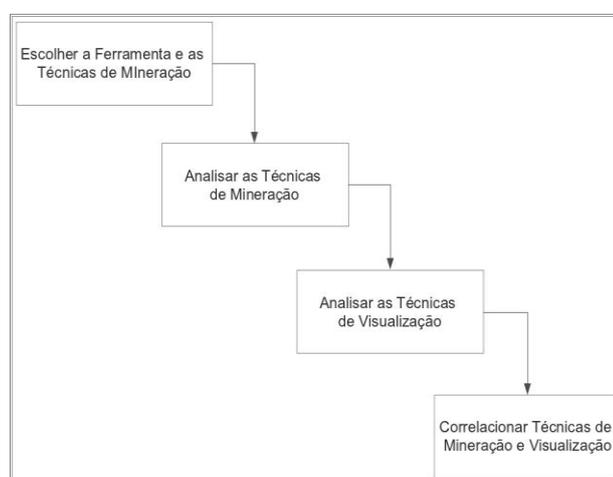


Figura 1 – Processo Geral de Desenvolvimento do Trabalho
Fonte: Autoria Própria (2013)

A primeira etapa, *Escolher a Ferramenta e as Técnicas de Mineração*, consiste em determinar entre as ferramentas e técnicas de mineração existentes na literatura qual será utilizada para a realização do experimento proposto neste trabalho.

A segunda etapa, *Analisar as Técnicas de Mineração*, tem como finalidade identificar as informações e o formato de exposição do conhecimento gerado por cada técnica.

Na próxima etapa, *Analisar as Técnicas de Visualização*, estuda-se as diversas técnicas de visualização com o objetivo de levantar os tipos de informações, a estrutura de visualização e o suporte de informações vinculadas a cada técnica.

Por fim, a última etapa, *Correlacionar Técnicas de Mineração e Visualização*, propõe uma nova forma de como os conhecimentos gerados pelas técnicas de mineração podem ser expostos ao usuário para facilitar sua compreensão durante a análise dos resultados. Desta forma, indicam-se as técnicas de visualização mais adequadas para cada técnica de mineração.

Detalha-se a seguir cada uma dessas etapas.

3.2.1 Escolher a Ferramenta e Técnicas de Mineração de Dados

Nessa etapa, escolheram-se duas ferramentas de gestão de informações, a saber, *Weka* (WEKA, 2012) e *WizRule* (LOH, 2009), as quais são mais usadas no campo da pesquisa científica na exemplificação de modelos de exploração de dados.

A escolha das ferramentas se deu por meio de uma análise obtida pelo trabalho de Rosas (2009) em que se utilizou parâmetros estabelecidos por Scapin e Bastien (1993) com a intenção de avaliar as condições dos *softwares*. Os parâmetros avaliados foram:

- **Convite:** trata das informações que permitem ao usuário identificar o estado em que ele se encontra na interação, as ações alternativas e as ferramentas de ajuda e modo de acesso. Uma interface convidativa é composta por títulos claros para as janelas e caixas de diálogo; informações sobre o estado dos componentes do sistema; sobre o preenchimento de um formulário, com os dados do usuário e opções de ajuda.
- **Legibilidade:** diz respeito à dificuldade ou facilidade na leitura das informações textuais. Tipo de letra, espaçamento entre palavras, parágrafos bem definidos.
- **Agrupamento de Itens:** o posicionamento, a ordenação, a organização e a forma como os objetos, imagens e comandos são dispostos na tela são aspectos importantes para que o usuário possa ter uma rápida compreensão de um programa.
- **Carga de Trabalho:** consiste no tempo que o usuário leva para obter a resposta que precisa. Títulos das telas, comandos, excesso de digitação ou não, necessidade de ir e voltar às janelas para se entender a informação e presença de informações descartáveis.
- **Resposta Imediata:** a resposta sobre a ação executada pelo usuário ou a indicação de que se está “processando a informação” são pontos observados e relacionados com a qualidade do software pelos usuários.
- **Controle do Usuário:** o computador deve somente executar aquilo que o usuário solicitar e no momento que for requisitado, podendo interromper ou cancelar o processo.
- **Adaptabilidade:** para que todos os tipos de usuário tenham direito a usabilidade, o sistema deve apresentar em sua interface maneiras variadas de se realizar a tarefa, permitindo que o usuário adapte à sua vontade ou necessidade.
- **Alerta de Erros:** a gestão de erros diz respeito a todos os mecanismos que permitem

evitar, reduzir e alertar sobre a ocorrência de erros e que favoreçam sua correção.

- **Consistência:** procedimentos, rótulos, comandos são mais facilmente reconhecidos, localizados e utilizados quando o seu formato, localização e nomenclatura manterem um padrão de uma tela ou seção para outra.
- **Significado dos Códigos:** quando o usuário desconhece o significado do código ou denominação, podem levá-lo a interpretações equivocadas e por consequência cometer erros.

Com base nesses critérios Rosas (2009) analisou as duas ferramentas – *Weka* e *WizRule* – por meio de seu uso e de observação. Constatou na primeira análise que dos dez parâmetros estudados, nove estavam presentes nas interfaces gráficas tanto do sistema *Weka* quanto do sistema *WizRule*, sendo que apenas o parâmetro *Convite* tinha sido parcialmente atendido. Porém, uma nova avaliação foi realizada, desta vez considerando-se os conceitos sobre técnicas de visualização de informações e concluiu-se que os itens *Legibilidade* e *Significado dos códigos* se mostram insuficientes para compreensão satisfatória das informações em ambas as ferramentas.

O resultado da segunda análise está ilustrada no Quadro 2.

Critérios	Sistemas de mineração de dados					
	Weka			WizRule		
	Possui	Não possui	Possui parcialmente	Possui	Não possui	Possui parcialmente
Convite			X			X
Legibilidade		X			X	
Agrupamento de itens	X			X		
Carga de trabalho	X			X		
Feedback	X			X		
Controle do usuário	X			X		
Adaptabilidade	X			X		
Alerte de erros	X			X		
Consistência	X			X		
Significado dos códigos		X			X	

Quadro 2 – Sistemas de mineração e os critérios de usabilidade

Fonte: Adaptado de Rosas (2009)

Os resultados obtidos pela segunda análise, especificamente *Legibilidade* e *Significado dos códigos*, tornaram-se o ponto de partida para a realização deste trabalho. Isto porque, para entender a mensagem contida na informação é fundamental que esta seja transmitida o mais completa possível, sendo que o usuário não pode dividir sua atenção tentando desvendar a linguagem utilizada pelo *software* para localizar a informação.

Depois da conclusão desta análise, a implementação das técnicas de mineração, descritas na subseção 2.1.1, levaram a escolha das ferramentas *Weka* e *WisRule*.

A ferramenta *WizRule*, basicamente, processa os registros em busca de informações que revelem a relação Se-Então (*If-Then*). Desta forma, para tornar a análise mais abrangente a ferramenta escolhida foi a *Weka* (versão 3.7.5) por contemplar mais técnicas de mineração, tais como: Classificação, Clusterização e Associação. *Weka* ainda apresenta a opção Seleção de Atributos, no entanto, esta técnica não se tornou foco desta pesquisa, pois ela somente mostra quais são os atributos mais relevantes em uma base de dados, não exibindo regras que ajudam o usuário na tomada de decisão.

Definidas a ferramenta e as técnicas, foi preciso decidir quais algoritmos seriam processadas as bases de dados para o experimento em cada técnica. Para isto, utilizou-se os trabalhos de Schuch et al. (2010); Rodrigues Filho (2001); Domingues (2004); Romão et al. (1999) e Mafra et al. (2008), nos quais citam-se, principalmente, os algoritmos: *J48* para técnica de Classificação, *Apriori* para Associação e *K-means* para a Clusterização. Um estudo mais dedicado será explicitado na próxima subseção.

3.2.2 Analisar as Técnicas de Mineração de Dados

Antes de utilizar qualquer ferramenta de exploração de dados, se faz importante reconhecer que informações são fornecidas pelas técnicas de mineração e de que forma essas informações são apresentadas ao usuário. Por isto, nesta etapa realizou-se um estudo aprofundado sobre cada uma das técnicas e seus respectivos algoritmos, os quais foram determinados na etapa anterior.

Após o entendimento da técnica e de seu algoritmo, identificou-se para eles o “Modelo de Conhecimento” gerado pela ferramenta *Weka* que engloba as características de tarefas e método. As *tarefas* representam os tipos de conhecimento que são extraídos da base de dados; os *métodos* representam as formas de como o conhecimento é apresentado para o usuário.

Assim, as tarefas são:

- Grupos: conjunto de registros que possuem alguma semelhança entre si e diferenças significativas com relação às informações de outros grupos.
- Relacionamento: representam o grau de ligação de um registro com outro; pode apresentar uma relação de hierarquia ou apenas indicar que possuem características em

comum.

- Regras de associação: são informações derivadas dos relacionamentos representados por árvores, redes ou classes.
- Padrões: são usados para unir ou distinguir informações e podem também servir como base para busca de erros ou falhas nos registros. Os itens Grupos, Relacionamento e Regras de associação podem ser formados pela verificação de comportamento repetitivo dentro dos registros.

E os métodos são:

- **Árvore:** demonstra os níveis de hierarquia presentes num conjunto de registro que forma a informação. É um dos recursos visuais mais fáceis de ser compreendido, pois sua construção obedece a lógica “se-então” e a partir dela pode-se derivar regras que são aplicadas em futuras análises.
- **Classes:** reúnem informações que possuem partes semelhantes verificadas pelo algoritmo aplicado.
- **Relação de dependência:** apresentam uma condição de existência de um registro em virtude de outro. São representados pela relação antecedente/consequente, em que o antecedente pode conter inúmeras variáveis que levam a um resultado, o consequente.

Tendo as tarefas e os métodos reconhecidos, o próximo passo é construir uma matriz que indique esses aspectos conforme a técnica de mineração de dados. O Quadro 3 esquematiza a análise das tarefas e métodos fornecidos pelas técnicas de mineração e seus respectivos algoritmos.

Modelo de Conhecimento		Técnicas de mineração de dados		
		Classificação	Clusterização	Associação
		Algoritmos de mineração		
		J48	K-means	Apriori
Tarefas	Grupos			
	Relacionamento			
	Regras de associação			
	Padrões			
Métodos	Árvore			
	Classes			
	Relação de dependência			

Quadro 3 – Tarefas e métodos dos algoritmos das técnicas de mineração de dados

Fonte: Autoria própria (2013)

3.2.3 Analisar as Técnicas de Visualização de Informações

Quando se pretende transmitir uma mensagem na forma gráfica ao invés da textual, é importante saber se o recurso visual escolhido tem capacidade para exibir toda a carga informativa necessária à sua compreensão. Desta forma, a efetividade e a expressividade da técnica de visualização de informações dependem das características intrínsecas a técnica de visualização e à informação a ser representada. As técnicas de visualização que foram analisadas nesta pesquisa estão ilustradas na Figura 2.

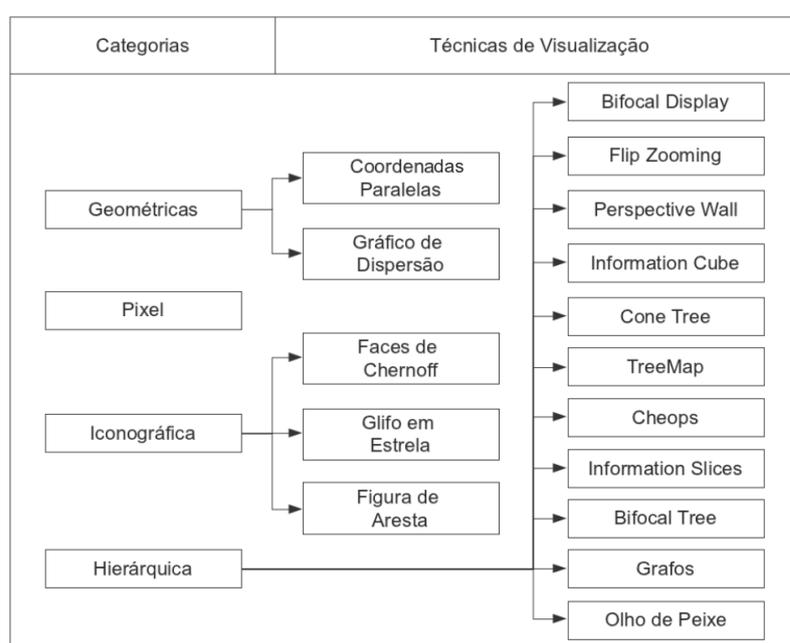


Figura 2 – Categorias e técnicas de visualização de informações
Fonte: Autoria própria (2013)

As técnicas de visualização foram analisadas levando em consideração as tarefas, métodos e suporte de informações, assim:

- Tarefa: representa o tipo de conhecimento que é exibido para o usuário, podendo ser: Grupo, Relacionamento, Regras de Associação e Padrões.
- Método: representa o formato de exibição do conhecimento, o qual pode ser classificado como: Árvore, Classes e Relação de dependência.
- Suporte de informações: representa a quantidade de informações que pode ser exibida na interface gráfica sem que ocorra a sobrecarga, a qual prejudica a compreensão. Neste requisito foi adotada classificação do suporte apresetnada por Yamaguchi (2010) que indicou os limites de 10 até 100 registros para *Pequeno*, de 1.000 até 10.000 para *Médio* e de 1.000.000 até 10.000.000 para *Grande*.

As características tarefa e métodos foram escolhidas por corresponderem as mesmas características analisadas nas técnicas de mineração.

Para a análise do tipo de conhecimento (tarefa), do formato de apresentação do conhecimento (método) e do suporte de informações inerentes às técnicas de visualização seguiu-se a matriz ilustrada no Quadro 4.

Características das informações		Técnicas de visualização																
		Cp	Gd	Px	Fc	Ge	Fa	Bd	Fz	Pw	Ic	Ct	Tm	Cs	Is	Bt	Gf	Op
Tarefas	Grupos																	
	Relacionamento																	
	Regras de associação																	
	Padrões																	
Métodos	Árvore																	
	Classes																	
	Relação de dependência																	
Suporte de informações	Pequeno																	
	Médio																	
	Grande																	

Quadro 4 – Tarefas, métodos e suporte de informações das técnicas de visualização

Fonte: Autoria própria (2013)

Cp: Coordenadas paralelas – Gd: Gráfico de dispersão – Px: Pixel – Fc: Faces de chernoff – Ge: Glifo em estrela – Fa: Figura de arestas – Bd: Bifocal display – Fz: Flip zooming – Pw: Perspective wall – Ic: Information cube – Ct: Cne tree – Tm: Treemap – Cs: Cheops – Is: Information slices – Bt: Bifocal tree – Gf: Grafo – Op: Olho de peixe.

3.2.4 Correlacionar as Técnicas de Mineração e Visualização

O objetivo desta etapa foi estabelecer uma relação entre as técnicas de mineração e as de visualização, com o objetivo de permitir que o conhecimento exibido pelas técnicas de mineração seja melhor compreendido pelo usuário.

A relação entre as técnicas se deu por meio da identificação das características comuns entre elas tais como: Tarefas e Métodos. Esta relação é importante para que aplicação dos recursos gráficos contidos nas técnicas de visualização estejam de acordo com o conhecimento obtido pelos algoritmos de mineração. Este procedimento está esquematizado pelo Quadro 5.

Modelo de Conhecimento		Técnicas de mineração			Técnicas de visualização de informações																
		Classificação	Clusterização	Associação																	
		Algoritmos																			
		J48	K-means	Apriori	Cp	Gd	Px	Fc	Ge	Fa	Bd	Fz	Pw	Ic	Ct	Tm	Cs	Is	Bt	Gf	Op
Tarefas	Grupos																				
	Relacionamento																				
	Regras de associação																				
	Padrões																				
Métodos	Árvore																				
	Classes																				
	Relação de dependência																				

Quadro 5 – Correlação entre técnicas de mineração de dados e técnicas de visualização de informações

Fonte: Autoria própria (2013)

Cp: Coordenadas paralelas – Gd: Gráfico de dispersão – Px: Pixel – Fc: Faces de chernoff – Ge: Glifo em estrela – Fa: Figura de arestas – Bd: Bifocal display – Fz: Flip zooming – Pw: Perspective wall – Ic: Information cube – Ct: Cne tree – Tm: Treemap – Cs: Cheops – Is: Information slices – Bt: Bifocal tree – Gf: Grafo – Op: Olho de peixe.

Após o preenchimento da matriz ilustrada no Quadro 5, deve-se determinar quais técnicas de visualização podem ser usadas na representação do conhecimento resultante da execução dos algoritmos.

Em seguida, caracteriza-se os elementos visuais inerentes à cada técnica de visualização que melhor se adequar aos algoritmos. Para isto, utiliza-se a matriz ilustrada no Quadro 6.

Elementos visuais	Técnicas de de visualização de informações																
	Cp	Gd	Px	Fc	Ge	Fa	Bd	Fz	Pw	Ic	Ct	Tm	Cs	Is	Bt	Gf	Op
Árvore																	
Ângulos																	
Atributos faciais																	
Base circular																	
Brilho																	
Cores diversas																	
Coordenadas (x,y)																	
Cones																	
Compactação																	
Comprimento																	
Cubos																	
Deslocamento																	
Disco semicircular																	
Distorção																	
Divisão de tela																	
Expansão																	
Forma radial																	
Linhas																	
Janela auxiliar																	
Pixel																	
Pontos																	
Projeção 2D																	
Projeção 3D																	
Quadrados																	
Rede																	
Retângulos																	
Sobreposição																	
Símbolos																	
Sombra																	
Tamanho variado																	
Transparência																	
Triângulos																	

Quadro 6 – Elementos visuais usados pelas técnicas de visualização de informações

Fonte: Autoria própria (2013)

Cp: Coordenadas paralelas – Gd: Gráfico de dispersão – Px: Pixel – Fc: Faces de chernoff – Ge: Glifo em estrela – Fa: Figura de arestas – Bd: Bifocal display – Fz: Flip zooming – Pw: Perspective wall – Ic: Information cube – Ct: Cne tree – Tm: Treemap – Cs: Cheops – Is: Information slices – Bt: Bifocal tree – Gf: Grafo – Op: Olho de peixe.

Ao término deste processo segue a proposta dos protótipos que utilize técnicas de visualização, com seus elementos visuais, para representar as informações resultantes da execução dos algoritmos no ambiente *Weka*.

O próximo capítulo descreve os resultados atingidos por meio do presente experimento.

4 ADEQUAÇÃO DAS TÉCNICAS DE VISUALIZAÇÃO ÀS TÉCNICAS DE MINERAÇÃO DE DADOS

Este capítulo apresentará os resultados obtidos por este trabalho. A seção 4.1 relata a descrição das bases de dados usadas no experimento. A seção 4.2 descreve a análise dos resultados fornecidos pelas técnicas de mineração de dados executadas nas bases de dados por meio da ferramenta *Weka*. A seção 4.3 apresenta o estudo e os resultados sobre as técnicas de visualização. A seção 4.4 correlaciona as técnicas de visualização mais adequadas às técnicas de mineração. Ao final, na seção 4.5 apresenta os protótipos da utilização de técnicas de visualização nos resultados originados das técnicas de mineração pela ferramenta *Weka*.

4.1 BASES DE DADOS PARA O EXPERIMENTO

Foram escolhidas bases que permitissem a mineração pelas técnicas inerentes a ferramenta *Weka* e que tivessem atributos suficientes para a execução dos algoritmos, ou seja, possa-se formar grupos, relacionamentos, regras de associação e padrões, e ainda permitissem a elaboração de árvore, rede de relacionamento, classes ou relação de dependência em seus resultados.

As bases de dados utilizadas foram de domínio público, as bases 1 (Consumidor) e 2 (Seguros) estão disponível para *download* no sitio do *Machine Learning Repository* (UCI, 2012), a base 3 (Supermercado) está disponível em *Sample Weka Data Sets* (CIS, 2012).

As bases são necessárias para tornar o trabalho mais próximo da realidade. A seguir segue a descrição das três bases de dados que foram utilizadas no presente trabalho.

- Base 1: *Consumidor*

Essa base contém dados sobre vendas de produtos e possui ao total 1.966 instâncias com um total de 49 atributos, sendo 46 atributos numéricos e 3 atributos categóricos. Atributos numéricos só aceitam valores que caracterizem quantidade; atributos categóricos aceitam valores que qualificam um estado ou situação.

A composição dos dados nesta base permite a aplicação de tarefas de classificação.

- Base 2: *Seguros*

Base relacionada a uma companhia de seguros contento perfis completos dos clientes referentes as informações sobre compra ou não da apólice de seguro. O arquivo contém 86 atributos, todos do tipo categórico, que foram substituídos por valores numéricos para exploração do algoritmo, esta transformação se deu para tornar os registros adequados aos algoritmos de clusterização.

- Base 3: *Supermercado*

Esta base tem objetivo de identificar o comportamento de compra dos clientes de um supermercado, possui 4.627 instâncias e seus registros têm 217 atributos, destes 27 são categóricos e o restante de natureza indefinida. Esta base foi indicada para algoritmos de associação.

Todas estas bases já estão formatadas na extensão específica do ambiente *Weka* que é *.arff* (*Attribute Relation File Format*). Para a exploração das bases de dados, nenhuma alteração foi realizada nas configurações básicas da ferramenta.

A relação bases, técnicas de mineração e algoritmos usados neste trabalho é resumida no Quadro 7.

Base de dados	Técnica de Mineração	Algoritmo
Consumidor	Classificação	J48
Seguros	Clusterização	k-means
Supermercado	Associação	Apriori

Quadro 7 – Relação Base x Técnica de Mineração x Algoritmo

Fonte: Autoria própria (2013)

Os resultados obtidos neste processo de mineração são relatados na próxima seção.

4.2 ANÁLISE DOS RESULTADOS GERADOS PELAS TÉCNICAS DE MINERAÇÃO

A aplicação de recursos visuais nos resultados produzidos pelas técnicas de mineração requer, primeiramente, a definição das tarefas e métodos, que constituem o Modelo de Conhecimento, fornecidos pelos algoritmos de mineração.

A *Weka*, ferramenta selecionada para realização do experimento, permite explorar os registros utilizando-se três técnicas: classificação, clusterização e associação. Elas executam o processo e seus algoritmos exploram os registros em busca de regras, padrões, grupos ou

relacionamentos e apresentam esses resultados em diversos formatos, tais como árvore, classes ou relação de dependência.

Para estabelecer quais dessas informações são funções dos algoritmos escolhidos foram estudadas as definições das técnicas e algoritmos de mineração de dados, das quais obteve-se os resultados apresentados no Quadro 8.

Modelo de Conhecimento		Técnicas de mineração de dados		
		Classificação	Clusterização	Associação
		Algoritmos de mineração		
		J48	K-means	Apriori
Tarefas	Grupos		●	
	Relacionamento	●		
	Regras de associação			●
	Padrões	●	●	●
Métodos	Árvore	●		
	Classes		●	
	Relação de dependência			●

Quadro 8 – Tarefas e métodos dos algoritmos das técnicas de mineração de dados

Fonte: Autoria própria (2013)

De acordo com o Quadro 8, a distribuição das tarefas ocorre, na ordem:

- Grupos: estão presentes na técnica de Clusterização e é função operacional do algoritmo k-means.
- Relacionamento: tarefa funcional exercida pela técnica de Classificação com o algoritmo J48.
- Regras de associação: é a informação extraída com a técnica de Associação por meio do algoritmo Apriori.
- Padrões: comum a todas as técnicas e algoritmos.

Em relação aos métodos, o Quadro 8 demonstra que:

- Árvore: é o formato de exibição do algoritmo J48.
- Classes: é a forma de expor as informações mineradas do algoritmo k-means.
- Relação de dependência: é o método utilizado pelo algoritmo Apriori.

Ou seja, na *Weka* as técnicas com estes algoritmos específicos seguem a relação ilustrada na Figura 3.

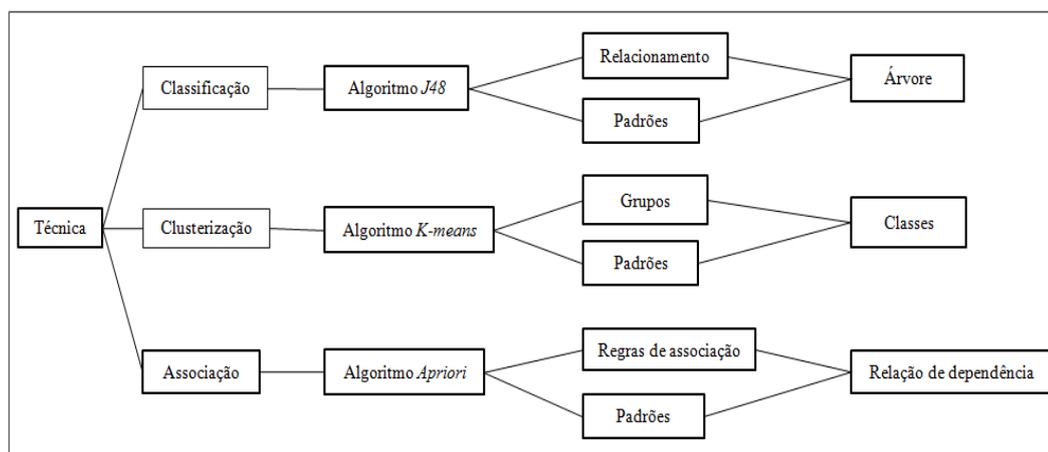


Figura 3 – Relação tarefas e métodos das técnicas de mineração

Fonte: Autoria própria (2013)

Após o término do processo de mineração pelas três técnicas e pelos três algoritmos a *Weka* gera um relatório com os resultados encontrados. Observando os resultados fornecidos se pode concluir que a localização dos resultados seguem um padrão dividido da seguinte forma:

- Cabeçalho: indica a natureza da informação, sua composição (número de atributos e quais são eles), as instâncias ou linhas de registros que compõe os dados, o nome da técnica escolhida e o algoritmo usado.
- Conhecimento Novo: representa o “Modelo de Conhecimento” obtido pela execução das técnicas e algoritmos de mineração. Estabeleceu-se o nome “Conhecimento Novo” porque o objetivo das técnicas de mineração não é extrair somente uma informação a partir dos dados armazenados, mas sim processar e cruzar os dados para gerar algo útil e novo. Os grupos, regras de associação, padrões e relacionamentos são representados graficamente por árvore, classes e relação de dependência.
- Informações adicionais: dão suporte às informações constantes na árvore, nas regras ou nas classes. As informações exibidas são: acurácia, tempo de processo e percentual de erros encontrados nos registros.

Considerando o padrão de visualização de resultado oferecido pela ferramenta citado anteriormente, concluiu-se:

- Classificação (algoritmo J48): cabeçalho, conhecimento novo (relacionamentos e padrões representados por árvore) e informações adicionais.
- Clusterização (algoritmo k-means): cabeçalho, conhecimento novo (grupos e padrões representados por classes) e informações adicionais.
- Associação (algoritmo apriori): cabeçalho, conhecimento novo (regras de associação e

padrões representados por relação de dependência) e informação adicionais.

Para tornar mais compreensível esta divisão e a relação mostrada na Figura 3, a seguir, são apresentados três experimentos usando as bases de dados descritas na seção 4.1, na ferramenta *Weka*. Nesta amostragem serão feitos comentários sobre as informações encontradas e sobre a maneira de exibição dessas informações.

4.2.1 Experimento com a Base Consumidor

A Figura 4 mostra o resultado da aplicação do algoritmo J48 para a base Consumidor, considerando o item Cabeçalho.

```

Classifier output
=== Run information ===
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: consumidor
Instances: 1966
Attributes: 49
AMOUNT
INCOME
HOMEVAL
FREQUENT
REGENCY
MARITAL
NTITLE
AGE
TELIND
APRTHWT
MOBILE
DOMESTIC
APPAREL
LEISURE
KITCHEN
LUXURY
PROMO7
PROMO13
COUNTY
RETURN
MENSWARE
FLATWARE
DISHES
HOMEACC
LAMPS
LINENS
BLANKETS
TOWELS
OUTDOOR
COATS

```

Figura 4 – Cabeçalho do resultado de mineração da base Consumidor

Fonte: *Weka* (2012)

O cabeçalho é dividido em quatro partes: *Scheme*, refere-se ao nome da técnica e do algoritmo que estão sendo usados para minerar. *Relation*, o nome da base de dados. *Intances*, o total de instâncias que a base de dados possui. *Attributes*, a quantidade de atributos e a listagem deles. Esta divisão é para todos os resultados das técnicas e algoritmos aplicados.

No caso da base Consumidor, o cabeçalho informa que a técnica de mineração é Classificação, o algoritmo aplicado foi o J48, o nome da base é Consumidor, tem 1966 linhas

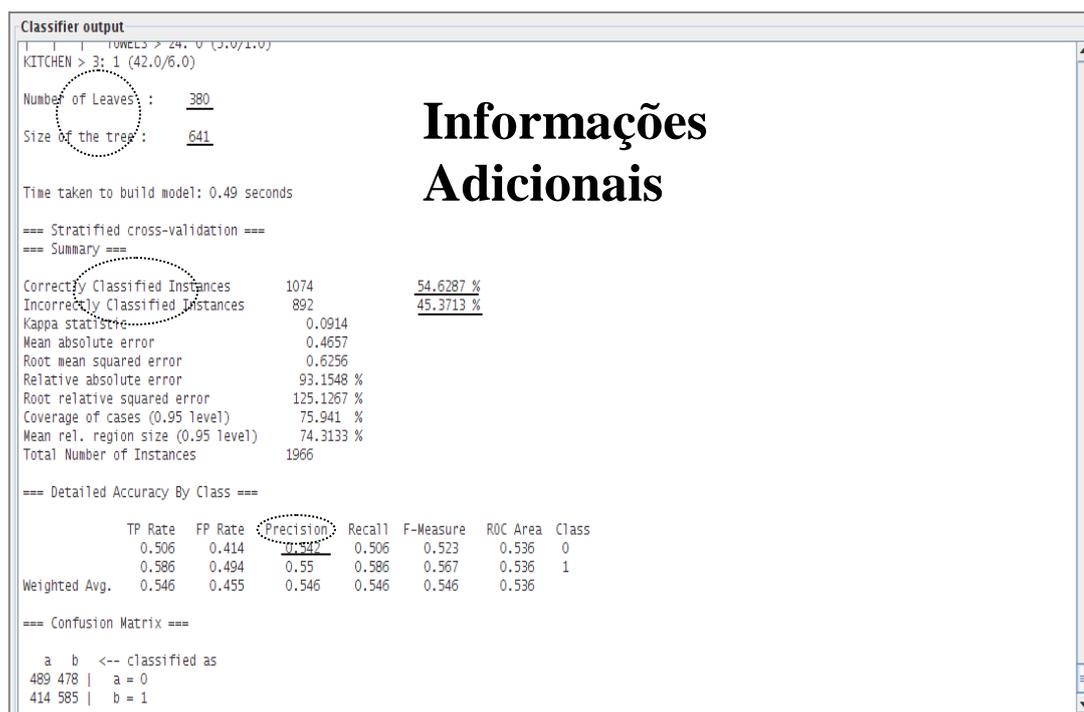


Figura 6 – Informações Adicionais fornecidas pela técnica Classificação

Fonte: Weka (2012)

Algumas informações adicionais são exibidas tais como: tamanho da árvore (380), quantos nós ou folhas foram construídas (641); o número de análises corretas (1074) e o número de incorretas (892). Outra informação relevante é vista mais abaixo onde se indica a porcentagem de precisão da análise do J48, que foi de 54%.

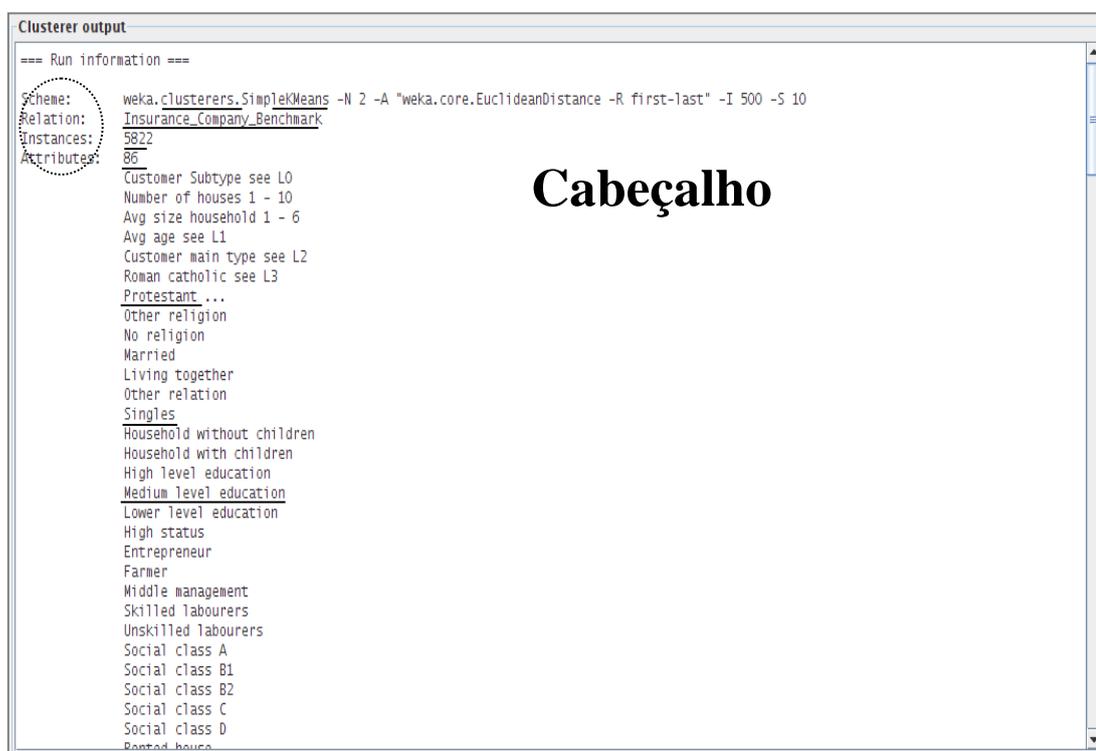
Um dos passos desta pesquisa é analisar as condições de visualização do conhecimento obtido pela técnica de mineração, o Conhecimento Novo (Figura 5). Alguns problemas de interação com o usuário foram observados e descritos a seguir:

- Não há visão total do conhecimento extraído: na tela onde se mostra o início da árvore, só se pode ver parte dela, 28 níveis de hierarquia, sendo que ao todo são 380. Ou seja, para continuar a leitura da árvore, o usuário deve arrastar a barra de rolagem da ferramenta.
- Não há possibilidade de interação: operações como de reduzir os níveis de hierarquia ou dos relacionamentos, ampliar o tamanho da visualização, focar em apenas um nível se quiser se concentrar em apenas uma relação não são contemplados pela ferramenta.
- Se a árvore for muito grande, como no caso deste experimento, o usuário fica impossibilitado de manter uma linha de lógica da localização.
- Informações de peso igual: não são usados cor ou tamanhos de fonte diferentes que indiquem relevância de informações ou do subnível, toda a árvore segue um mesmo padrão de apresentação. O único simbolismo que denota relevância é de quem é *raiz* e

de quem são *nós-filho*.

4.2.2 Experimento com a Base Seguros

Minerando a base relacionada a uma companhia de seguros, a Figura 7 mostra o resultado da aplicação do algoritmo *K-means* para a base Seguros, considerando o item Cabeçalho.



```
Clusterer output
=== Run information ===
Scheme: weka.clusterers.SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10
Relation: Insurance_Company_Benchmark
Instances: 5822
Attributes: 86
Customer Subtype see L0
Number of houses 1 - 10
Avg size household 1 - 6
Avg age see L1
Customer main type see L2
Roman catholic see L3
Protestant ...
Other religion
No religion
Married
Living together
Other relation
Singles
Household without children
Household with children
High level education
Medium level education
Lower level education
High status
Entrepreneur
Farmer
Middle management
Skilled labourers
Unskilled labourers
Social class A
Social class B1
Social class B2
Social class C
Social class D
Rented house
```

Cabeçalho

Figura 7 – Cabeçalho fornecido pela técnica Clusterização
Fonte: Weka (2012)

As informações que seguem são: em *Scheme*, a técnica é de clusterização pelo algoritmo *simplekmeans* (nome dado ao *kmeans* nesta ferramenta). *Relation*, o nome da base é *Insurance Company Benchmark*. *Intances*, tem total de 5.822 instâncias. *Attributes*, com 86 atributos entre eles *Protestant*, *Singles* e *Medium level education*.

A Figura 8 mostra os grupos de classes obtidos pelo algoritmo *k-means* para a base Seguros na parte Conhecimento Novo.

Nesta figura os atributos são separados em classes: total, 0 e 1. Observa-se também a quantidade de registros lançados em cada classe.

Clusterer output

```

=== Clustering model (full training set) ===

kMeans
=====

Number of iterations: 17
Within cluster sum of squared errors: 11999.911117664975
Missing values globally replaced with mean/mode

Cluster centroids:
Attribute          Full Data          Cluster#
                   (5822)            (1892)            (3930)
-----
Customer Subtype see L0      24.2533      10.6464      30.8041
Number of houses 1 - 10     1.1106       1.1242       1.1041
Avg size household 1 - 6    2.6788       2.9593       2.5438
Avg age see L1              2.9912       2.935        3.0183
Customer main type see L2   5.7736       2.8214       7.1949
Roman catholic see L3       0.6965       0.972        0.9639
Protestant ...              4.6269       4.6543       4.6137
Other religion               1.0699       0.9725       1.1168
No religion                  3.2585       3.1094       3.3303
Married                      6.1834       7.0185       5.7814
Living together              0.8835       0.7326       0.9562
Other relation               2.2905       1.5576       2.6433
Singles                      1.8877       1.2215       2.2084
Household without children  3.2303       3.3303       3.1822
Household with children     4.3002       4.8087       4.0555
High level education         1.461        2.6591       0.8842
Medium level education       3.3513       4.1823       2.9511
Lower level education        4.5725       2.6047       5.5198
High status                  1.8951       3.2764       1.23
Entrepreneur                 0.398        0.5724       0.314

```

**Conhecimento
Novo**

Figura 8 – Conhecimento Novo fornecido pela técnica Clusterização
Fonte: Weka (2012)

As informações que dão suporte às classes estão alocados em Informações Adicionais, na Figura 9.

Clusterer output

```

number of private third party insurance      0.405      0.4513      0.3695
Number of third party insurance (firms)      0.0148      0.0185      0.013
Number of third party insurance (agriculture) 0.0206      0.0143      0.0237
Number of car policies                       0.5622      0.5936      0.5471
Number of delivery van policies              0.0105      0.0122      0.0097
Number of motorcycle/scooter policies        0.0411      0.0439      0.0397
Number of lorry policies                     0.0022      0.0026      0.002
Number of trailer policies                   0.0125      0.01        0.0137
Number of tractor policies                   0.0337      0.0275      0.0366
Number of agricultural machines policies     0.0062      0.0032      0.0076
Number of moped policies                     0.0704      0.0603      0.0753
Number of life insurances                    0.0766      0.1062      0.0623
Number of private accident insurance policies 0.0053      0.0032      0.0064
Number of family accidents insurance policies 0.0065      0.0106      0.0046
Number of disability insurance policies      0.0046      0.0063      0.0038
Number of fire policies                      0.5701      0.6025      0.5545
Number of surfboard policies                 0.0005      0.0011      0.0003
Number of boat policies                      0.006       0.0095      0.0043
Number of bicycle policies                   0.0318      0.0418      0.027
Number of property insurance policies        0.0079      0.0116      0.0061
Number of social security insurance policies 0.0143      0.0248      0.0092
class                                         0           0           0

Time taken to build model (full training data) : 3.24 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      1892 ( 32%)
1      3930 ( 68%)

```

**Informações
Adicionais**

Figura 9 – Informações Adicionais fornecida pela técnica Clusterização
Fonte: Weka (2012)

As Informações adicionais são: 5.822 instâncias que continham o perfil dos clientes da companhia, 68% ou 3930 estão na classe 1 e 32% ou 1892 na classe 0.

Alguns problemas de interação com o usuário também são notados no formato de apresentação dos resultados na Figura 8 (Conhecimento Novo):

- Todos os atributos dos registros e as classes resultantes só podem ser totalmente compreendidos se a janela de apresentação for movida para cima ou para baixo.
- Novamente não há alteração na apresentação que indique qual informação é mais importante.
- Não se pode analisar as classes individualmente com seus componentes.
- Não se tem uma visão do tamanho que cada registro ocupa dentro das classes.

4.2.3 Experimento da Base Supermercado

O Cabeçalho e as Informações Adicionais da base de dados Supermercado está ilustrado pela Figura 10.

```

Associator output
=== Run information ===
Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    supermarket
Instances:   4627
Attributes:  217
              [List of attributes omitted]
=== Associator model (full training set) ===

Apriori
=====
Minimum support: 0.15 (694 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:
Size of set of large itemsets L(1): 44
Size of set of large itemsets L(2): 380
Size of set of large itemsets L(3): 910
Size of set of large itemsets L(4): 633
Size of set of large itemsets L(5): 105
Size of set of large itemsets L(6): 1

Best rules found:
1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 <conf:(0.92)> lift:(1.27) lev:(0.03) [155] conv:(3.35)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf:(0.92)> lift:(1.27) lev:(0.03) [149] conv:(3.28)
3. baking needs=t frozen foods=t fruit=t total=high 720 ==> bread and cake=t 705 <conf:(0.92)> lift:(1.27) lev:(0.03) [150] conv:(3.27)

```

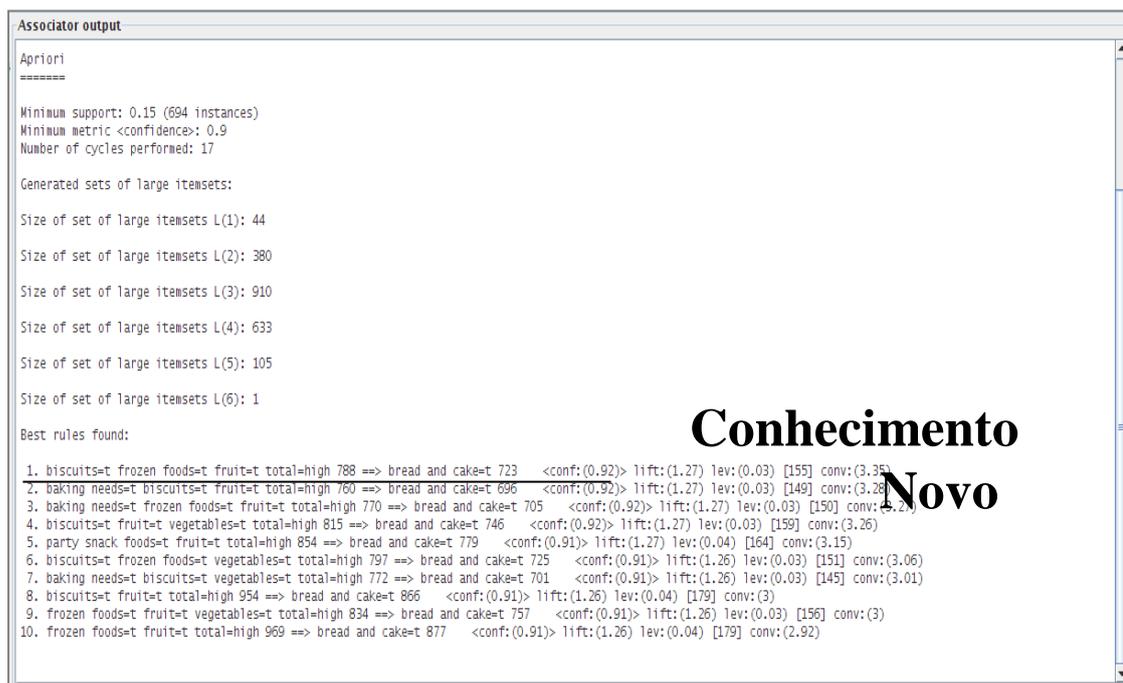
Figura 10 – Cabeçalho e Informações Adicionais fornecidos pela técnica Associação

Fonte: Weka (2012)

Nesta figura o *Scheme* informa que a mineração é por Associação, o algoritmo é Apriori. Em *Relation*, o nome da base é *Supermarket*. Para *Instances* tem-se o total de 4.627 instâncias. A quantidade de atributos desta base é 217.

Nas Informações Adicionais é mostrado que foram encontradas regras de associação em 694 instâncias das 4.627, e a confiança de que estão corretas as relações é de 0,9.

A Figura 11 mostra o Conhecimento Novo gerado pelo algoritmo Apriori.



```

Associator output
Apriori
=====
Minimum support: 0.15 (694 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:
Size of set of large itemsets L(1): 44
Size of set of large itemsets L(2): 380
Size of set of large itemsets L(3): 910
Size of set of large itemsets L(4): 633
Size of set of large itemsets L(5): 105
Size of set of large itemsets L(6): 1

Best rules found:
1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 <conf:(0.92)> lift:(1.27) lev:(0.03) [155] conv:(3.35)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf:(0.92)> lift:(1.27) lev:(0.03) [149] conv:(3.28)
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 <conf:(0.92)> lift:(1.27) lev:(0.03) [150] conv:(3.27)
4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746 <conf:(0.92)> lift:(1.27) lev:(0.03) [159] conv:(3.26)
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779 <conf:(0.91)> lift:(1.27) lev:(0.04) [164] conv:(3.15)
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 <conf:(0.91)> lift:(1.26) lev:(0.03) [151] conv:(3.06)
7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701 <conf:(0.91)> lift:(1.26) lev:(0.03) [145] conv:(3.01)
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(3)
9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757 <conf:(0.91)> lift:(1.26) lev:(0.03) [156] conv:(3)
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(2.92)

```

Figura 11 – Conhecimento Novo fornecido pela técnica Clusterização
Fonte: Weka (2012)

Nas Regras simbolizadas pela relação $x \Rightarrow y$, neste caso, aponta que grupos de produtos são adquiridos conjuntamente. Por exemplo, a Regra 1 refere-se a relação de compra de biscoitos ($biscuits=t$), comida gelada ($frozen\ foods=t$) e fruta ($fruit=t$) que ocorre 788 vezes é acompanhada pela compra de pão ($bread=t$) e bolo ($cake=t$) em 723 ocorrências, com 92% de certeza.

Sobre a forma de apresentação dos resultados do Conhecimento Novo, pela Figura 11 observou-se que:

- As regras precisam ser atentamente estudadas para reconhecer a relação de dependência.
- Não é possível filtrar as informações para reduzir o número de relações ou detalhar uma relação preferencial.
- Muitas regras podem causar sobrecarga, o que pode prejudicar a compreensão do conhecimento.

- É possível induzir que regras lineares podem se tornar difíceis de ler caso o número de variáveis seja grande.

Observou-se que todos os formatos de apresentação dos resultados obrigam o usuário a ter conhecimento da linguagem utilizada pela ferramenta para exibir cada parte do resultado, ocasionando a usuário não-especialista uma demora na localização da informação que o motivou à mineração pelo *Weka*.

Para tentar simplificar este procedimento é que a aplicação de técnicas de visualização se mostra útil, na qual os recursos gráficos podem contribuir para a melhoria da compreensão das informações.

4.3 ANÁLISE DAS TÉCNICAS DE VISUALIZAÇÃO

Após a identificação de problemas de interação com usuário na ferramenta *Weka* ao executar os algoritmos, analisou-se cada técnica de visualização para se ter uma abrangência geral de seu uso para depois correlacioná-la aos algoritmos de mineração.

As técnicas de visualização foram analisadas por três aspectos: tarefas (tipo de conhecimento extraído), métodos (formato de apresentação) e suporte de informações. O resultado desta análise está disposta no Quadro 9.

Características das informações		Técnicas de visualização																
		Cp	Gd	Px	Fc	Ge	Fa	Bd	Fz	Pw	Ic	Ct	Tm	Cs	Is	Bt	Gf	Op
Tarefas	Grupos	●	●	●		●	●		●		●	●	●	●	●	●	●	
	Relacionamento	●					●				●	●	●	●	●	●	●	
	Regras de associação						●										●	
	Padrões	●	●	●	●	●	●		●		●	●	●	●	●	●	●	●
Métodos	Árvore										●	●	●	●	●	●	●	
	Classes	●	●	●		●	●		●				●				●	
	Relação de dependência						●										●	
Suporte de informações	Pequeno				●	●												
	Médio		●				●		●								●	
	Grande	●		●							●	●	●	●	●	●		

Quadro 9 – Tarefas, métodos e suporte das técnicas de visualização

Fonte: Autoria própria (2013)

Cp: Coordenadas paralelas – Gd: Gráfico de dispersão – Px: Pixel – Fc: Faces de chernoff – Ge: Glifo em estrela – Fa: Figura de arestas – Bd: Bifocal display – Fz: Flip zooming – Pw: Perspective wall – Ic: Information cube – Ct: Cne tree – Tm: Treemap – Cs: Cheops – Is: Information slices – Bt: Bifocal tree – Gf: Grafo – Op: Olho de peixe.

A seguir descrevem-se os resultados da análise:

4.3.1 Técnicas de Visualização Geométricas

Em geral, as técnicas geométricas facilitam a identificação de padrões nas informações.

A técnica Coordenadas Paralelas se mostra melhor para representar informações que possuam algum tipo de relacionamento com formação de agrupamentos e podem ser exibidos através de classes.

A técnica Gráfico de Dispersão pode ser usada para a representação de classes mas o relacionamento nos grupos não fica perceptível em sua estrutura gráfica.

No quesito quantidade de informações, o Gráfico de Dispersão possui capacidade menor do que Coordenadas Paralelas que permite a exibição de grandes volumes de informações.

4.3.2 Técnicas de Visualização Orientadas a Pixel

Representações por pixels têm a vantagem de suportar grandes quantidades de informações. Por ela é possível visualizar a tarefa de formação de agrupamentos e padrões existentes. Em métodos, é ideal para a exibição de classes.

4.3.3 Técnicas de Visualização Iconográficas

Na representação de informações por ícones e glifos, verificou-se que as *Faces de Chernoff* conseguem demonstrar padrões existentes nas informações. Mas devido a sua estrutura gráfica a percepção de grupos, relacionamentos ou regras de associação fica prejudicada. Outro problema ligado a sua característica estrutural é que ela não atende aos métodos estabelecidos, tornando esta técnicas não ideal para o objetivo deste trabalho.

Glifo em Estrela possibilita a visualização de padrões que resultem na formação de grupos distintos e seu gráfico permite a representação das classes.

Dentre as três técnicas, Figura de arestas foi a que melhor correspondeu aos tipos de informações analisados. Ela consegue explicitar os padrões por agrupamentos, relacionamentos e regras de associação. Ela também permite a identificação visual de classes e de relações de dependências.

Quando se trata de quantidade de informações, Figura de Aresta é a que mais suporta informações, ainda assim possui capacidade média, enquanto as outras duas, *Faces de Chernoff* e *Glifo em estrela*, suportam pequeno volume de informações em suas representações.

4.3.4 Técnicas de Visualização Hierárquicas

Com relação as tarefas, os padrões são representados por grupos em oito das onze técnicas (*Flipz Zooming*, *Information Cube*, *Cone Tree*, *TreeMap*, *Cheops*, *Informations Slices*, *Bifocal Tree* e *Gafos*), além disso, sete dessas mesmas técnicas também explicitam os relacionamentos contidos nas informações (*Information Cube*, *Cone Tree*, *TreeMap*, *Cheops*, *Informations Slices*, *Bifocal Tree* e *Gafos*). Destaque para a técnica de Grafo que, apesar de constar como técnica hierárquica, sua representação gráfica também permite a apresentação de informações sem uma hierarquia ou nó *raiz*, por isso, consegue apresentar regras de associação.

Para transmitir a informação, sete técnicas utilizam o método de árvore (*Flipz Zooming*, *Information Cube*, *Cone Tree*, *TreeMap*, *Cheops*, *Informations Slices*, *Bifocal Tree* e *Gafos*), a visualização de classes é atendida pelas técnicas *Flip Zooming*, *TreeMap* e *Gafos*. Esta última ainda permite a interpretação de relações de dependência.

Genericamente, as técnicas desta categoria podem dispor grandes volumes de informações na tela, as limitações estariam no desenho gráfico e nos objetos utilizados para representar as informações, Grafo, no entanto, alcança a saturação de informações mais rapidamente.

Três técnicas não atendem aos requisitos analisados, *Bifocal Display*, *Perspective Wall* e *Olho de peixe* pois possuem a função de apenas realçarem as informações exibidas.

4.4 CORRELAÇÃO DAS TÉCNICAS DE VISUALIZAÇÃO COM TÉCNICAS DE MINERAÇÃO

Identificados os “Modelos de Conhecimento” (tarefas e métodos) presentes nas técnicas de mineração e na visualização de informações, o próximo passo é combiná-los. Para que uma técnica de visualização seja apontada como adequada é preciso que ela atenda a todas as tarefas e métodos implementados pelas técnicas e algoritmos de mineração. Assim foram identificadas quais são as mais ideais, conforme o Quadro 10.

Modelo de Conhecimento		Técnicas de mineração			Técnicas de visualização de informações																
		Classificação	Clusterização	Associação																	
		Algoritmos																			
		J48	K-means	Apriori	Cp	Gd	Px	Fc	Ge	Fa	Bd	Fz	Pw	Ic	Ct	Tm	Cs	Is	Bt	Gf	Op
Tarefas	Grupos		●		●	●	●		●	●		●		●	●	●	●	●	●	●	
	Relacionamento	●			●					●				●	●	●	●	●	●	●	
	Regras de associação			●						●										●	
	Padrões	●	●	●	●	●	●	●	●	●		●		●	●	●	●	●	●	●	●
Métodos	Arvore	●											●	●	●	●	●	●	●	●	
	Classes		●		●	●		●	●		●				●					●	
	Relação de dependência			●						●										●	

Quadro 10 – Correlação entre técnicas de mineração de dados e técnicas de visualização de informações

Fonte: Autoria própria (2013)

Cp: Coordenadas paralelas – Gd: Gráfico de dispersão – Px: Pixel – Fc: Faces de chernoff – Ge: Glifo em estrela – Fa: Figura de arestas – Bd: Bifocal display – Fz: Flip zooming – Pw: Perspective wall – Ic: Information cube – Ct: Cne tree – Tm: Treemap – Cs: Cheops – Is: Information slices – Bt: Bifocal tree – Gf: Grafo – Op: Olho de peixe.

De acordo com o quadro conclui-se que:

- Para o algoritmo J48, que contempla as tarefas de análise de relacionamento, detecção de padrões e usa o método de construção de árvores, as técnicas de visualização mais adequadas são: *Information Cube*, *Cone Tree*, *TreeMap*, *Cheops*, *Information Slices*, *Bifocal Tree* e Grafos, conforme ilustra o grafo de relacionamento da Figura 12.

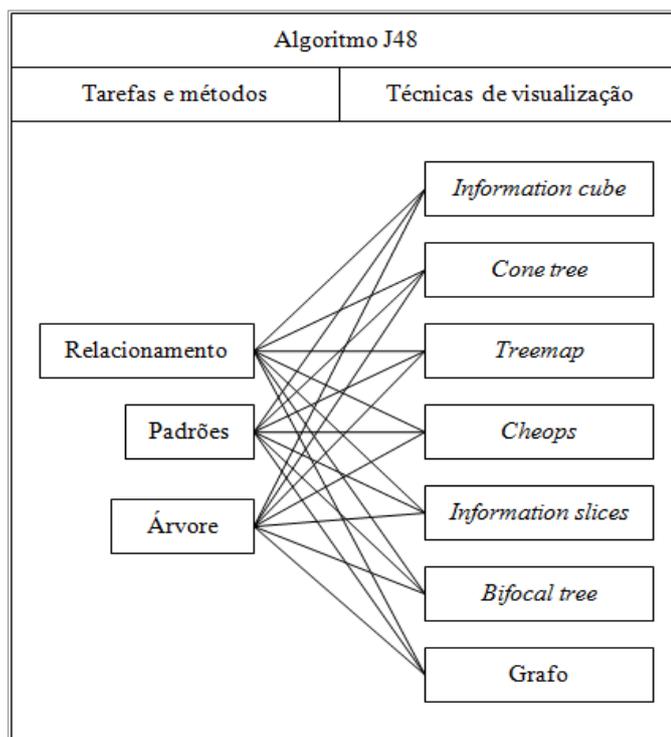


Figura 12 – Técnicas de visualização que correspondem às características do algoritmo J48

Fonte: Autoria própria (2013)

- Para o algoritmo K-means, o qual opera as tarefas de formação de grupos, detecção de padrões e utiliza o método de apresentar as informações em classes, as técnicas de visualização Coordenadas Paralelas, Gráfico de Dispersão, Pixel, Glifo em estrela, Figura de aresta, *Flip Zooming*, *TreeMap* e Grafos se mostram mais apropriadas, como mostra a Figura 13.

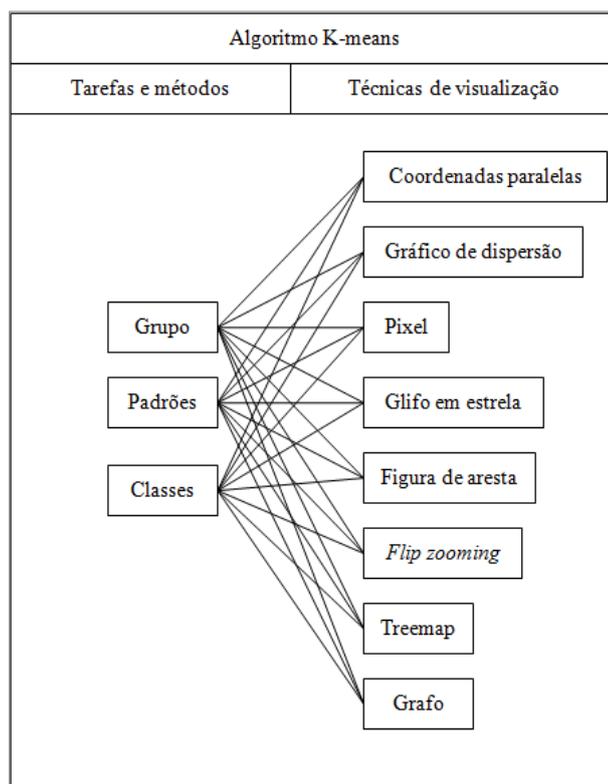


Figura 13 – Técnicas de visualização que correspondem as características do algoritmo K-means
Fonte: Autoria própria (2013)

- Para o algoritmo Apriori que pratica as tarefas de verificação de regras, detecção de padrões e demonstra seus resultados pela relação de dependência, as técnicas de visualização mais adequadas são Figuras de Aresta e Grafos e estão indicadas na Figura 14.

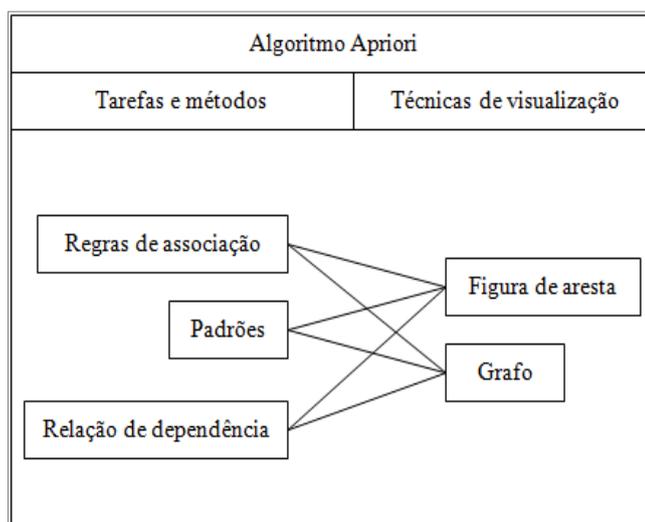


Figura 14 – Técnicas de visualização que correspondem as características do algoritmo Apriori
Fonte: Autoria própria (2013)

Após a identificação das melhores técnicas para cada algoritmo, é importante reconhecer quais são os elementos visuais que pode ser usados para representar a informação

em cada técnica. O Quadro 11 ilustra o resultado desta identificação. Por exemplo, para a Técnica *TreeMap* os elementos visuais que devem ser usados são: árvore, retângulos, tamanhos variados, cores diversas e divisão da tela de exibição, além de deslocamento.

Elementos visuais	Técnicas de de visualização de informações																
	Cp	Gd	Px	Fc	Ge	Fa	Bd	Fz	Pw	Ic	Ct	Tm	Cs	Is	Bt	Gf	Op
Árvore										●	●	●	●	●	●	●	
Ângulos					●	●											
Atributos faciais																	
Base circular											●						
Brilho																	
Cores diversas	●	●	●		●	●						●	●	●	●	●	
Coordenadas (x,y)	●	●															
Cones											●						
Compactação													●				
Comprimento					●	●										●	
Cubos										●							
Deslocamento								●			●	●		●	●		
Disco semicircular														●			
Distorção																	
Divisão de tela								●				●					
Expansão														●	●		
Forma radial																	
Linhas	●				●	●									●	●	
Janela auxiliar																	
Pixel			●														
Pontos		●															
Projeção 2D														●			
Projeção 3D										●							
Quadrados																	
Rede																●	
Retângulos								●			●	●					
Sobreposição																	
Símbolos																	
Sombra																	
Tamanho variado								●				●		●			
Transparência										●							
Triângulos													●				

Quadro 11 – Elementos visuais usados pelas técnicas de visualização de informações

Fonte: Autoria própria (2013)

Cp: Coordenadas paralelas – Gd: Gráfico de dispersão – Px: Pixel – Fc: Faces de chernoff – Ge: Glifo em estrela – Fa: Figura de arestas – Bd: Bifocal display – Fz: Flip zooming – Pw: Perspective wall – Ic: Information cube – Ct: Cne tree – Tm: Treemap – Cs: Cheops – Is: Information slices – Bt: Bifocal tree – Gf: Grafo – Op: Olho de peixe.

Os elementos visuais das respectivas técnicas foram aplicados na elaboração de protótipos para cada algoritmo com o objetivo de melhorar a visualização do Conhecimento Novo.

4.5 PROTÓTIPO DE INTERFACE GRÁFICA PARA VISUALIZAÇÃO DO CONHECIMENTO NOVO

Para a construção dos modelos de apresentação aplicando elementos visuais das técnicas de visualização de informações foi utilizada a ferramenta de prototipação *Balsamiq Mockup* (BALSAMIQ, 2013).

Esta ferramenta permite esboçar graficamente modelos de telas como, por exemplo, páginas de internet ou interface de um programa e, uma característica relevante desta ferramenta é que todos os seus desenhos possuem traços imperfeitos dando a ideia de rascunho.

A seguir serão exibidos protótipos de representação visual dos “Modelos de Conhecimento” do Conhecimento Novo gerado pelos algoritmos de mineração J48, K-means e Apriori no ambiente *Weka* utilizando os recursos fornecidos pela ferramenta *Balsamiq Mockup*.

4.5.1 Protótipo para Representação Visual do Algoritmo J48

A Figura 5 ilustrou o conhecimento obtido pelo uso do Algoritmo J48 na ferramenta *Weka* para a base Consumidor. A Figura 15 mostra como a visualização do resultado pode ser exibido, usando a técnica de visualização *Bifocal Tree*.

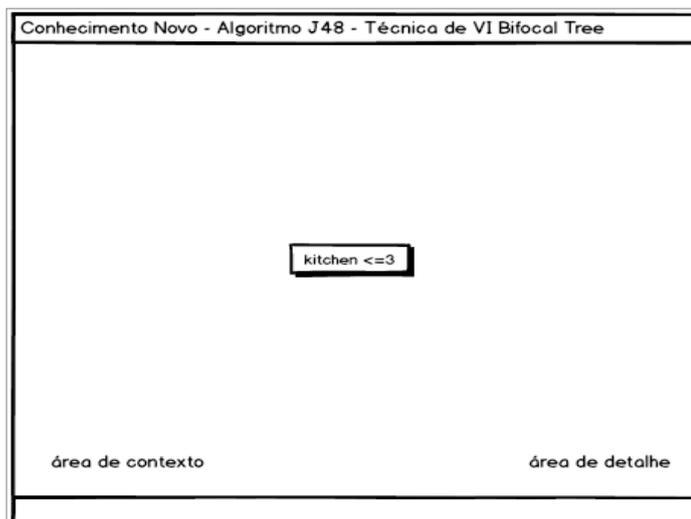


Figura 15 – A raiz dá início à árvore Bifocal Tree
Fonte: Autoria própria (2013)

Nesta técnica, o usuário terá conhecimento dos campos de observação da informação. Do lado esquerdo fica a área de contexto e do lado direito a área do detalhe, além de conhecer o atributo apontado pelo algoritmo J48 como a raiz da árvore de decisão que é “kitchen <=3”.

Depois do usuário clicar na raiz, Kitchen <=3 se desloca para a área de detalhe e mostra os nós-filhos, como na Figura 16.

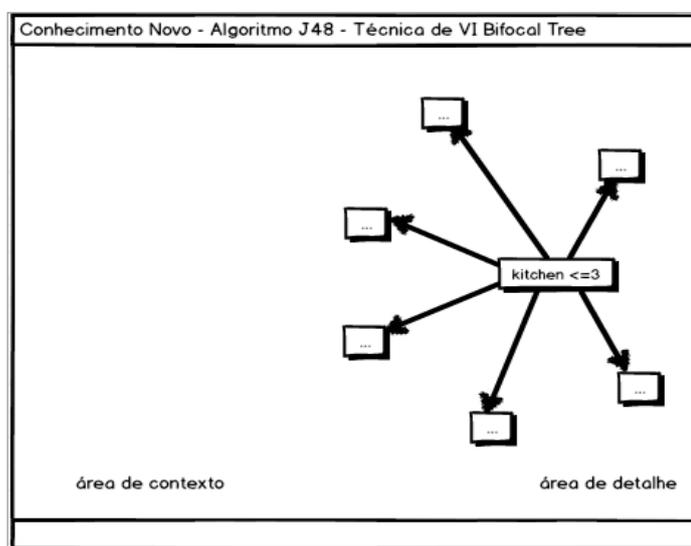


Figura 16 – Nós-filhos da raiz exibidos na área de detalhe
Fonte: Autoria própria (2013)

Quando o usuário seleciona um dos nós-filhos na área de detalhe, a raiz se desloca para a área de contexto e o nó-filho selecionado para a área de detalhe onde exibe seu conteúdo, como ilustra a Figura 17. Cores podem ser usadas para diferenciar a informação selecionada das demais.

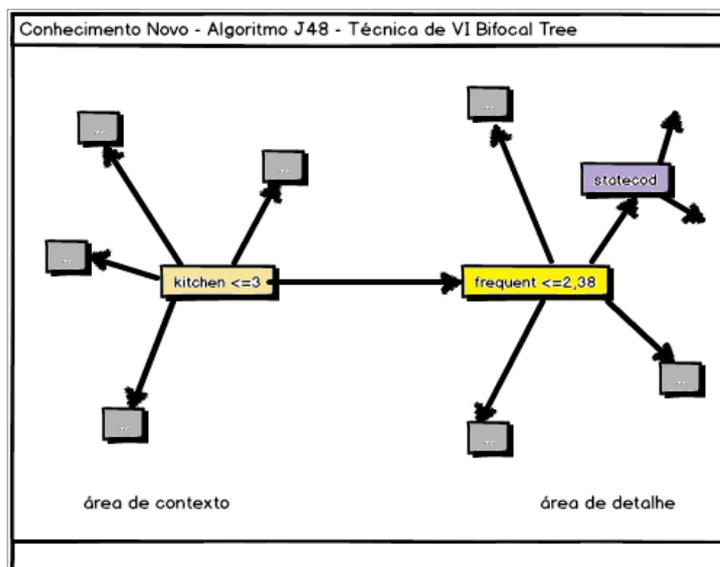


Figura 17 – Raíz “kitchen <=3” na área de contexto e o nó-filho “frequent <=2,38” na área de detalhe
Fonte: Autoria própria (2013)

Nesta figura, nota-se que o nó-filho “frequent <=2,38” torna-se nó-pai.

Selecionando-se o nó-filho “statecod” na área de detalhe, um novo deslocamento da informação para a área de contexto ocorre e novos detalhes são mostrados na área de detalhe (Figura 18).

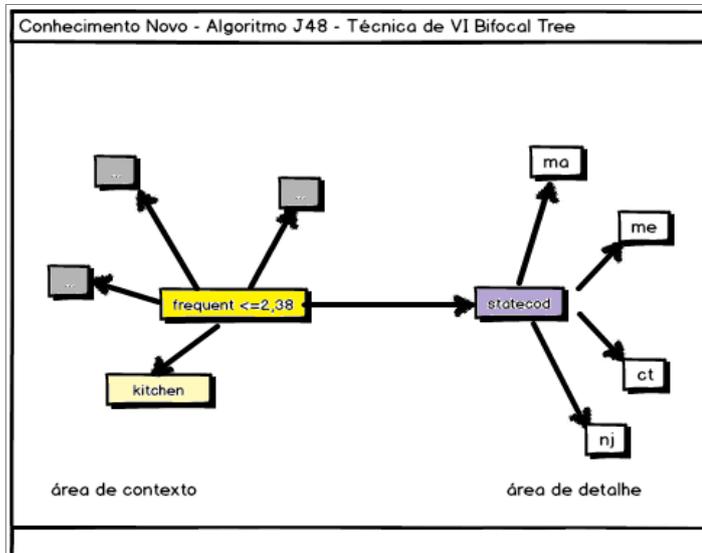


Figura 18 – Nó-pai “frequent <=2,38” na área de contexto e nó-filho “statecod” na área de detalhe
Fonte: Autoria própria (2013)

Nesta Figura identifica-se quatro nós-filhos ligados ao nó “statecod”.

Continuando o processo, selecionando o nó-filho “ma” na área de detalhe, “statecod” vai para a área de contexto e na área de detalhe os detalhes do nó-filho “ma” são expostos, conforme a Figura 19.

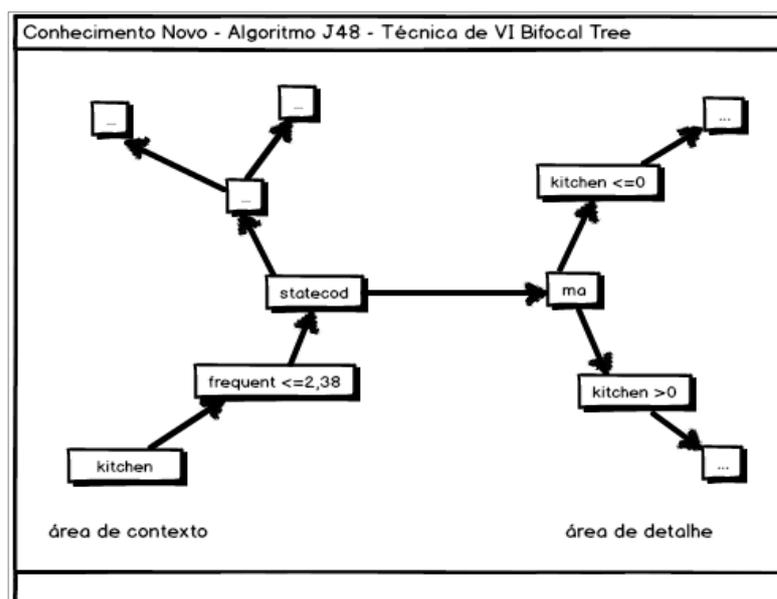


Figura 19 – Nó-pai “statecod” na área de contexto e o nó-filho na área de detalhe
Fonte: Autoria própria (2013)

A vantagem desta técnica é que a informação vai sendo agrupada a cada nó selecionado, as áreas de contexto e de detalhe permitem que o usuário guie-se até chegar a informação desejada sem que ocorra um acúmulo de informações inúteis na tela.

Outra técnica identificada como adequada para os resultados do J48 é a *Treemap*.

Esta técnica de visualização constrói a árvore dividindo o plano de exibição em retângulos de tamanhos variados um dentro do outro. Por esta técnica, cada ação do usuário clicando nos níveis da árvore de decisão modifica toda a apresentação, fazendo com que o foco de pesquisa do usuário ocupe toda tela.

A Figura 20 mostra a construção da árvore de decisão por *Treemap*.

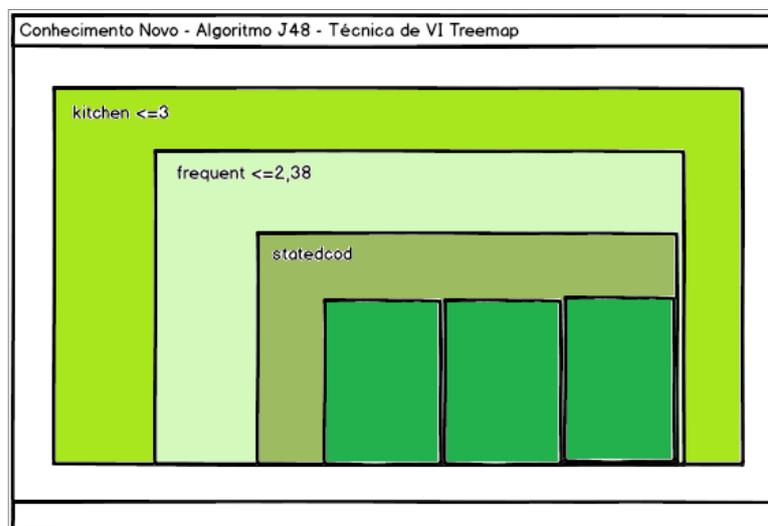


Figura 20 – Raiz “kitchen <=3” tem a maior divisão
Fonte: Autoria própria (2013)

A Figura 21 mostra a raiz “kitchen ≤ 3 ” representada pelo retângulo maior e dentro dele divisões menores pertencentes aos nós filhos “frequent $\leq 2,38$ ” e “statecod”. Pode-se usar cores para representar cada nó. Quando o usuário seleciona o nó-filho “statecod”, a apresentação muda e este se torna o nó-pai tomando o maior espaço da tela e seus nós-filhos são exibidos, como mostra a Figura 21.

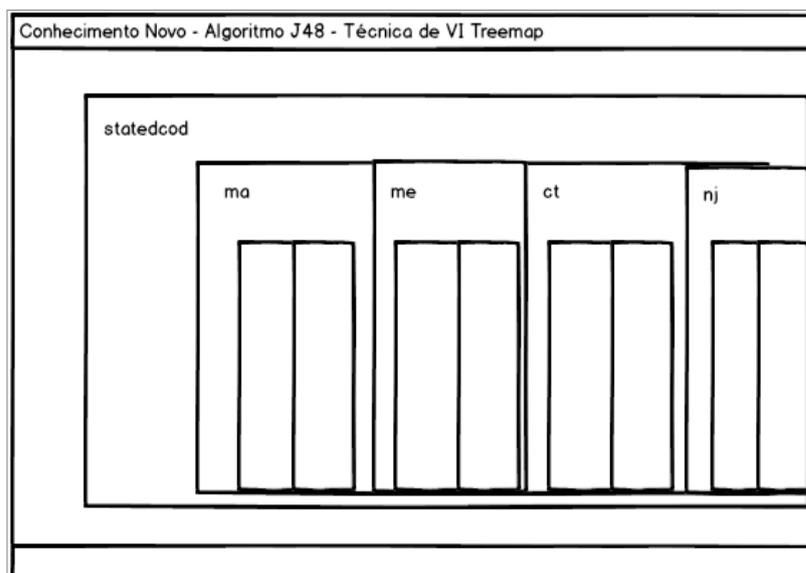


Figura 21 – Nó-pai “statecod” e seus nós-filhos
Fonte: Autoria própria (2013)

A figura mostra que o nó “statecod” possui quatro nós-filhos e estes por sua vez possuem divisões, seus nós-filhos.

Selecionando-se o nó-filho “ma”, este toma conta da tela e seu conteúdo é apresentado por outros retângulos, de acordo com a Figura 22.

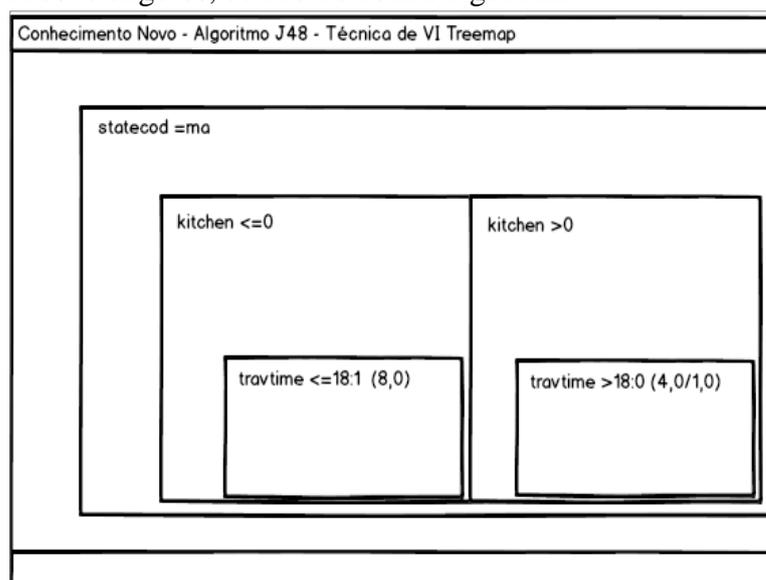


Figura 22 – Nó-pai “statecod =ma” e seus nós-filhos
Fonte: Autoria própria (2013)

Pela Figura 22, é possível notar mais detalhes das informações presentes nos subníveis da árvore de decisão.

A árvore de decisão pode ser representada aplicando outra técnica de visualização, a *Cheops*. Esta técnica utiliza o mesmo princípio das anteriores, a construção visual de uma árvore. No topo localiza-se a raiz da árvore e abaixo dela os nós filhos são alocados de acordo com suas ligações. Nesta técnica, cores podem ser utilizadas para indicar informações importantes ou de conteúdo semelhante, observe a Figura 23.

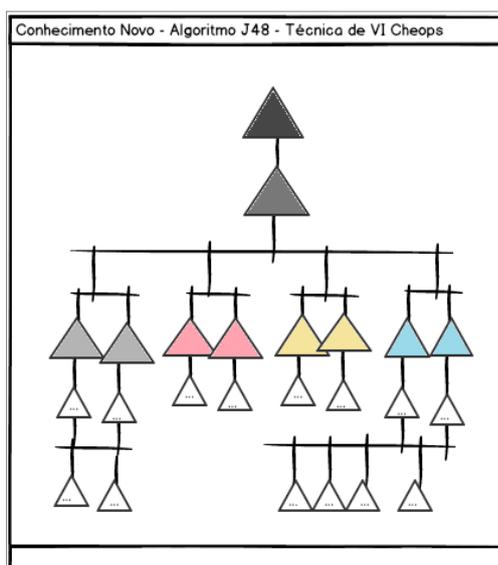


Figura 23 – Estrutura da árvore de decisão por Cheops
Fonte: Autoria própria (2013)

No caso de semelhança nas informações *Cheops* permite compactá-las em um único ícone. Se o usuário quiser saber que informações estão sendo representadas pelos ícones basta selecionar um deles e seu conteúdo irá aparecer expandido ao lado, como mostra a Figura 24.

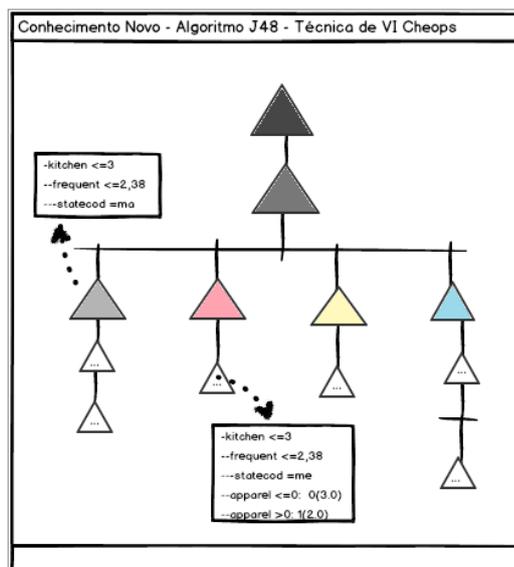


Figura 24 – Cheops, as informações compactadas aparecem com a seleção do icone
Fonte: Autoria própria (2013)

A Figura 25 mostra a mesma árvore de decisão obtida pelo algoritmo J48 sendo representada pela técnica de visualização Grafos.

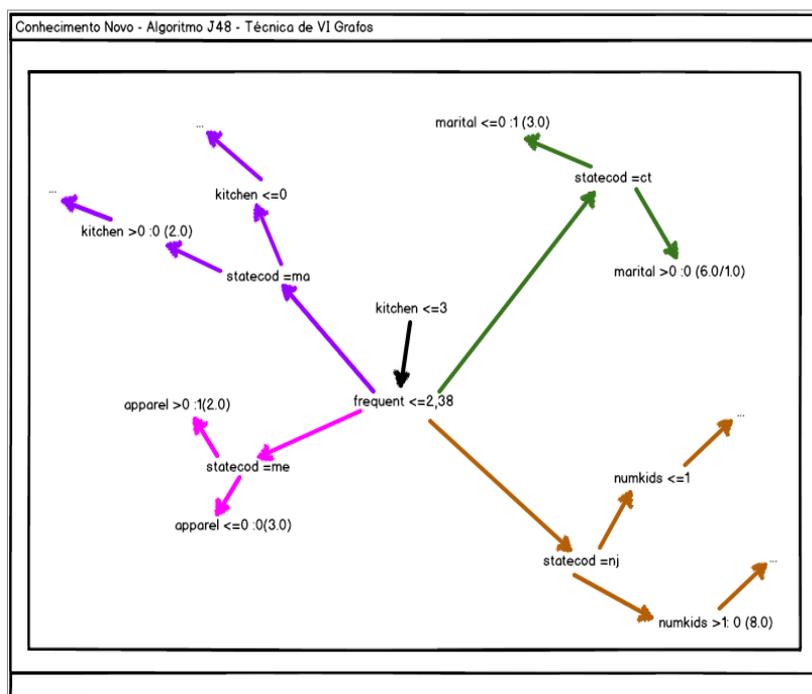


Figura 25 – Árvore de decisão pela técnica Grafos
Fonte: Autoria própria (2013)

Da raiz “kitchen ≤ 3 ”, ao centro, partem as ligações e os nós-filhos são dispostos por toda tela. Cores e setas também podem ser usadas para identificar os relacionamentos.

Information Slides é mais uma técnica de visualização que foi assinalada como capaz de atender as tarefas e métodos característicos do algoritmo J48, o esquema da representação das informações do Conhecimento Novo está na Figura 26.

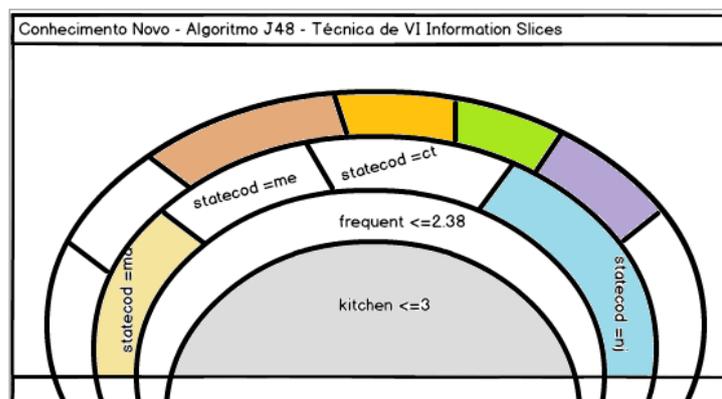


Figura 26 – Árvore de decisão pela técnica Information Slices
Fonte: Autoria própria (2013)

Esta técnica monta a árvore de decisão a partir da base e vai subindo de acordo com o número de nós-filhos existentes. Caso o usuário queira ver com mais detalhes um subnível da hierarquia, basta selecionar o nó e outro disco, semelhante ao da figura, surgirá próximo com os componentes deste subnível, conforme Figura 27.

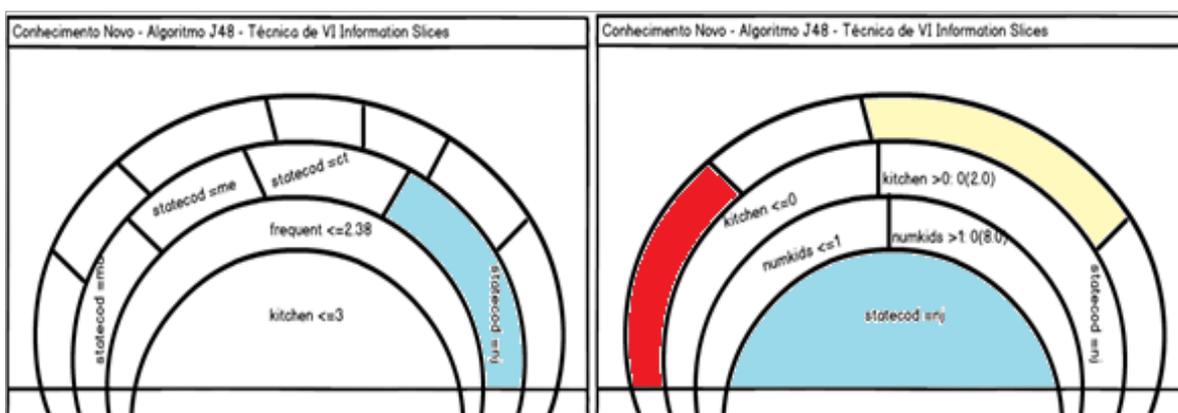


Figura 27 – Disco à direita mostra os subníveis referentes ao nó-filho selecionado
Fonte: Autoria própria (2013)

A última técnica apontada como ideal para representar informações extraídas pelo algoritmo J48 foi a *Information Cube*. A figura 28 ilustra como seria sua apresentação.

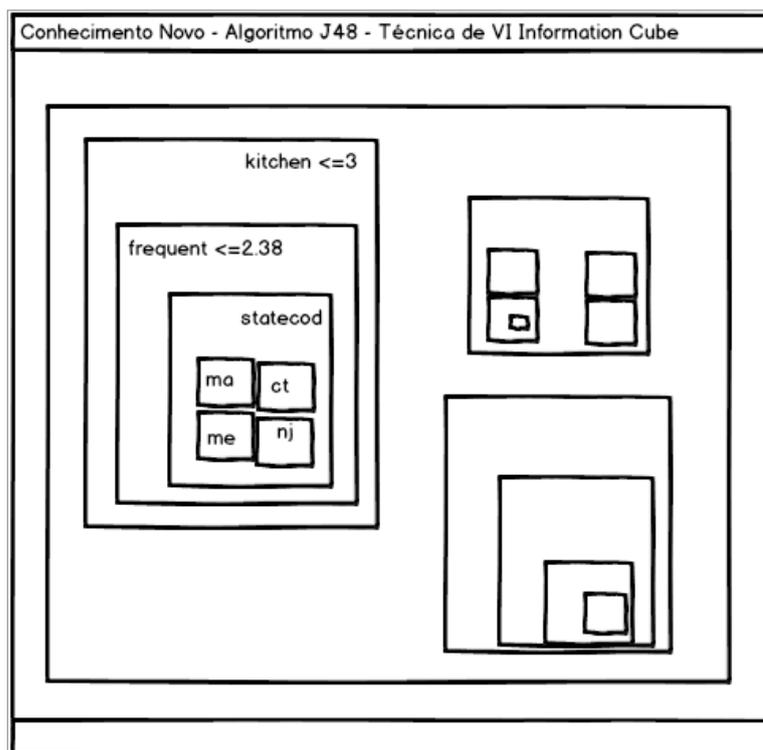


Figura 28 – Esquema de apresentação da técnica Information Cube
Fonte: Autoria própria (2013)

As informações são representadas por cubos translúcidos que mostram a hierarquia das informações sendo que os nós-filhos (cubo menor) ficam dentro do nó-pai (cubo maior).

4.5.2 Protótipo para Representação Visual do Algoritmo K-means

O conhecimento extraído pelo algoritmo *k-means* com a ferramenta *Weka*, ilustrado na Figura 8, também pode ser representado por técnicas de visualização de informações.

A Figura 29 mostra como os clusteres ou classes podem ser representados aplicando a técnica de visualização *Flip Zooming*.

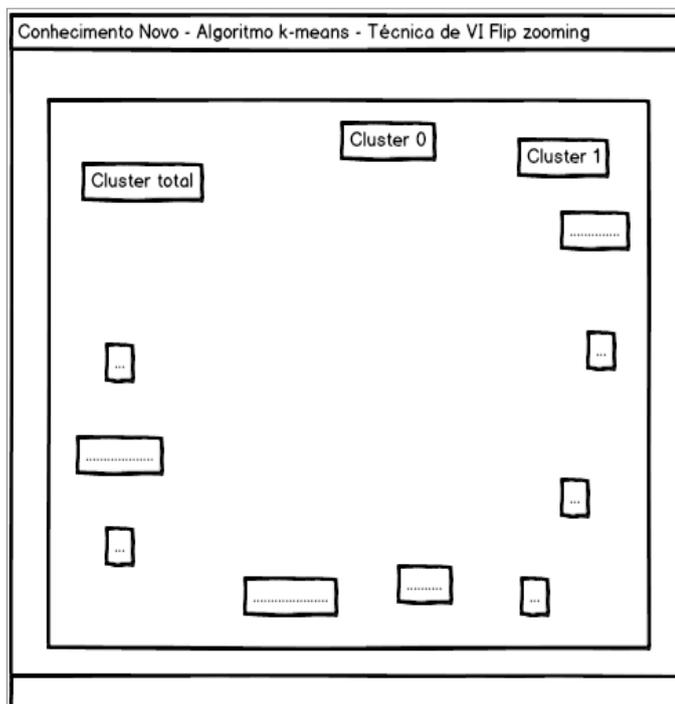


Figura 29 – Classes distribuídas na janela de exibição
Fonte: Autoria própria (2013)

Esta técnica dispõe os grupos num plano e permite que se mova uma classe para o centro para ver as informações detalhadas, conforme a Figura 31.

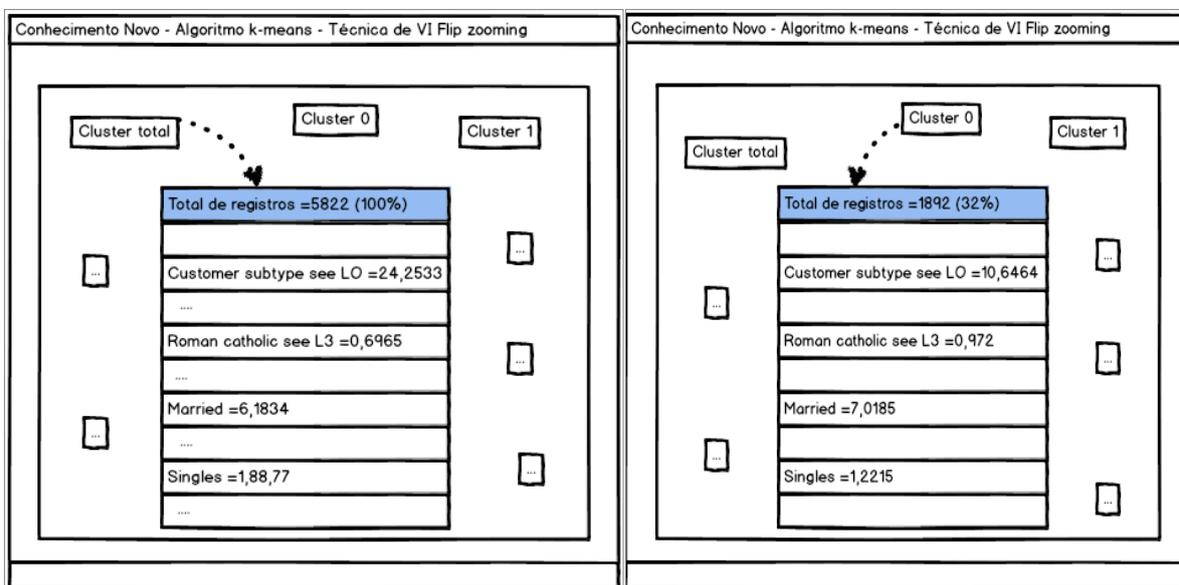


Figura 30 – Conteúdo detalhado do cluster movido para o centro
Fonte: Autoria própria (2013)

A Figura 31 mostra como as classes podem ser representadas aplicando a técnica de visualização Coordenadas Paralelas.

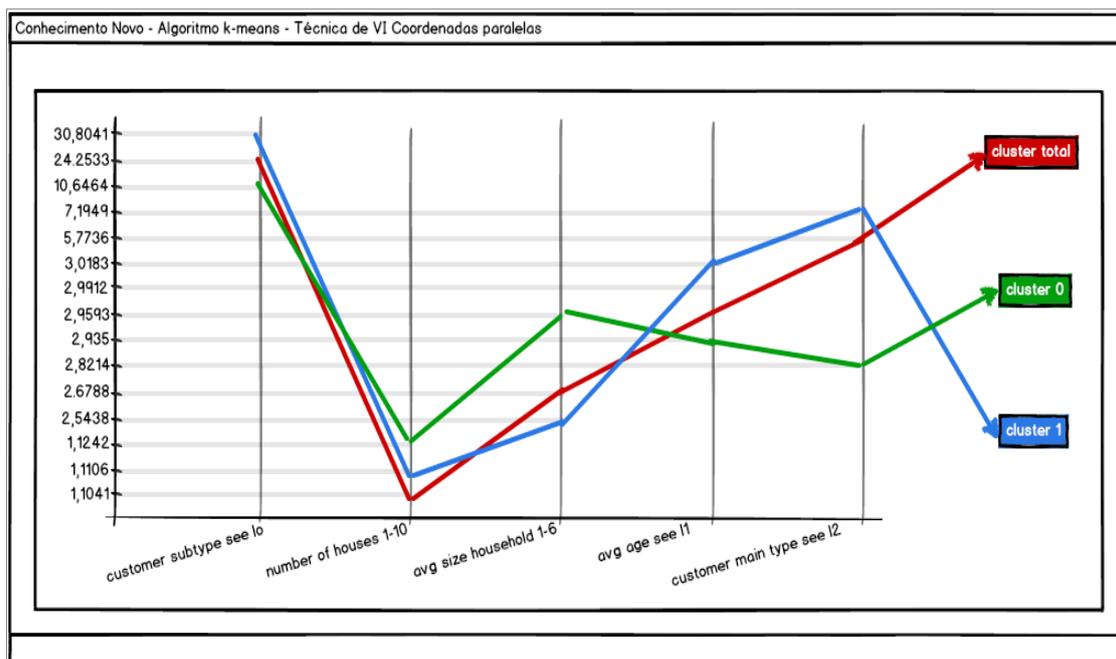


Figura 31 – Classes representadas pela técnica Coordenadas Paralelas
Fonte: Autoria própria (2013)

Esta técnica permite diferenciar as informações de cada classe por cor, também. A linha de cor igual indica as informações pertencentes a cada grupo.

A aplicação da técnica de visualização Grafos na representação das classes extraídas pelo *k-means* é ilustrada na Figura 32.

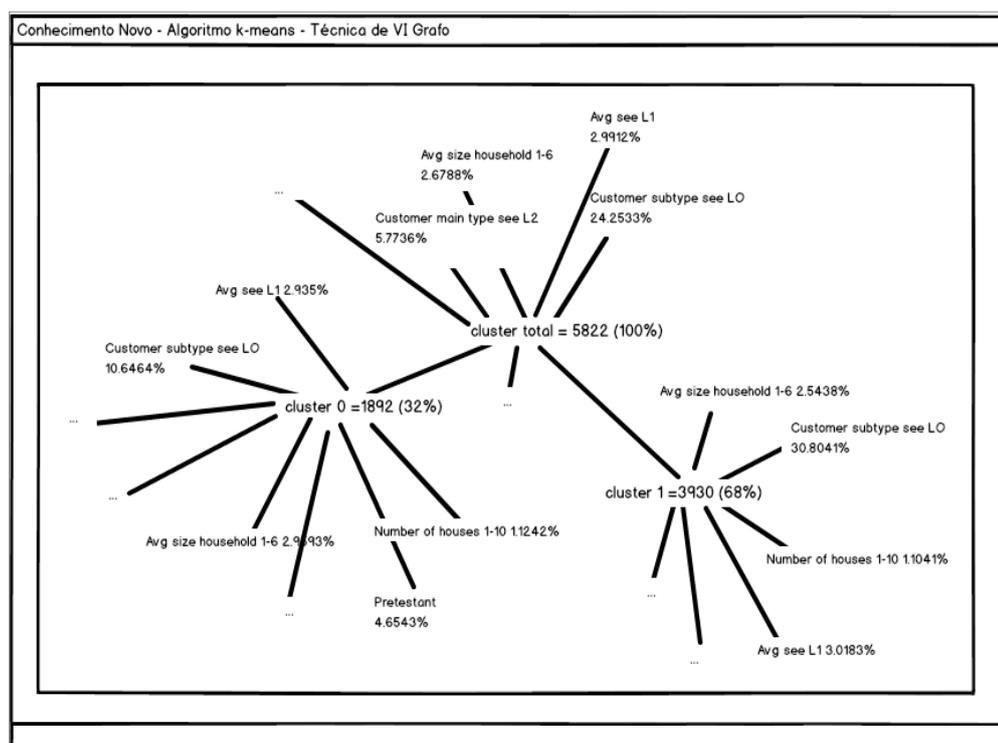


Figura 32 – Conhecimento Novo representado pela técnica Grafos
Fonte: Autoria própria (2013)

Pela figura é possível visualizar as classes totalmente, mas os grafos também podem ser construídos individualmente para cada classe, se for preciso.

Classes também podem ser representadas pela técnica de visualização *Treemap*. A Figura 33 mostra o esquema de representação dos grupos de *clusters* por divisão de tela.

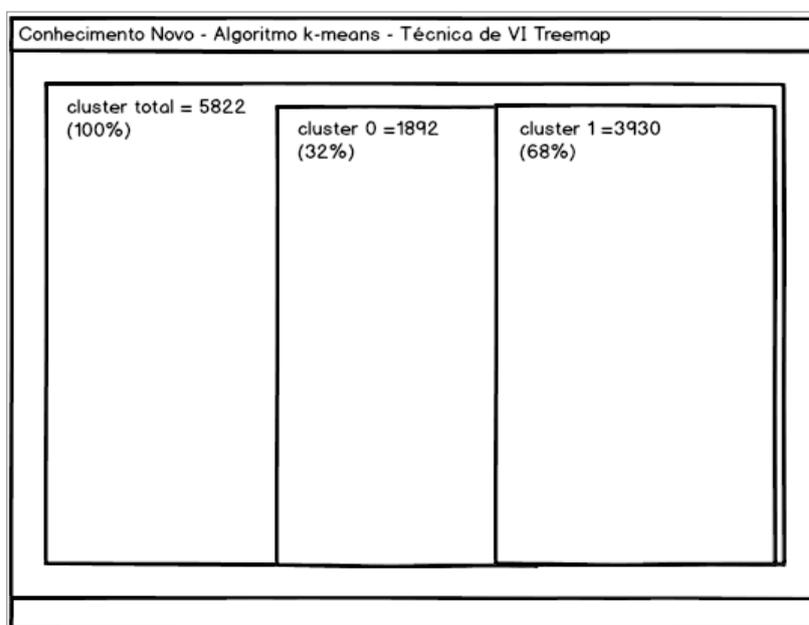


Figura 33 – Representação das clusters do algoritmo K-means por Treemap
Fonte: Autoria própria (2013)

O *cluster* total é representado pelo retângulo maior. Dentro, os demais clusters são representados por retângulos de tamanho diferentes que indicam maior porcentagem de resultados.

Selecionando-se *cluster* Zero, o desenho da apresentação modifica e o conteúdo do *cluster* fica evidente, como a Figura 34 mostra.

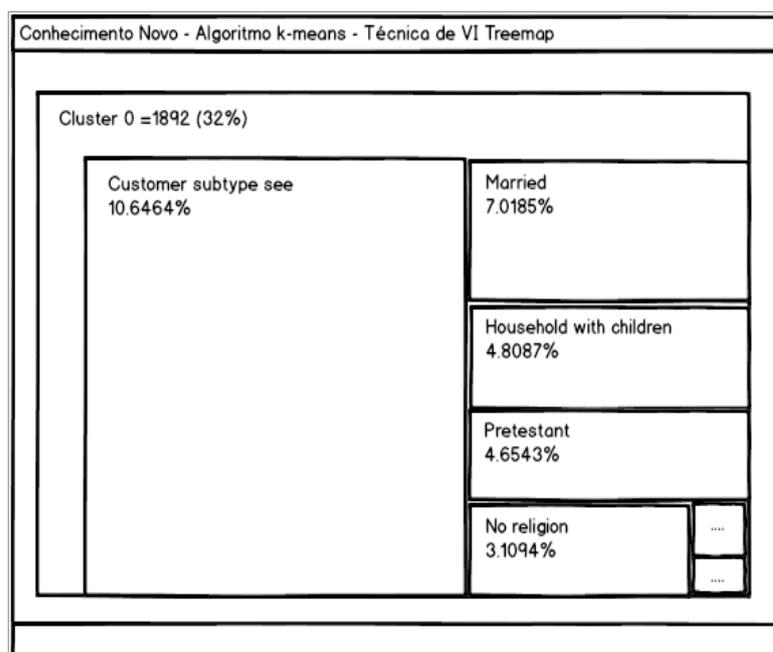


Figura 34 – Representação das clusters do algoritmo K-means por Treemap
Fonte: Aatoria própria (2013)

A Figura 35 demonstra como os clusters podem ser apresentados pela técnica orientada a pixel.

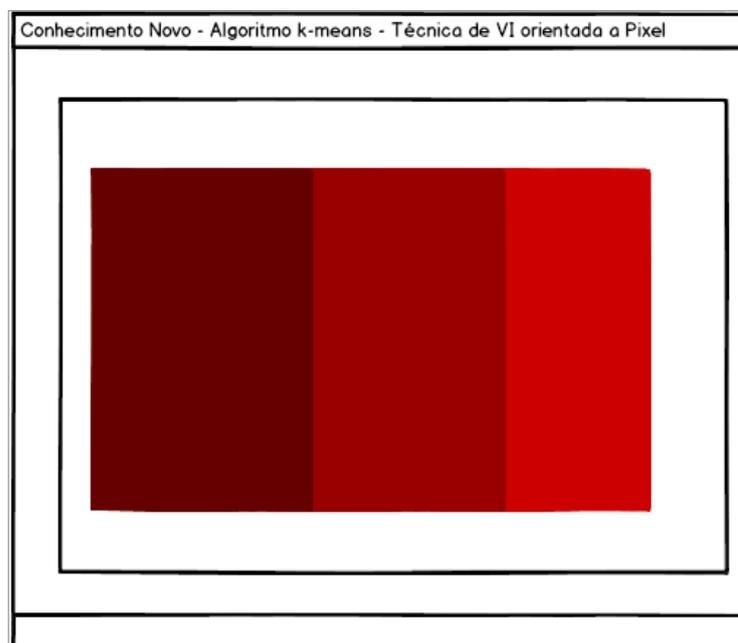


Figura 35 – Pixéis de cores diferentes para classes diferentes
Fonte: Aatoria própria (2013)

A figura ilustra a substituição dos atributos de informação por pixeis coloridos e agrupados. O tamanho do bloco de pixeis dependerá da quantidade de informações pertencente em cada classe. É possível, também representar cada atributo por um pixel de cor diferente, como mostra a Figura 36.

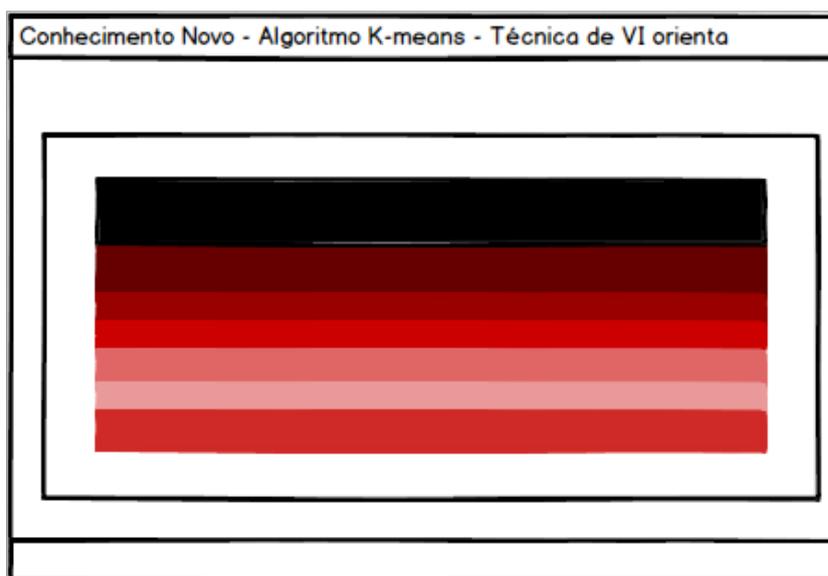


Figura 36 – Cada atributo de um cluster é representado por um pixel colorido
Fonte: Autoria própria (2013)

A espessura da faixa vai depender do valor do atributo. Esta técnica permite verificar o tamanho que o atributo ocupa dentro do *cluster*, seja devido a espessura ou a tonalidade do pixel.

4.5.3 Protótipo para Representação Visual do Algoritmo Apriori

A Figura 10 mostra as regras de relação de dependência gerada pela técnica Apriori com a ferramenta *Weka*.

Um exemplo da aplicação da técnica Figura de Arestas na representação das relações de dependência, das regras de associação é ilustrada na Figura 37.

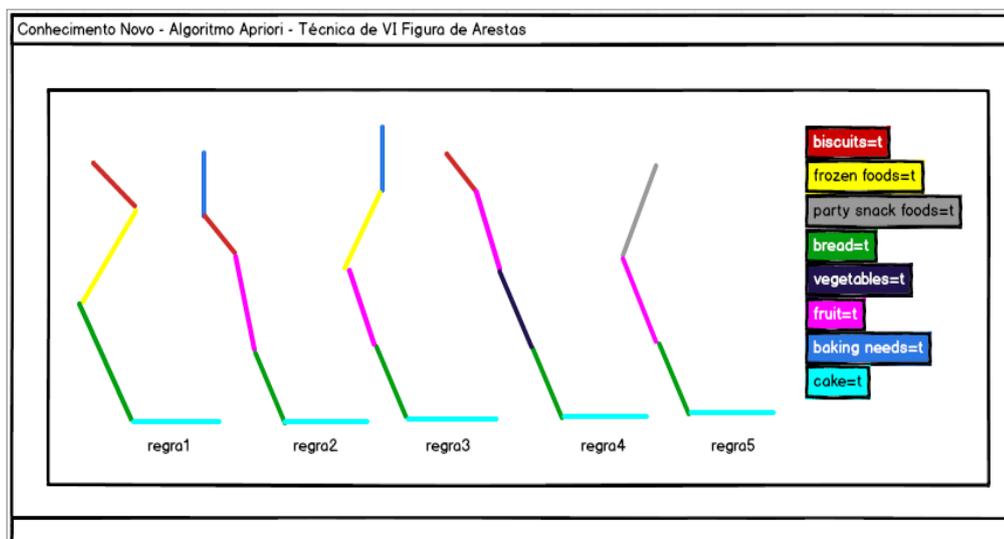


Figura 37 – Relações de dependência por cores na Figura de Arestas
 Fonte: Autoria própria (2013)

O desenho da técnica leva em consideração a relação indicada.

Outra forma de representar as regras de associação é por meio da técnica de visualização Grafos, onde cores e setas podem ajudar na compreensão das relações de dependências.

A Figura 38 exemplifica a aplicação desta técnica.

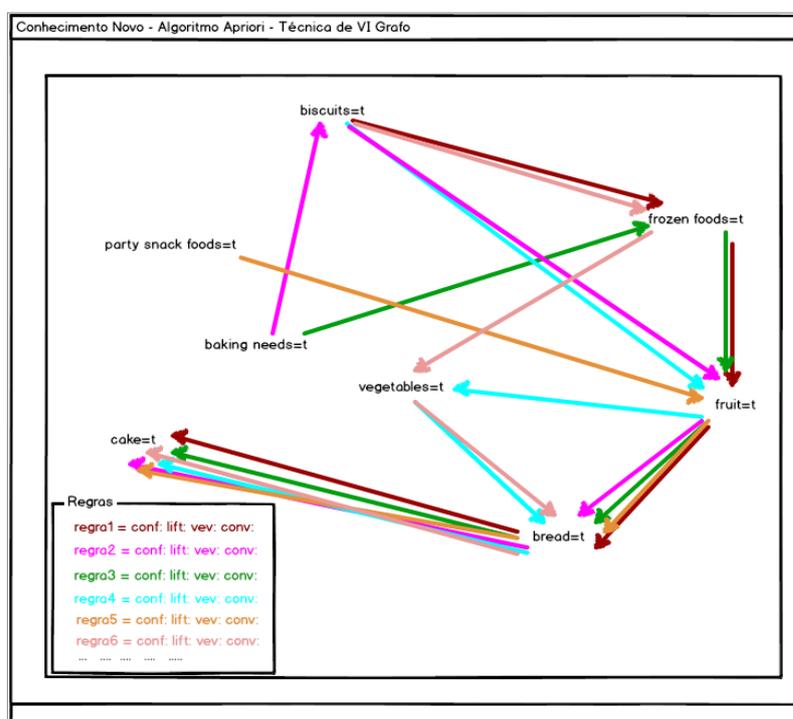


Figura 38 – Relação de dependência representada pela técnica Grafo
 Fonte: Autoria própria (2013)

4.6 CARACTERÍSTICAS RELEVANTES DO USO DAS TÉCNICAS DE VISUALIZAÇÃO EM UMA FERRAMENTA DE DESCOBERTA DE CONHECIMENTO

É notório o problema que pode se transformar a interpretação de informações depois do processo de mineração de dados pela ferramenta *Weka*. O formato utilizado por esta ferramenta pode tornar lento o entendimento, absorção e utilização do conhecimento exposto. Num ambiente organizacional a demora na captação e disseminação de informações relevantes pode significar prejuízo ao planejamento estratégico. Diante desta situação, os administradores precisam buscar meios que tornem viável a aplicação das informações nas decisões cotidianas.

Este capítulo consistiu na identificação das informações constantes nos resultados minerados e na construção de uma correlação destas informações com as estruturas fornecidas pelas técnicas de visualização de informações. Este processo foi exemplificado na apresentação de protótipos das telas das técnicas de visualização que se mostraram mais adequadas para as informações fornecidas pelos algoritmos de mineração na divisão denominada *Conhecimento Novo*.

Se comparadas as apresentações das informações constantes no formato tradicional da ferramenta *Weka* (verificadas nas figuras 5, 8 e 11 das subseções 4.2.1, 4.2.2 e 4.2.3, respectivamente) com os modelos construídos com elementos visuais é possível notar uma melhora na captação das informações geradas pela mineração de dados, ou seja, o conhecimento ficou mais fácil de ser entendido. As relações, grupos e relevância das informações foram mais rapidamente notados.

A capacidade de reduzir ou expandir a carga informativa e focar-se em apenas uma informação por vez, ou seguir apenas um relacionamento se mostrou uma vantagem na utilização das técnicas *Flip Zooming*, *Cone Tree*, *TreeMap*, *Cheops*, *Information Slices*, *Bifocal Tree* e Grafos, que possuem esta capacidade. Além disso, cores se mostraram auxiliadoras na compreensão das informações de desigual relevância ou na distinção proposital tanto nas técnicas citadas anteriormente quanto nas Coordenadas Paralelas, Orientada a Pixel e Figura de Arestas.

Os relacionamentos e padrões representados por árvores, modelos de conhecimento gerados pelo algoritmo J48, são melhores visualizadas por sete técnicas de natureza hierárquica.

Quando os grupos e padrões são representados por classes (modelos de conhecimento característicos do algoritmo *k-means*), as duas técnicas da categoria geométrica, as orientadas a pixel, duas iconográficas e três da hierárquica se mostraram mais adequadas.

Para os modelos de conhecimento extraídos pelo algoritmo Apriori que são regras de associação e padrões por meio da relação de dependência, duas técnicas (uma iconográfica e uma hierárquica) foram verificadas como as melhores para este tipo.

Os protótipos apresentados, por serem construções simples, não contem todos os recursos inerentes às técnicas de visualização de informações; por exemplo, a técnica *Information Cube*, ainda pode ter sua projeção em modo tridimensional tornando ainda mais amigável a representação do conhecimento.

No entanto, é possível perceber que não se faz necessário conhecer os registros profundamente, pois na representação visual já se nota a natureza das informações.

5 CONCLUSÃO

O objetivo deste trabalho foi apresentar uma correlação entre as informações e os formatos de apresentação de conhecimento extraído pelas técnicas e algoritmos escolhidos (classificação-*J48*, clusterização-algoritmo *k-means*, associação-algoritmo *apriori*) implementados pela ferramenta *Weka* com a estrutura de gráficas das técnicas de visualização de informações. Apesar de terem sido apenas um algoritmo de cada técnicas de mineração, nada impede que se utilize desta correlação como base para implementação em outros algoritmos e em outras ferramentas, guiando a escolha de recursos visuais para outras representações.

A mineração de dados se apresenta como principal meio para busca de informações e de grande valia no processo de gestão e tomada de decisão. No entanto, seus desenvolvedores precisam estar atentos para não focarem apenas na funcionalidade e esquecer-se de tornar estas ferramentas mais fáceis de utilização por parte do usuário, pois, problemas ligados a interpretação podem desmotivar e afastar os potenciais usuários.

Uma forma de contornar esta situação é a aplicação de recursos visuais familiares ao usuário na substituição de texto por símbolos.

A área de Visualização de Informações torna mais acessível e fácil a interpretação de informações, deixando o usuário mais dedicado ao problema de pesquisa e menos ao entendimento de como funciona a ferramenta.

A Visualização de Informações pode ser aplicada na seleção dos dados antes da mineração, durante o processo de mineração ou no final quando ocorre a interpretação dos resultados. Neste trabalho deu-se ênfase na aplicação nos resultados obtidos com a mineração de dados executada pela ferramenta de descoberta de conhecimento *Weka*.

Para isso foi necessário identificar os modelos de conhecimento (tarefas e métodos) gerados pelas técnicas de mineração e verificar quais eram realizadas pelos algoritmos *J48*, *K-means* e *Apriori*. Para tornar mais fácil o estudo das informações o resultado fornecido pela *Weka* foi dividido em Cabeçalho, Conhecimento Novo e Informações Adicionais.

Depois disso, foram analisadas as técnicas de visualização e verificadas quais suportavam as tarefas e métodos estabelecidos anteriormente. Com isso foram analisadas as técnicas tendo como resultados: *Information Cube*, *Cone Tree*, *TreeMap*, *Cheops*, *Information Slices*, *Bifocal Tree*, Grafos para o algoritmo *J48*; Coordenadas Paralelas, Gráfico de

Dispersão, Pixel, Glifo em Estrela, Figura de Aresta, *FlipZooming*, *TreeMap* e Grafos para o algoritmo *K-means*, e; Figura de Arestas e Grafos para o algoritmo *Apriori*.

Finalmente, foram construídos telas de exibição com o objetivo de exemplificar a representação do conhecimento presente em Conhecimento Novo dos algoritmos e foi possível constatar a melhoria na apresentação e percepção das informações exibidas.

Este trabalho contribui para o grupo de pesquisa (GPSI, 2013) ao realizar experimentos com o objetivo de mostrar a vantagem de se aplicar recursos gráficos na compreensão da informação.

Importante salientar que não há uma técnica que satisfaça todas as expectativas de uma visualização de informações completa, mas é recomendável que ocorram combinações entre as técnicas de visualização para atingir o máximo de interpretação do resultado.

5.1 TRABALHOS FUTUROS

Outros trabalhos podem ser desenvolvidos a partir do tema abordado nesta pesquisa, são eles:

- Aplicação de técnicas de visualização em resultados gerados por outros algoritmos de mineração.
- Medição do tempo de percepção e de ação dos administradores após a utilização de recursos visuais para representação de situações-problema.
- Existem muitas ferramentas de descoberta de conhecimento, a *Weka* foi alvo deste estudo, portando há ainda outras que podem ser analisadas com mesmos objetivos.
- Construção e implementação prática das técnicas de visualização em ferramentas de mineração de dados.
- Aperfeiçoamento dos recursos gráficos presentes nas técnicas de visualização de informações.
- Incorporação dos protótipos na ferramenta que está sendo desenvolvida pelo GPSI, a saber: Framework de Formação de Preço de Venda.

REFERENCIAS

ALEXANDRE, Dulclerci Sternadt; TAVARES, João Manuek R.S. Factores de percepção visual humana na visualização de dados. In: CONGRESSO INTERNACIONAL EM MÉTODOS NUMÉRICOS EM ENGENHARIA E CONGRESSO IBERO LATINO-AMERICANO SOBRE MÉTODOS COMPUTACIONAIS EM ENGENHARIA (CMNE/Cilamce). Porto, 2007. Disponível em: <<http://repositorio-aberto.up.pt/bitstream/10216/357/2/13662.pdf>>. Acesso em: 15 out. 2012.

ANDREWS, Keith; HEIDEGGER, Helmut. Information Slices: Visualising and Exploring Large Hierarchies using Cascading, Semi-Circular Discs. Late Breaking Hot Topic Paper. IICM, Graz University of Technology, Austria. **IEEE Symposium on Information Visualization**. Research Triangle Park, North Carolina, Oct. 1998. Disponível em: <www.iicm.tugraz.at/liberation/iicm.../ivis98.pdf>. Acesso em: 24 abril 2012.

BALSAMIQ. 2013. **Ferramenta de Prototipação**. Disponível em: <www.build.balsamiq.com>. Acesso em 01 jun. 2013.

BARBOSA, Denise Chaves Carvalho; MACHADO, Maria Augusta. Mineração de Dados usando o Software WizRule em Base de Dados de Compras de TI. **Revista Eletrônica de Sistemas de Informação**. Curitiba, v. 10, p. 37-57, 2007. Disponível em: <<http://revistas.facecla.com.br/index.php/reinfo/article/download/184/93>>. Acesso em: 24 set. 2012.

BATISTA, Émerson de O. **Sistemas de Informação: o uso consciente da tecnologia para o gerenciamento**. p.296. 8 ed. São Paulo: Saraiva, 2006.

BOSCARIOLI, Clodis; TABUSADANI, Fernando Yukio; BIDARRA, Jorge. O uso integrado de K-NN e Scatter Plots 2D na mineração visual de dados. **Publicatio UEPG Ciências Exatas e da Terra, Ciências Agrárias e Engenharias (Impresso)**. Ponta Grossa, v. 14, p. 187-197, dez. 2008. Disponível em:
<<http://www.revistas2.uepg.br/index.php/exatas/article/view/944/779>>. Acesso em: 24 set. 2012.

BURNHAM, Teresinha F.; ALVES, Renato M.; MORAES, Isabel O. de; MORAES, Ramone L. de. Aprendizagem organizacional e gestão do conhecimento. **VI Encontro Nacional de Ciência da informação**. Salvador, 2005. Disponível em:
<<http://www.repositorio.ufba.br/ri/handle/ri/3876>>. Acesso em: 09 abril 2011.

CARDOSO, Olinda N. P.; MACHADO, Rosa T. M. Gestão do conhecimento usando data mining: estudos de caso na Universidade Federal de Lavras. **Revista de Administração Pública**. Fundação Getúlio Vargas. Rio de Janeiro, v. 42, n. 3, p. 495-528, maio/jun. 2008.

CASTANHEIRA, Luciana Gomes. Aplicação de Técnicas de Mineração de Dados em Problemas de Classificação de Padrões. 2008. 95f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais. Bel Horizonte, 2008. Disponível em:
<www.cpdee.ufmj/documentos/Defesas/777/Dissertacao_LucianaCastanheira.pdf>. Acesso em: 12 out 2012.

CAVA, Ricardo A.; LUZZARDI, Paulo R. G.; FREITAS, C. M. D. S. The Bifocal Tree: A technique for the visualization of hierarchical information structures. In: WORKSHOP ON HUMAN FACTORS IN COMPUTER SYSTEM. IHC, 2002, Fortaleza, CE. Anais... Fortaleza: SBC, 2002. Disponível em:
<http://www.researchgate.net/publication/2833884_The_Bifocal_Tree_a_Technique_for_the_Visualization_of_Hierarchical_Information_Structures/file/9fcfd50603d582810f.pdf>. Acesso em: 25 abril 2012.

CIS. **Computer and Information Science: Sample Weka Data Sets**. Disponível em: <http://storm.cis.hordham.edu/~gweiss/data-mining/datasets.html>. Acesso em: 02 mar. 2013.

DARIO, Dineide. **Aplicação de Técnicas de Visualização de Informação em Ferramentas para Apoio à Avaliação Formativa em Sistemas de EaD**. 2010. 124f. Dissertação (Mestrado em Ciência da Computação), Universidade Estadual de Campinas. Campinas, 2010. Disponível em: <http://www.bibliotecadigital.unicamp.br/document/?code=000783707>>. Acesso em: 12 out. 2012.

DIAS, Mateus P. **A Contribuição da Visualização da Informação para a Ciência da Informação**. 2007. 126f. Dissertação (Mestrado em Ciência da Informação) – Pontifícia Universidade Católica de Campinas. Campinas, 2007. Disponível em: http://www.bibliotecadigital.puc-campinas.edu.br/tde_arquivos/2/TDE-2007-08-02T080844Z-1353/Publico/Mateus%20Pereira%20Dias.pdf>. Acesso em: 24 abril 2012.

DIAS, Mateus P.; CARVALHO, José O. F. de. A visualização da informação e a sua contribuição para a Ciência da Informação e Biblioteconomia. **Revista Pesquisa Brasileira em Ciência da Informação e Biblioteconomia (PBCIB)**. Rio de Janeiro, v. 8, n. 5. 2007. Disponível em: <http://periodicos.ufpb.br/ojs/index.php/abcib/article/view/6109>>. Acesso em: 22 abril 2012.

DIAS, Maria. M.; YAMAGUCHI, Juliana; RABELO, Emerson; FRANCO, Clécia. Visualization Techniques: Which is the Most Appropriate in the Process of Knowledge Discovery in Data Base?. Capítulo de livro. **Advances in Data Mining Knowledge Discovery and Applications, Associate Prof. Adem Karahoca (Ed.)**, ISBN: 978-953-51-0748-4, InTech, DOI: 10.5772/50163. Available from: <http://www.intechopen.com/books/advances-in-data-mining-knowledge-discovery-and->

[applications/visualization-techniques-which-is-the-most-appropriate-in-the-process-of-knowledge-discovery-in-data](#)>. Acesso em 12 out. 2012.

DOMINGUES, Marcos Aurélio. Generalização de regras de associação. 171f. 2004. Dissertação (Mestrado em Ciências da Computação e Matemática Computacional) – Universidade de São Paulo, São Carlos, 2004. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-10082004-154242/publico/dissertacao.pdf>> Acesso em: 04 nov. 2012.

ESTIVALET, Luiz F. **O Uso de Ícones na Visualização de Informações**. 2000. 90f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/1629>>. Acesso em: 13 maio 2012.

FARIA, Ana Cristina; LAUDELINO, Julien Ariani de Souza; DOMINGUES, Maria José Carvalho Souza. A relevância do sistema de informação em pequena empresa do segmento de construção civil. In: III SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA. **Anais...** Rio de Janeiro, v.01, p.01-09, out. 2006.

FONSECA, Eva Fabiani de Mello; GARCIA, Osmarina Pedro Garcia. O Sistema de Informação Gerencial e sua Importância no Desenvolvimento das Empresas. **Revista Ciências Sociais em Perspectiva**. Cascavel, v. 6, n. 11, p. 11-25, 2sem. 2007. Disponível em: <<http://xa.yimg.com/kq/groups/20771158/1296670760/name/O+SIG+E+SUA+IMPORTAN CIA+PARA+AS+EMPRESAS.pdf>>. Acesso em: 09 ago. 2011.

FREITAS, Carla M. D. S.; CHUBACHI, Olinda M.; LUZZARDI, Paulo R. G.; CAVA, Ricardo A. Introdução à Visualização de Informações. **Revista de Informática Teórica e Aplicada**. Porto Alegre, v. 8, n.2, p.143-158. 2001. Disponível em:

<<http://www.lume.ufrgs.br/bitstream/handle/10183/19398/000300210.pdf?sequence=1>>.

Acesso em: 09 ago 2011.

GALVÃO, Noemi D.; MARIN, Heimar de F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paul Enferm (online)**. São Paulo, v. 22, n. 5, p. 686-690. set/out, 2009.

Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-21002009000500014&lng=en&nrm=iso>. Acesso em: 12 abril 2012.

GOMES, Leandro F. O.; TAVARES, João Manuel R. S. Percepção humana na visualização de grandes volumes de dados. In: CONGRESSO IBERO-AMERICANO DE ENGENHARIA MECÂNICA - CIBEM. **Anais...** Porto, 2011. Disponível em:

<http://paginas.fe.up.pt/~tavares/downloads/publications/artigos/CIBEM_2011_Leandro.pdf>. Acesso em 13/09/2012.

GONÇALVES, Bruno dos Santos; PICELLI, Haeliton; GONÇALVES, Leandro dos Santos; GERBONI, Mateus da Silva. **Curso de Sistemas de Informações: aplicação de técnicas de data mining na bolsa e valores**. 2012. União das Instituições Educacionais do Estado de São Paulo: Hortolândia, 2011, 43p. Disponível em: <<http://pt.scribd.com/doc/73215706/Data-Mining-Final>>. Acesso em: 14 set. 2012.

GOLDSCHMIDT, Ronaldo. PASSOS, Emmanuel. **Data Mining: Um Guia Prático – conceitos, técnicas, ferramentas, orientações e aplicações**. 4. ed. Elsevier: Rio de Janeiro, 2005. 255p.

GOUVEIA, Luiz B.; RANITO, João. **Sistemas de informação de apoio à gestão**. Coleção Inovação e governança nas autarquias. Porto: SPI, 2004. 96p. Disponível em:

<www2.spi.pt/inovaut/docs/Manual_VII.pdf>. Acesso em: 13 set. 2012.

GPSI. **Grupo de Pesquisa em Sistemas de Informação**. Departamento de Sistemas de Informação. Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2013.

HOFFMANN, Rosa Cristina. OLIVEIRA, Patricia Santos Marcondes de; ZEFERINO, Renato Zanelato. A Utilização Estratégica dos Sistemas de Informações Gerenciais no Ramo Hoteleiro da Cidade de Ponta Grossa – Paraná. **Revista de Engenharia e Tecnologia**, v. 4, n. 1, p. 18-28, abril 2012. Disponível em: <<http://www.revistaret.com.br/ojs-2.2.3/index.php/ret/article/view/49/135>>. Acesso em: 02 fev. 2013.

MARTINS JUNIOR, Walter; JUNIOR, Pedro Paulo de Andrade; RESENDE, Luis Mauricio; ALMEIDA, Leonardo Ferreira Pedrosa de. A Importância dos Sistemas de Informações Gerenciais em Arranjos Produtivos Locais. In: II CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO. **Gestão do Conhecimento nas Engenharias**. Ponta Grossa, nov. 2012.

KHAN, Muzammil; KHAN, Sarwar Shah. Data and Information Visualization Methods, and Interactive Mechanisms: A Survey. **International Journal of Computer Applications**. New York, v. 34, n. 1, p. 01-14, nov. 2011. Disponível em: <www.research.ijcaonline.org/volume34/number1/pxc3875722.pdf>. Acesso em: 13 set. 2012.

KEIM; Daniel; WARD, Matthew. **Visual Data Mining Techniques**. University of Konstanz, Germany and Worcester Polytechnic Institute. Intelligent Data Analysis: An Introduction. Berlin. **Springer**, 2002. Disponível em: <<http://kops.uni-konstanz.de/volltexte/2009/6984/>>. Acesso em: 05 abril 2012.

LOH, Stanley. **Manual do Software WizRule para Data Mining**. Disponível em: <<http://paginas.ucpel.tche.br/~loh/textos.htm>>. Acesso em 20 nov. 2009.

LUZZARDI, Paulo R. G. **Cr terios de avalia o de t cnicas de visualiza o de informa es hier rquicas**. 2003. 127f. Tese (Doutorado em Ci ncia da Computa o) – Programa P s-Gradua o em Computa o, Instituto de Inform tica, Universidade Federal do Rio Grande do Sul. Porto Alegre. 2003. Dispon vel em: <www.lume.ufrgs.br/bitstream/handle/10183/4764/000414947.pdf?sequence=1> Acesso em: 05 abril 2012.

MAFRA, Felipe N.; COSTA, Dayane M.; FERNANDES, Leandro A. C.; AQUINO JR, Pl nio T. Compositor de personagens fict cios para caracteriza o de usu rios. In: CONFER NCIA IADIS IBERO-AMERICANA. **Anais...** WWW/Internet, IADIS, p. 289-296, 2008. Dispon vel em: <www.iadis.net/dl/final_uploads/200819L036.pdf>. Acesso em: 05 abril 2012.

MEGAPUTER. **PolyAnalyst 6**. Megaputer Intelligence Inc. USA, 2007. Dispon vel em: <http://www.megaputer.com/site/down/PolyAnalyst_6_brochure.pdf>. Acesso em: 11 maio 2012.

MORAES, Andr  F.; BASTOS, Lia C.; BITTENCOURT, Rog rio G. Data Mining: Uma Experi ncia em Cadastro T cnico Multifinalit rio Urbano. In: II CONGRESSO BRASILEIRO DE COMPUTA O. **Anais...** Itaja , v. 1, p. 122-129, Itaja , 2002. Dispon vel em: <<http://www.cbcomp.univali.br>>. Acesso em: 19 abril 2012.

NASCIMENTO, Hugo A. D. do; FERREIRA, Cristiane B. R. Visualiza o de Informa es: uma abordagem pr tica. In: XXV CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTA O. **Anais...** S o Leopoldo, 2005. Dispon vel em: <www.lbd.dcc.ufmg.br/colecoes/jai/2005/002.pdf>. Acesso em: 05 abril 2012.

NETO, Marcos A. S.; VILLWOCK, Rosangela; SCHEER, S rgio; STEINER, Maria T. A.; DYMINSKI, Andr  S. T cnicas de minera o visual de dados aplicados aos dados de

instrumentação da barragem de Itaipu. **Gestão e Produção**. São Carlos, v. 17, n. 4, p. 721-734. set/out. 2010. Disponível em:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-530X2010000400007&lng=en&nrm=iso. Acesso em: 14 ago. 2012.

NETO, Alvim A. de O. **IHC – Interação Humano Computador; Modelagem e Gerência de Interfaces com o Usuário**. Florianópolis: Visual Books, 2008. 120p.

NETO, Marco A. S.; BURIOL, Tiago Martinuzzi ; VILLWOCK, Rosangela ; SCHEER, Sérgio ; STEINER, Maria Terezinha Arns ; DYMINSKI, Andrea Sell . Técnicas de Mineração Visual de Dados Aplicadas ao Monitoramento Estrutural da Barragem de Itaipu. In: XXIX CILAMCE (IBERIAN LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING), 2008, Maceió, AL. **Anais...**, v. 1, p. 1-17, 2008.

NEVES, José M. S. das. A implantação de tecnologias da informação como fator de competitividade nos sistemas produtivos e nos negócios. In: XIII SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO. **Anais...** Bauru, nov. 2006. Disponível em: www.simpep.feb.unesp.br/anais/anais_13/artigos/251.pdf. Acesso em: 14 ago. 2012.

PERMONIAM, Viviane A. **Visualização Exploratória de Dados do Desempenho na Aprendizagem em um Ambiente Adaptável**. 2008. 124f. Tese (Doutorado em Engenharia Elétrica) – Universidade de São Paulo, Escola de Engenharia de São Carlos, São Carlos, 2008.

PENTEADO, Rosângela de F. S. CARVALHO, Hélio G. de. Gestão do conhecimento e aplicativos de gerenciamento de ideias voltados à inovação organizacional: uma análise de

softwares. **Revista Gestão Industrial**. Universidade Tecnológica Federal do Paraná, Ponta Grossa, v.5, p. 216-233, 2009.

PUTZ, Werner. **The hierarchical Visualization System**: A general framework for visualizing information hierarchies using the example of information Pyramids. 2005. 141f. Thesis (Master), Graz University of Technology. Austria, 2005. Disponível em: <www.iicm.tugraz.at/thesis/wputz.pdf> Acesso em: 20 ago. 2012.

RABELO, Emerson. **Avaliação de técnicas de visualização para mineração de dados**. 2007. 103f. Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual de Maringá, Programa de Pós-Graduação em Ciência da Computação, Maringá, 2007. Disponível em: <www.din.uem.br/~mestrado/diss/2007/rabelo.pdf>. Acesso em: 17 ago. 2012.

RAUTER, André. BENATO, Karina. Visualização da informação aplicada à estratégia competitiva de uma Instituição Educacional. **Gestão da Produção, Operações e Sistemas – GEPROS**. Bauru. ano 2, n. 1, p. 107-115, set/dez. 2006. Disponível em: <<http://revista.feb.unesp.br/index.php/gepros/article/view/125/87>>. Acesso em: 18 ago. 2012.

REZENDE; Denise; ABREU, Aline França de. 2000. **Tecnologia da informação aplicada a sistemas de informação empresariais**: o papel estratégico da informação e dos sistemas de informação nas empresas. São Paulo: Atlas, 2000, 311p.

RODRIGUES FILHO, José Alberto Florentino. **Data mining**: conceitos, técnicas e aplicação. 2001. 125f. Dissertação (Mestrado em Engenharia de Produção) – Universidade de São Paulo, São Paulo, 2001. Disponível em: <www.beto.pro.br/DtMining.pdf>. Acesso em: 13 set. 2012.

ROMANI, Luciana A. S. **InterMap**: Ferramenta para Visualização da Interação em Ambientes de Educação a Distância na Web. 2000. 123f. Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual de Campinas. Campinas, 2000. Disponível em: <www.docsagencia.cnptia.embrapa.br/informatica/InterMap>. Acesso em: 17 ago. 2012.

ROMÃO, Wesley; NIEDERAUER, Carlos A. P.; MARTINS, Alejandro; TCHOLAKIAN, Aran; PACHECO, Roberto C. S.; BARCIA, Ricardo M. Extração de regras de associação em C&T: o algoritmo apriori. In: XIX ENCONTRO NACIONAL EM ENGENHARIA DE PRODUÇÃO. **Anais...** Rio de Janeiro, 1999. Disponível em: <www.abepro.org.br/biblioteca/ENEGEP1999_A0901.pdf> Acesso em: 04 dez. 2011.

ROSAS, Daniel Henrique P. **Análise de Ferramentas de Mineração de Dados Usando Critérios de Usabilidade**. 2009. 57f. Monografia (Especialização em Gestão Industrial) – Universidade Tecnológica Federal do Paraná, Programa de Pós Graduação em Engenharia de Produção, Ponta Grossa, 2009.

SAS. **SAS Interprise Miner para Desktop**. Disponível em: <http://www.sas.com/offices/europe/portugal/images/meta/FactSheet_enterpriseminer.pdf>. Acesso em: 11 maio 2012.

SCAPIN, D; BASTIEN, C. Ergonomic criteria for the evaluation of human-computer interfaces. **Rapport technique de l'INRIA**. 1993. Disponível em: <www.inria.fr/rrrt/rt-0156.html>. Acesso em: 25 nov. 2009.

SCHUCH, Regis; DILL, Sérgio Luiz; SUASEN, Paulo Sérgio; PADOIN, Edson Luis; CAMPOS, Mauricio de. Mineração de dados em uma subestação de energia elétrica. In: PROCEEDINGS OF THE 9TH BRAZILIAN CONFERENCE ON DYNAMICS, CONTROL AND THEIR APPLICATIONS – DINCON'10. **Anais...** Serra Negra, p. 804-810. jun. 2010.

Disponível em: <www.sbmac.org.br/dincon/trabalhos/PDF/energy/68015.pdf>. Acesso em: 17 ago. 2012.

SFERRA, Heloisa Helena; CORREA, Angela M. C. Conceitos e Aplicações de Data Mining. **Revista de Ciência e Tecnologia**. Piracicaba, v. 11, n. 22, p.19-34, jul/dez 2003. Disponível em: <<http://www.unimep.br/phpg/editora/revitaspdf/rct22art02.pdf>>. Acesso em: 12 nov. 2012.

SILVA, Celmar Guimarães. Considerações sobre o uso de visualização de informações no auxílio à gestão da informação. In: XXXIV SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE. **Anais do XXVII Congresso da SBC**. Rio de Janeiro, p. 2070-2084. jul. 2007. Disponível em: <www.lbd.dcc.ufmg.br/colecoes/semish/2007/002.pdf>. Acesso em: 17 ago. 2012.

SPSS. **Clementine 11.1 User's Guide**. SPSS Inc, 2007. Disponível em: <http://www.forms.manchester.ac.uk/applications-media/document/clementine/11.1/ClementineUsersGuide_11.1.pdf>. Acesso em: 11 maio 2012.

TURRIONI, João B.; MELLO, Carlos H. P. **Metodologia de Pesquisa em Engenharia de Produção**: estratégias, métodos e técnicas para condução de pesquisas quantitativas e qualitativas. Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Itajuba. Itajubá: UNIFEI, 2012. 199p. Disponível em: <http://www.carlosmello.unifei.edu.br/Disciplinas/Mestrado/PCM-10/Apostila-Mestrado/Apostila_Metodologia_Completa_2012.pdf>. Acesso em: 08 jul 2012.

UCI. **Machine Learning Repository**. Browse Through: Datasets. Disponível em <<http://archive.ics.uci.edu/ml/datasets.html>>. Acesso em: 22 maio 2013.

VALIATI, Eliane Regina de Almeida. **Avaliação de Usabilidade de Técnicas de Visualização de Informações Multidimensionais**. 2008. 220f. Tese (Doutorado em Computação) – Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Computação, Porto Alegre, 2008. Disponível em: <<http://hdl.handle.net/10183/13699>>. Acesso em: 19 ago. 2012.

VIEIRA, Jessica M de L.; CORRÊA, Renato, F. Recuperação de informação através de recursos visuais. In: ENCONTRO NACIONAL DE ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, GESTÃO E CIÊNCIA DA INFORMAÇÃO. **Os desafios do profissional da informação frente às tecnologias e suportes informacionais do século XXI: lugares de memória para a biblioteconomia**. Paraíba, jul. 2010. Disponível em: <<http://dci.ccsa.ufpb.br/enebd/index.php/enebd/article/view/19/22>>. Acesso em 19 ago. 2012.

YAMAGUCHI, Juliana Keiko. **Diretrizes para a escolha de técnicas de visualização aplicadas no processo de extração do conhecimento**. 2010. 182f. Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual de Maringá, Maringá, 2010. Disponível em: <www.din.uem.br/~mestrado/diss/2010/yamaguchi.pdf>. Acesso em: 19 ago. 2012.

WEKA. **Data Mining With Open Source Machine Learning Software**. Disponível em <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 20 abril 2012.

WITTEN, Ian H.; FRANK, Eibe Frank. **Data Mining - Practical Machine Learning Tools and Techniques**. Ian H. Witten, Eibe Frank. 2 Ed. Morgan Kaufmann Publishers. Elsevier: USA, 2005.